National University of Ireland – Galway

College of Sciences, School of Chemistry

The study of protein chemical modification using polarized Excitation Emission Matrix (pEEM)

Thesis presented for the degree of PhD
of the National University of Ireland – Galway

by **Ana Luiza de Faria e Silva**
Student ID 16234190

Supervisor: **Prof. Alan G. Ryder**

October 2020

Acknowledgment

First, I wish to thank my supervisor, Prof. Alan Ryder for the trust and for giving me the opportunity to be part of this group and carry out this work. Thank you for the support, encouragement, and guidance throughout this journey.

My appreciation also extends to my colleagues in the Nanoscale Biophotonics Laboratory, specially to Saioa and Yannick for sharing their knowledge with me and Marina for being my friend. Likewise, to the staff members of the School of Chemistry for all the help.

I also wish to thank my financial supporters, for making this work possible. This project was supported in part by a research grant from Science Foundation Ireland and in part under the European Regional Development Fund (under Grant Number 14/IA/2282, Advanced Analytics for Biological Therapeutic Manufacture to Alan G. Ryder).

I would like to express my most sincere gratitude to my parents Luiz and Heloisa for their love, prayers, and sacrifices for educating and preparing me for future; to my sister Maria Clara who has always been my best friend and supporter, and to Anna who was my family in Ireland during this time and made every step of this journey more enjoyable. Without you, this achievement would not have been possible.

And finally, I wish to acknowledge the support of my family, specially my grandparents, and all my friends, and extend my appreciation to all who have directly or indirectly helped me in the completion of this work.

Table of Contents

Commonly used abbreviations

AI	Aggregation Index
AUC	Area under the curve
DLS:	Dynamic light scattering
EJCR	Elliptical Joint Confidence Region
FRET:	Förster resonance energy transfer
G:	G factor
HIC	Hydrophobic Interaction Chromatography
HMWS	High molecular weight species
HPLC	High performance liquid chromatography
IFE:	Inner filter effects
Lys	Lysine
LZ	Lysozyme
MDF:	Multidimensional fluorescence
PAT:	Process analytical technology
pEEM:	Polarized Excitation Emission Matrix
PEG	Polyethileneglycol
Phe:	Phenylalanine
r :	Anisotropy
REP	Relative error of prediction
rIgG	Rabbit Immunoglobulin G
RMSE	Root mean square error of prediction
ROBPCA	Robust Principal component analysis
RS	Rayleigh Scatter
RSD	Relative standard deviation
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SimI	Similarity Index
SM	Starting material(s)
stdev	Standard deviation
Trp / W	Tryptophan
Tyr / Y	Tyrosine
UV-Vis:	Ultraviolet-Visible absorbance spectroscopy

Abstract

The advances in drug development and the increased use of biologics as therapeutics drives the need for more effective Process Analytical Technology (PAT) tools to measure biophysical changes during manufacturing and formulation processes. This is particularly true for proteins undergoing chemical modification for the production of new entities such as Antibody drug conjugates (ADCs), PEGylated proteins or radioimmunoconjugates. The measurement of structural quality in solution can be challenging and time consuming, but intrinsic protein fluorescence (IPF) because of its high sensitivity, ease of use, and when implemented in via multi-dimensional techniques like polarized Excitation Emission (pEEM) spectroscopy, its high information content, might offer a solution.

Conjugation reactions potentially exposes the protein to chemical and physical stressors such as the modification of specific amino acids and agitation, which can affect protein stability triggering unfolding, aggregation, and/or fragmentation, and potentially lead to reduction in process yields, loss of activity and/or immunogenicity issues. This, combined with the increased structural complexity of proteins and its conjugates, requires the use of more sensitive quality control (QC) tools to better identify and assess biophysical and structural changes during manufacturing.

Here we demonstrate how pEEM measurements can be used in combination with simple chemometric tools like Principal Component Analysis (PCA) and partial least squares (PLS) to discriminate the raw product solutions according to the degree of PEGylation and also to predict conjugation degree (PEG to protein ratio, PPR) with good accuracy (root mean square error, RMSE for calibration ~10%, relative error of prediction, REP< 20%), in comparison to the reference technique Size Exclusion Chromatography (SEC) (SEC error: ~7.2%). The use of variable selection tools allows for similar predictions obtained with faster (from ~7min to less than 60s) and simpler two-dimensional spectra, which makes the method a more viable PAT tool for in or online measurements.

Furthermore, pEEM offers a reproducible and fast alternative to simultaneously measure protein concentration (RMSE<0.01 g/L), asses structural variance, and particle/aggregate content. It allows one to generate quantitative

prediction models for non-reversible aggregation content as described by SEC, and obtain qualitative information about reversible aggregate content, which cannot be obtained from SEC

Finally, we show the feasibility of using pEEM for assessing and monitoring structural changes during the reaction. By using IgG PEGylation as a model system, we show that the three levels of information extracted from pEEM can be used for in-situ assessment of conjugation reactions: Rayleigh scatter (RS) to probe aggregate/particle formation, both transient and permanent; intrinsic fluorescence emission to identify subtle chemical and structural variations in the parent protein, and *aniso*-EEM maps that corroborated protein structural changes.

Chapter 1. Introduction

1.1. Biotherapeutics

Biotherapeutics are defined as materials manufactured by biological means and are used for therapeutic purposes. Peptides, oligonucleotides, vaccines, recombinant proteins, antibodies, its fragments and conjugates are some examples of these biopharmaceuticals.

Biologics have been widely used as therapeutic agents since 1982 when the US Food and Drug Administration (FDA) approved the first recombinant protein, R-insulin, to treat *Diabetes mellitus.*[1] The use of these macromolecules as therapeutics brings numerous advantages in comparison to small-molecule active pharmaceutical ingredients (APIs). For instance, the complex structure, specific conformation, and large size provide high selectivity, which is usually linked with less frequent side effects.[1, 2] However, the increased complexity also leads to lower stability and usually more demanding production process and characterization, often requiring more knowledge, time and resources. Some of the main differences between small and big molecule agents are listed in Table 1.1.

Table 1.1- Differences between small molecule and big molecule therapeutics or biotherapeutics in terms of size, structural complexity, selectivity, stability, production process, and characterization needs.

Parameter	Small molecules	Biotherapeutics
Size/MW[a]	<1kDa	>10kDa
Structure	Well defined structure	Higher order complex tridimensional structure.
Selectivity	Lower selectivity	High selectivity
Stability	Higher stability	Lower stability
Production	Via chemical synthesis: relatively simple process and purification. Less expensive.	Via recombinant DNA or hybridoma technology: complex processes with a large number of steps. Usually coasty.
Characterization	Various and well-defined suitable characterization methods. Usually characterized by simple HPLC[a] or spectroscopy.	Complex composition and heterogeneity make complete characterization more challenging. Usually a combination of techniques is necessary.

[a] MW stands for molecular weight and HPLC stands for high performance liquid chromatography

1.2. Proteins

This section provides details on the two proteins used in this study. Lysozyme was chosen because it is a smaller, well studied, and less complex model protein, while IgG is widely used for therapeutic purposes and represents the case of a more structurally complex protein.

1.2.1. Lysozyme (LZ)

Lysozyme, also known as muramidase or N-acetylmuramichydrolaseis, is a relatively small and stable enzyme, which is widely used in research as a model system for studying protein structure, stability, function and chemical modification.[3, 4] It is naturally found in the body in secretions like tears, saliva, and milk, playing an important role as antimicrobial agent because of its ability to cleave some structural subunits of gram-positive bacterial cell walls.[5]

It is a 14kDa polypeptide, composed of 129 amino acids arranged in two domains and stabilized by four disulphide bonds. The main source for its monomeric form is hen egg white which is widely available commercially.[6, 7]

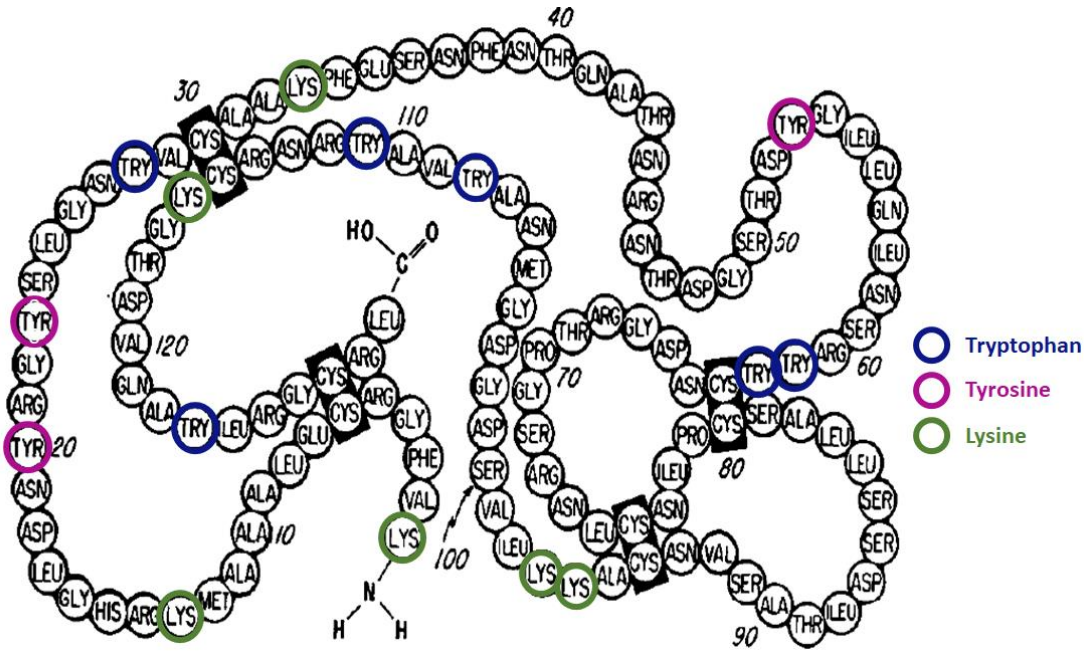


Figure 1.1- The structure of egg white lysozyme indicating the positions of the four disulphide bonds. Content reproduced from Canfield at al [7] with permission from American Society for Biochemistry and Molecular Biology. The figure was modified to highlight Trp (blue), Tyr (pink) and Lys (green) residues.

Lysozyme intrinsic fluorescence emission arises from the presence of six tryptophan and three tyrosine residues, being Trp62 and Trp108 the dominant native state emitters.[8] The six lysine (lys) residues within the structure are the most common targets for conjugation via amine chemistry and this type of reaction has been extensively discussed in the literature. This makes lysozyme conjugation an ideal model system for our study. (Figure 1.1)

1.2.2. Immunoglobulin G (IgG)

Antibodies (Ab) also referred to as immunoglobulins (Ig) are critical components of the immune systems produced by all vertebrates in a variety of types. Monoclonal antibodies (mAbs) have been applied as therapeutic agents since 1986 when the US FDA approved the first mAb, Muromonab-CD3, to prevent rejection reactions after transplantation.[9]

IgG is the most widely used Ig type for therapeutic purposes. It is a globular Y-shaped protein composed of heavy (H, 50 kDa) and light (L, 25 kDa) polypeptide chains that form two fragment antigen-binding (Fab) regions and one domain responsible for the effector function (Fc, Fragment crystallizable).[10, 11] Covalent (intra and interchain di-sulphide bridges) and non-covalent forces (hydrogen bonds, ionic bonds, van der Waals interactions) are responsible for stabilizing the structure and keeping the native arrangement of the antibody.[12] (Figure 1.2)

Each antibody arm is called a Fab fragment and is composed of a light chain (LC) and the end portion of a heavy chain (HC). It can be divided into two parts with different structural functions: the variable and constant regions. The terminal portion in both H and L chains (the N-terminus region of Fab) is referred to as the variable (V or F_V) region because of its intrinsic variability in amino acid composition (it is composed of 110-130 residues) which plays a critical role in the antigen-binding specificity. This specificity comes from the hypervariable regions, also known as CDR (Complementarity-Determining regions), which are linear polypeptide segments of both H and L chains (CDR1,2,3).[13, 14] The second region, the non-variable part of the Fab, is composed of the constant light (CL) and the first constant portion of the heavy chain (CH_1).[12] The base of the Y in an antibody, also referred to as the Fc domain, consists only of heavy chains (CH_2 and CH_3). It is connected to the Fab portion by the

hinge region, which is a polypeptide sequence found between CH_1 and CH_2 domains responsible for mAb flexibility, allowing free movement of the two antibody arms.

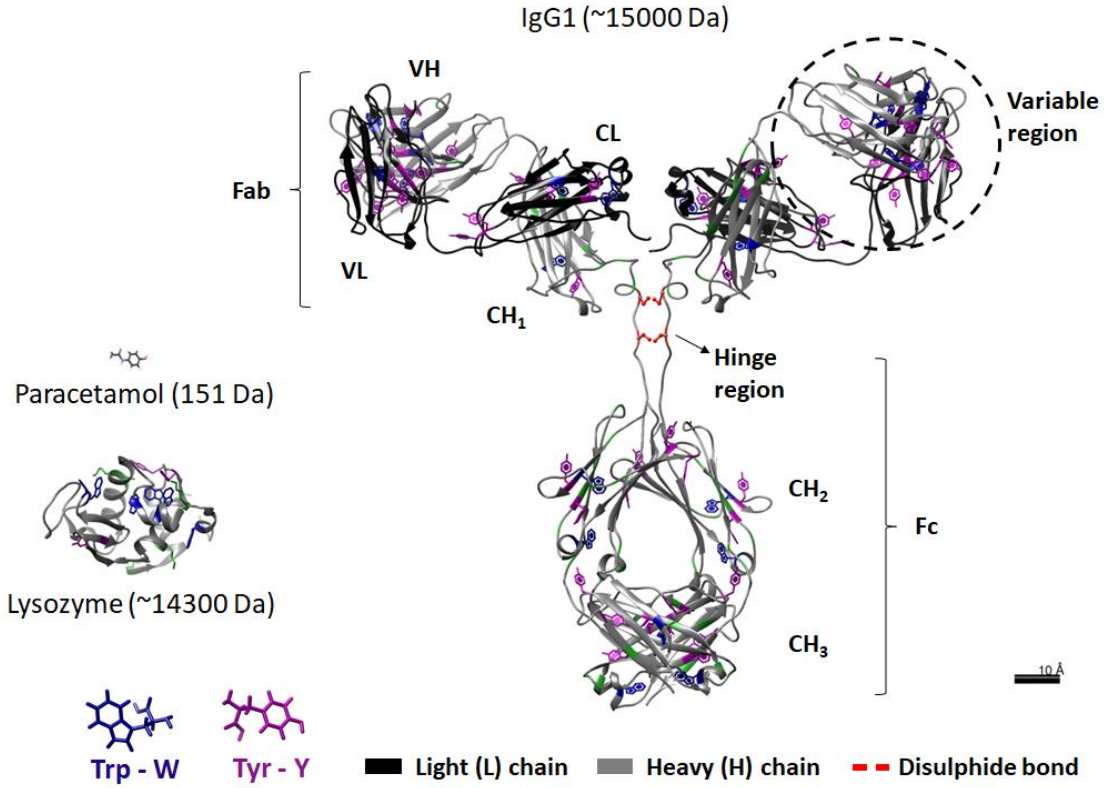


Figure 1.2- Size comparison between a small molecule therapeutic (Paracetamol), a small (Lysozyme) and a large and more complex protein (IgG). Crystal structure of hen egg-white lysozyme (PDB entry: 1DPX, Chimera software) and human IgG shows the location of main fluorophores (Trp, W and Tyr, Y).

The native structure of a protein is determined by, and maintained via a combination of covalent and non-covalent interactions. The strongest covalent forces are the disulphide bridges, which are formed by the oxidation of two cysteine residues to form covalent S-S (sulphur-sulphur) bonds both intra and interchains.[15] Interchain bridges are found between heavy chains or attaching heavy and light chains in the Fab domain, while those called intrachain bonds provide stabilization to peptides in a single chain. Apart from the structural function, disulphide bridges are also target for conjugation of small molecules.[16] Non-covalent forces (e.g. Hydrogen bonds, ionic bonds, van der Waals interactions, and hydrophobic bonds) are also critical in maintaining the three-dimensional structure of proteins and allow the maintenance of the structural arrangement even if the disulfide bonds are broken.[17-19] The general arrangement of domains in an Ig is shown in Figure 1.2.

Classes of immunoglobulins:

There are five classes of immunoglobulins[b]: IgA, IgM, IgE, IgG, IgD, with IgG being the most abundant type found in the serum (70-75%).[13] In addition to the high plasma abundance, the high affinity and long half-life in plasma make the IgG isotype an ideal candidate for generation of therapeutic antibodies.[20] IgGs can be found in the bloodstream in four different isoforms: IgG_1 (60-70%), IgG_2 (20-30%), IgG_3 (5-8%), and IgG_4 (1-3%). The main differences between them are the heavy chain type and sizes as well as the pattern of linkage in the hinge region.[12] (Appendix 7.1) A typical human IgG_1 (hIgG) molecule contains two 50 kDa HC and two 25 kDa light chains making a total of 150 kDa for the whole molecule and 16 disulphide bonds which stabilizes the individual mAb domains. There are two bonds in the hinge region linking the two HC, one in each antibody arm connecting HC and LC, and 12 intrachain bridges.[16]

The antibody of choice for this study was a rabbit IgG (rIgG) because it is easily found in an isolated isoform, relatively similar to $hIgG_1$ (the most used for therapeutic purposes) and cheaper compared to human monoclonal antibodies. When it comes to structure, rIgG differs to hIgG on the number of disulphide bonds: whereas hIgG has two interchain bonds attaching the two heavy sequences, the rIgG only possess a single interchain disulphide between heavy chains. rIgG also presents a slightly different arrangement of intra-chain bonds compared to hIgG, as shown by Figure 1.3.[21] For rIgG there are approximately 50 Tyr and 24 Trp residues contributing to the overall intrinsic fluorescence.

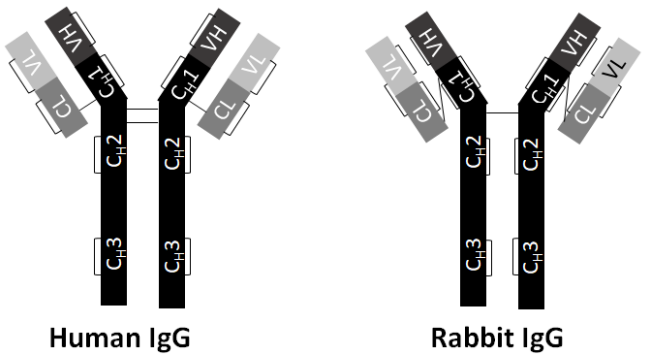


Figure 1.3- Schematic representation of human IgG1 and rabbit IgG highlighting the main structural differences in terms of disulphide bond location and number.[16, 21]

[b] See Appendix 7.1 for more details on the differences between the various classes in terms of size and structural arrangement

1.3. Protein Stability

Proteins are organized in four levels of structure: primary, secondary, tertiary and quaternary structure. (Figure 1.4) The primary structure is the linear amminoacid sequence linked via covalent peptide bonds, while the secondary structure refers to the local arrangement of residues within a polypeptide sequence, which is usually formed via hydrogen bonds. There are two major types of secondary arrangements: alpha (α) helix and beta (β) sheets. The spatial arrangement of the protein via (mostly) electrostatic interactions forms what is known as the tertiary structure and the process by which a protein enters this three-dimensional arrangement is called folding. Finally, the quaternary structure is the association of subunits of tertiary structure to form an oligomeric assembly that carries a function.[18, 22, 23]

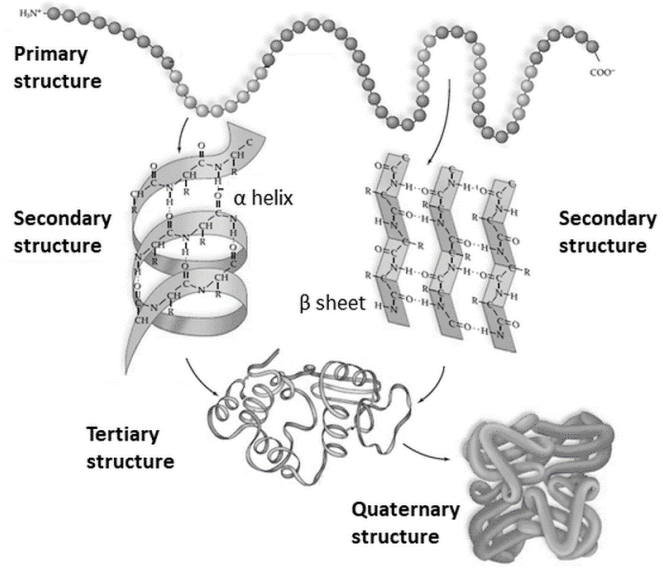


Figure 1.4- Diagram showing the different levels of protein structure: primary, secondary, tertiary and quaternary structure. Reproduced from [24] with permission from publisher.

Proteins are usually found in solution in an equilibrium between folded and unfolded species and protein stability can be described as the balance of forces determining if a protein will be in the called native conformation (folded form) or a denatured (unfolded) state.[18, 25] Hydrophobic, electrostatic, van der Waals interactions, hydrogen, and disulphide bonds and are known to stabilize the folded state, whereas conformational entropy and interactions between the residues and solvent are said to energetically stabilize the denatured state.[18, 19, 26] Because protein function depends upon the maintenance of the native arrangement, any factor that induces structural changes can potentially affect activity. For instance, temperature[27] and solution

condition (pH, ionic strength)[28-30] are some factors that can affect protein conformational stability.

The process of aggregation is a common issue in therapeutic mAb production and it is known to affect the quality, safety, and efficacy of the final product. Various are the processes where aggregation can occur during the manufacturing: from the fermentation, purification, and processing to the storage and transportation.[31] Many factors can affect aggregation propensity and several aggregation mechanisms have been described.[32, 33] For instance, protein structure plays an important role in stability as the presence of hydrophobic residues creates the called aggregation prone regions (APRs) which can facilitate aggregation processes particularly if protein secondary and tertiary structures are altered. Thus, processes that modify the structural arrangement of the protein will have an impact on aggregation proneness (e.g. deglycosylation).[34] Similarly, chemical degradation processes like oxidation or deamidation can induce the formation of high molecular weight species (HMWS) by destabilization of the protein and/or increasing surface hydrophobicity.[34, 35] Fragmentation, another common degradation pathway, can affect potency/efficacy of biologics and is also described as a triggering factor of aggregation, although the mechanisms and real contribution are still not well defined.[34, 36]

Cross-linking reactions like dityrosine formation and disulphide exchange are also known to facilitate aggregation.[32, 37] Changes in temperature, protein concentration, pH, solvent ionic strength, and oxygen exposure during freezing/thawing and agitation, or interactions with metal surfaces may lead to post-translational modification which could result in the irreversible formation of aggregates.[38]

Furthermore, protein chemical modification (e.g. conjugation of small molecules) can also affect aggregation propensity.[39] The attachment process usually involves chemically stressful conditions such as the modification of specific amino acids or disruption of covalent bonds and physical stressors such as agitation, which can result in unwanted product formation. These can trigger changes in folding/unfolding routes, induce aggregation, and or fragmentation, leading to loss of activity and immunogenicity issues,[40] and/or reduction in process yields.[41, 42] In

addition to specific properties of the small molecule surrogates like hydrophobicity may also enhance aggregation propensity.[31]

1.4. Protein modification

The chemical modification of proteins to produce new functional entities such as PEGylated proteins,[43] Antibody Drug Conjugates (ADCs)[44, 45] or radioimmunoconjugates[46] is becoming more widespread as the use of biotherapeutics continues to increase. Among all the benefits of the conjugation, the enhanced therapeutic potential, improved pharmacokinetic[c] and reduced immunogenic response are common advantages.[43, 45, 46] There are many biologics commercially available nowadays and PEGylation is one of the most widespread strategies used in marketed products. Table 1.2 shows some examples of conjugates currently in the market.

Table 1.2- List of some marketed protein conjugates, company, small and big molecules composing the structure, and conjugation chemistry (*vide infra*).[46-48]

Product name	Company	Type of conjugate	Biologic	Surrogate	Linker method
Kadcyla®[49]	Genentech	ADC	IgG1 (Trastuzumab)	Cytotoxic drug DM1[d]	Lysine conjugation
Adcetris®[50]	Seattle Genetics	ADC	IgG1 (Brentuximab)	MMAE[c]	Cysteine conjugation
Somavert®[51]	Pharmacia & Upjohn/ Pfizer	PEGylated conjugate	Human Growth hormone	Linear 5 kDa PEGs (4-6 residues)	Lysine conjugation
Zevalin®[52]	IDEC Pharmaceuticals	Radio-immuno conjugate	IgG1 Ibritumomab	Yttrium-90	Lysine conjugation
Cimzia®[53]	UCBPharma	PEGylated conjugate	mAb Fab' fragment	40kDa PEG	Cysteine Chemistry

Protein conjugates can be prepared using a variety of chemical strategies and both the level of conjugation and the sites within the antibody to which the small molecule is bound to, are key factors determining the final stability of the conjugates. Chemical modification via natural amminoacids is an attractive strategy because it does not require the prior modification of the protein. Lysine and cysteine residues the most common natural targets.(Figure 1.5) Conjugation to the amine group of lysine

[c] Pharmacokinetics is the "process of the uptake of drugs by the body, the biotransformation they undergo, the distribution of the drugs and their metabolites in the tissues, and the elimination of the drugs and their metabolites from the body over a period of time." [22]

[d] DM1 stands for Mertansine and MMAE stands for monomethyl auristatin E, which are highly potent cytotoxic agents. [49 50]

residues (amine chemistry) is one of the most used strategies of conjugation because it is present in many proteins, and the reaction is relatively simple, usually involving just one step. For instance, IgG1 has approximately 80 reactive amine groups[54] from lysine residues of which around 10 are accessible sites for chemical modification[55] whereas lysozyme has six residues. Conjugation is based on the reaction between exposed ε-amino groups with an active reagent with groups like N-hydroxysuccinimidyl (NHS), p-nitrophenylcarbonate (pNPC), or aldehyde. One of the drawbacks of this strategy is high heterogeneity of the final products which are usually composed of several species with different properties and physicochemical characteristics.[56] [57]

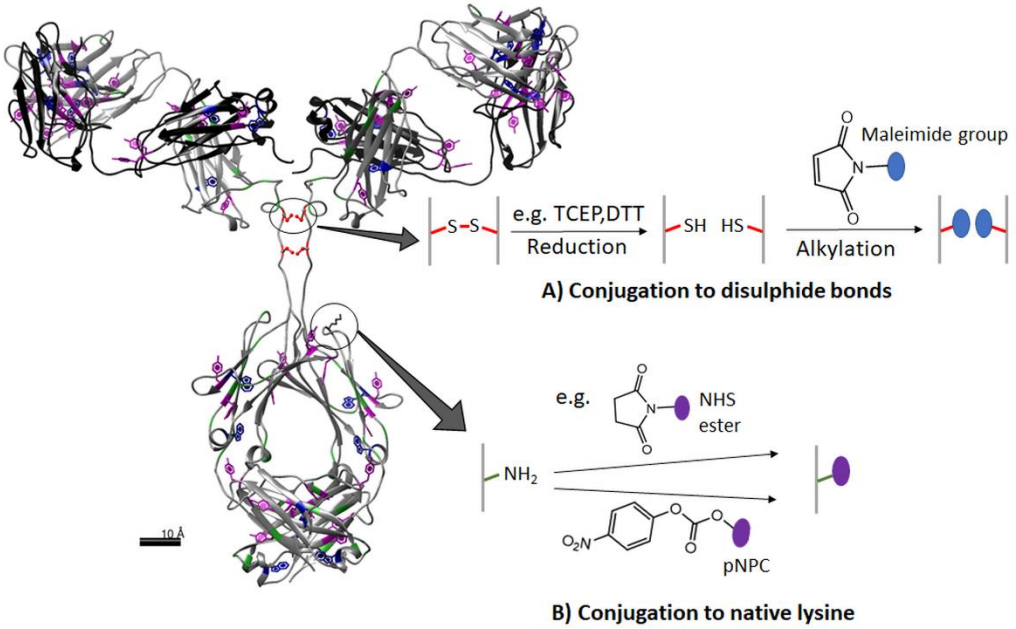


Figure 1.5- Schematic representation of protein conjugation via (A) disulphide bond and (B) native lysine modification showing the main sites of attachment (using IgG is used as an example), usual chemical routes and some of the common reagents used in each approach.

Conjugation to cysteine (thiol chemistry) is another common approach. Because of the lack of free thiol groups from cysteine in most proteins, the conjugation usually first involves the reduction of disulphide bonds under controlled conditions to create sulfhydryls (Cys-SH) available for conjugation to active reagents containing groups like maleimide. This route usually restricts the attachment sites to six or eight for IgG for example (depending on the number of disulphide bonds in the protein). The big advantage of this strategy is the lower heterogeneity in comparison to the lysine attachment mechanism.[56] Other strategies like engineered cysteine technology,

cysteine rebridging and non-natural amminoacid addition have also been used for producing bioconjugates,[58, 59] but here we focused only on the conjugation via natural lysine and disulphide bonds because they are the most common routes and the ones chosen for this study.

The conventional conjugation strategies usually lead to heterogeneous products with different numbers of conjugated payload, each of which may have different physicochemical properties making characterization and purification more difficult.[2] This, combined with the potential unfolding, fragmentation, and/or aggregation during conjugation[35] and the increased structural complexity of the modified proteins, makes the characterization of the parent proteins and its conjugates a demanding and challenging process.[60-63]

1.5. Analysis of proteins and conjugates

The pharmaceutical industry is moving towards the implementation of the called pharmaceutical Quality by design (QbD) which is "a systematic approach to development that begins with predefined objectives and emphasises product and process understanding and process control, based on sound science and quality risk management" according to the ICH Q8 document.[64] Process Analytical Technology (PAT) is one of the strategies to facilitate the QbD initiative. PAT is defined as in-process timely measurements of quality or performance parameters, implemented to design, analyse and control manufacturing with the final objective of ensuring final product quality.[65]

When it comes to proteins, the use of PAT tools becomes even more critical because of the increased instability, usually longer manufacturing times, which often leads to more prolonged exposure to stressors. Thus, there are some aspects of the samples, either physical, chemical, or biological, which have to be assessed and meet specifications in order to ensure the desired product quality. These are called Critical Quality Attributes (CQAs).[64] Table 1.3 summarizes some of the main CQAs for protein based samples.

Generally speaking, the most important quality aspects to be considered during a protein chemical modification process are homogeneity, purity, degree of conjugation, total protein concentration, and lot-to-lot variability for starting materials,

intermediates and the final conjugates. These CQAs are known to impact product potency, pharmacokinetics, and toxicity.[61,66,67] While small molecule adducts are usually characterized using various standard chromatographic or spectroscopic methods, protein and conjugates analysis is usually more demanding because of the increased structural complexity.

Table 1.3- List of potential CQAs for proteins/bioconjugates and analytical tools commonly used for assessing each.[61, 67]

CQA[e]	**Technique**
Total protein concentration	Absorbance Spectroscopy Colorimetric methods
Aggregation / Fragmentation	Size-exclusion chromatography (SEC) Dynamic Light scattering (DLS) Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)
Purity	Mass spectrometry (MS) SEC SDS-PAGE
Charge variants	Ion exchange chromatography (IEC)
Glycosylation	MS
Degree of conjugation (Surrogate to protein ratio / conjugates distribution)	Absorbance Spectroscopy SEC DLS SDS-PAGE

SEC is a well-established and the most common method for assessing HMWS content in protein-based samples,[68, 69] as well as size changes with conjugation.[70] SEC separates molecules based on their hydrodynamic radius (or Stokes radius) and their different interaction with the pores within the column packing. Bigger molecules tend to pass through the pores eluting earlier while small molecules will penetrate the pores showing longer retention times.[71] Despite its extensive use, it has many disadvantages. Firstly, sample contact with buffer components (e.g. salts/organic solvents) and column surface, and the dilution of the sample with analysis can induce changes in the sample and/or increase secondary interactions generating misleading results.[72] Also, the long column equilibration times and run times (mainly when it is necessary to use bigger columns to increase resolution) preclude its use as an online, inexpensive monitoring tool for protein analysis. Furthermore, SEC may not always be able to

[e] The CQAs listed in the table are only some of the aspects that potentially influence the quality of a naked protein or a conjugate. Despite being very common factors, a case-by-case evaluation is the ideal approach to determine the CQAs for a given sample, as well as the best assessment method in each case.

measure weakly bound soluble aggregates, which are often too small for measurement by light scattering techniques. Insoluble aggregates are also often too big to enter the column and will not be detected.[73, 74]

DLS is another technique vastly used to assess homogeneity of protein based samples[75, 76] and size changes with conjugation (e.g. PEGylation).[77] It measures the Brownian motion of dispersed particles which can be related to Hydrodynamic radius (R_h) using the Stokes-Einstein equation.

$$D = \frac{K_B T}{6 \pi \eta R_h} \qquad \text{Equation 1.1}$$

where K_b is the Boltzman constant, T is temperature and η is the absolute viscosity of the sample. DLS analysis, however, suffers from low robustness and resolution, when the particles are very close in size.[78]

SDS-PAGE is another size-based technique, which qualitatively assesses molecular weight variances based on differential migration through a gel under the influence of an applied electrical field. It has been widely used for assessing physical homogeneity and purity of proteins,[79-81] and also for assessing conjugation and purity/integrity of final conjugates.[77] One of the main drawbacks is protein denaturation prior to analysis and the lack of quantitative information.[82]

UV-Visible absorbance spectroscopy is a convenient and straightforward spectroscopic method for the routine, simple analysis of protein. The absorbance at 280 nm (A_{280}) is commonly used for assessing total protein concentration, and it is sometimes an alternative to colorimetric methods such as Bradford assay, which requires the addition of reagents to the sample and thus are destructive to a certain degree.[83] The basis for the quantitative analysis using absorbance spectroscopy is the Beer-Lambert Law which is the linear the correlation between absorbance and analyte concentration.[84] Absorbance spectra have also been used for online monitoring of conjugation[85] and assessing the degree of conjugation when the attached small molecules absorb at different wavelengths to the protein substrate.[86]

HIC has been used for assessing degree of conjugation based on changes in hydrophobicity with the attachment of small molecules to proteins (e.g. ADCs),[87] and to a lesser extent, for the analysis of protein aggregates or fragments.[88] HIC exploits the interaction of hydrophobic regions in the sample with a stationary phase containing

hydrophobic groups like butyl, phenyl, and alkyl, usually under a gradient of salt concentrations.[89] Charge variance in biopharmaceutical preparations can be caused by antibody fragmentation, oxidation, isomerization, deamidation, aggregation, and glycation.[90] Ion Exchange Chromatography (IEC) has been extensively used for the separation and determination of ionic compounds in protein stability and purity assays, based on electrostatic interactions between the analyte and the ionic groups attached to the stationary phase.[91]

MS can be considered the technique of choice for a more detailed analysis of proteins and has been used in many studies involving intact and chemically modified proteins.[75, 77,92, 93] The measurement involves the ionization of the sample in order to generate molecular fragments which are then measured in terms of ion abundance versus m/z (mass-to-charge ratio).[94] However, often requires extensive sample preparation, is usually linked to a chromatographic step and generates complex data, which are factors that limit its use as a routine tool.[95]

Other techniques like Differential Scanning Calorimetry (DSC), circular dichroism (CD) and Fourier-transform infrared spectroscopy (FT-IR) have also been employed in the protein characterization scenario, particularly for assessing conformational changes. DSC is a technique for evaluation of material transitions when the sample is subject to controlled changes in temperature.[96] It has been particularly useful for protein tertiary structure assessment (presence of native folded and unfolded/partially unfolded states) and to study the effect of conjugation on the conformational stability of proteins.[75, 97] CD and FT-IR are frequently used techniques for protein secondary structure and dynamic studies. CD spectroscopy takes advantage of the different interaction of molecules with circular left and right polarized light to rapidly assess α-helix and β-sheet content in protein structure.[70, 77, 98, 99] FT-IR on the other hand, exploits the absorption of infrared light to determine molecular composition and structure.[100-102]

Each technique has advantages and disadvantages; however, no single technique can provide all the information required to properly assess protein quality in terms of the main CQAs. Thus, the use of a combination of techniques is sometimes the only alternative, which will add to the analysis time, and may be impractical in some

situations. A summary of the pros and cons of some common techniques for protein analysis is provided in Table 1.4.

Table 1.4- Some of the advantages and limitations of the main analytical techniques used for assessing the key CQA of protein-based samples. Modified from [103]

Technique	Advantages	Limitations
SEC	- Robust; - Sensitive; - Precise.	-Time consuming and destructive; - Limited resolution/ limited particle size range; - Possible interactions sample-matrix and dilution;
DLS	- Non-destructive; - High sensitivity; - Low sample consumption.	- Does not provide quantitative information; - Low resolution (species with close size); - Very sensitive to contamination.
SDS-PAGE	- Relatively easy to perform; - Relatively low cost.	- Sample denaturation (Does not detect weakly bound species); - Quantification can be difficult; - Requires staining.
UV-Visible Absorbance spectroscopy	- Non-destructive; - High sensitivity; - Easy to perform; - Detection of scattering material outside absorption bands;	- Limited information particularly on particle properties
MS	- Provides detailed structural information - Very high resolution - High accuracy, precision, and high sensitivity	- Usually requires sample preparation; - Complex data generated; -Expensive equipment;
FT-IR	- Non-destructive; - Solid state analysis possible.	- High concentrations needed; - Limited information
CD	- Non-destructive; - Low protein amount required (far UV); - Possibly used for online detection.	- Limited resolution/information; - Complicated data interpretation;

1.6. Fluorescence

Fluorescence spectroscopy is a well-established technique employed in different areas such as food analysis, imaging and biopharmaceutical industries. It has become an important analytical tool for structural analysis of biopharmaceuticals [104] because of its high selectivity and sensitivity, low structural perturbation in comparison to

alternative methods, and possibility of implementation on-line/at-line, which makes it a potential PAT tool.

1.6.1. Principles of Fluorescence

When electrons in a molecule are excited by incident light, they are promoted from the lowest electronic state called ground state (S_0) to an excited singlet state (either S_1 or S_2) by absorbing a photon of light (represented by a purple arrow in (Figure 1.6). When in the excited state, the molecule can undergo non-radiative transitions[f] like internal coversion or vibrational relaxation which are relatively fast processes generally in the time scale of 10^{-11} to 10^{-9}s and 10^{-12} to 10^{-10}s respectively. The first (wavy black arrow) occurs when the molecule is excited to higher energy states (S_2 or higer) and returns to the lowest level S_1. Vibrational relaxation (dotted black line) represents the relaxation of the molecule to lower energy levels within the same state by dissipation of energy to the environment.[22, 105, 106]

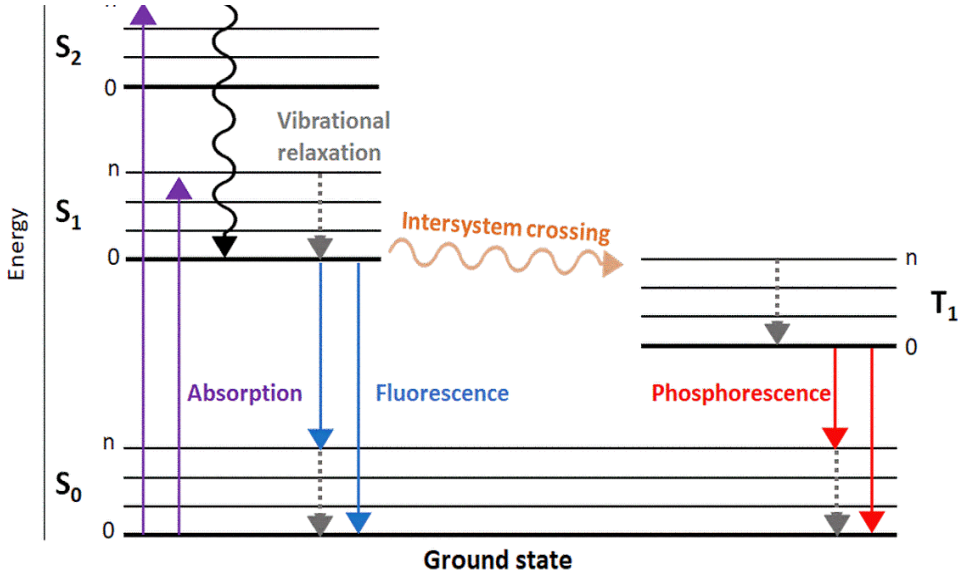


Figure 1.6- Representation of one form of Jablonski diagram showing the energy states of a molecule. The arrows indicate the possible electronic transitions to different states or to different vibrational levels within each electronic state.[105]

Because the energy gap between S_1 and the ground state is much bigger, in order to return to the ground state the molecule can undergo a radiative transition. One of the mechanisms by which a molecule can return to the more thermodynamically stable

[f] Non -radiative transitions are transition between two energy states which do not involve absorption or emission of a photon, while radiative transitions involve absorption/emission.

ground state is fluorescence (blue arrow), which is the emission of a photon (radiative transition) between two states of same spin multiplicity (usually from S_1 to S_0). Phosphorescence (red arrow) is a competing process and is characterized by the emission of a photon between two states of different spin multiplicity. It occurs after molecules undergo a non-radiative transition from S_1 to the triplet state (T_1) called intersystem crossing, and then return to the ground state by emitting a photon. While fluorescence is a rapid process with lifetime[g] (τ) typically ranging from 10^{-10} to 10^{-7} s whereas the phosphorescence has longer lifetime (usually 10^{-5} seconds or slower) due to the forbidden transition from T_1 to ground state.[22, 105, 106]

Molecules can also dissipate energy while in the excited stated by transfering energy to surrounding molecules in a process called quenching, which is a competing process with fluorescence and phosphorescence.[105, 106] This will be discussed in details in the next sections.

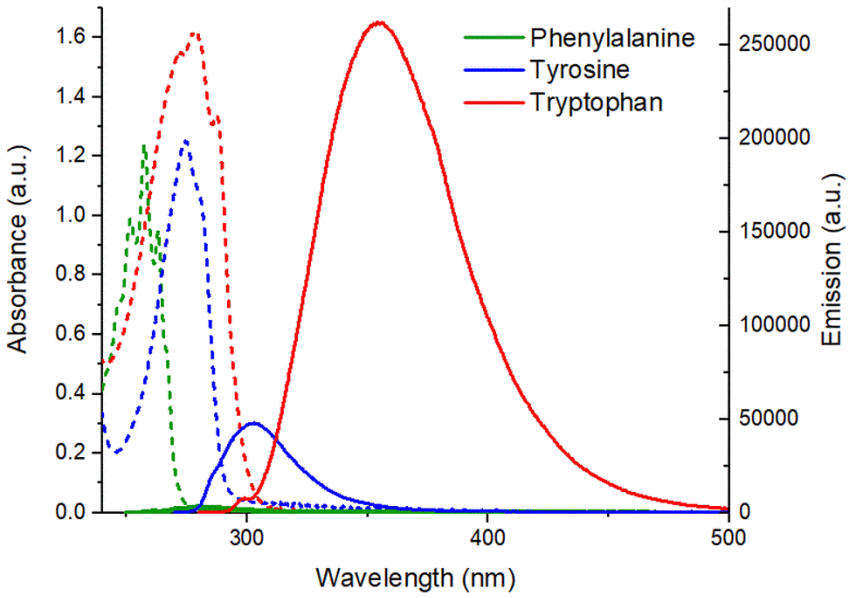


Figure 1.7- Absorption (dotted line) and Emission spectra (full line) of Phe in water (green), Tyr and Trp in Phosphate buffer pH 7 (blue and red respectively). Data was obtained from PhotochemCAD package version 2.1a.

As a consequence of the energy loss in the excited state due to vibrational relaxation, the emission spectrum is located at longer wavelengths (lower energy) compared to the absorption. This difference between the wavelength of maximum emission and maximum absorbance is called the Stokes shift.[106] (Figure 1.7)

[g] Lifetime is the average time an electron spends in the excited state. This concept will be discussed in details in the next sections.

1.6.2. Intrinsic Protein Fluorescence (IPF)

Fluorophores are molecules that can absorb and re emit light. There are many molecules with fluorescent properties but the use of intrinsic fluorophores offers the advantage of being minimally disruptive compared to the addition of extrinsic fluorophores.[107] There are three amino acids responsible for the fluorescence in proteins: phenylalanine (Phe, F), tyrosine (Tyr, Y), and tryptophan (Trp, W).[105] Each of these fluorophores, as well as extrinsic ones, have characteristic fluorescence properties like absorption and emission spectra, lifetime, Stokes shift, and quantum yield. (Table 1.5)

The fluorescence lifetime (τ) is defined as the average time that a fluorophore spends in the excited state prior to emission. This represents the time available for a fluorophore to interact with the surrounding environment before returning to the ground state.[105, 106] The fluorescence quantum yield (Φ_F), in simple terms, is related to the efficiency of the fluorescence process and is defined as the ratio between the number of photons emitted to the number of photons absorbed.[106]

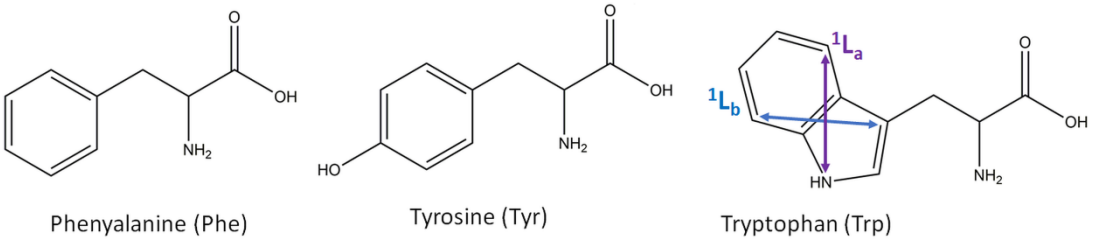


Figure 1.8- Chemical structure of the three intrinsic fluorophores Phenyalalanine (Phe), Tyrosine (Tyr), and Tryptophan (Trp) and Trp excitation states L_a and L_b. [105]

As can be seen in Figure 1.7, Trp shows the strongest absorbance and emission of the three aromatic amminoacids and it is usually the dominant emitter in proteins. Phe on the other hand, has the shortest excitation and emission wavelength (~λ_{ex}/ λ_{em}=260/282) and has a very low quantum yield (and molar absorptivity, Table 1.5). The weak Phe emission is overlapped by the much stronger Trp and Tyr emission, and furthermore Phe emission overlaps absorption bands of both Trp and Tyr leading to FRET *(vide infra)* further weakening of Phe emission. Thus, Phe emission is not observable in most proteins. Tyr has a quantum yield value nearer to that of Trp but, just as Phe, it transfers energy to Trp residues in a very efficient manner, which leads to weaker emission. By changing the excitation wavelength it is possible to

preferentially excite Trp (λ_{ex}=295-305nm) whereas when using excitation λ_{ex}=280 nm, both Tyr and Trp are excited.[105]

1.6.3. Factors affecting emission spectra

There are many factors related to the interaction of fluorophores with its surrounding environment that can affect fluorescence properties. Here the main properties and processes affecting the emission are discussed.

One of the main characteristics of Trp emission is the high sensitivity to its local environment (as opposed to Tyr), which makes its emission spectra particularly useful for monitoring conformational changes, denaturation, and binding in protein-based samples. Upon a change in the polarity of the environment from non-polar to completely water exposed, the emission maxima shift towards longer wavelengths (increase in the Stokes shift). In non-polar environment, Trp fluorescence is characterized by emission at λ_{em}=330-332nm, λ_{em}=340-342nm if water exposure is limited, and λ_{em}=350-353nm when completely exposed to water in polar environments. These different populations also exhibit different fluorescence lifetime and quantum yield. (Table 1.5) When various Trp residues are present in the molecule, the emission spectra will be an average of all species. [104, 108]

Table 1.5- Fluorescence properties of the three intrinsic fluorophores Phe, Tyr and the three classes of Trp. The data refers to the fluorescence measured at room temperature.[104, 105, 108]

Fluorophore		**~λ_{ex} (nm)**	**Absorptivity (M^{-1} cm^{-1})**	**~λ_{em} (nm)**	**Lifetime τ (ns)**	**Quantum Yield Φ_F**
Phe		260	195	282	6.8	0.02
Tyr		275	1405	304	3.6	0.13
Trp	Buried (no exposure to water)	>295	5500	330	2.1	0.11
	Limited water exposure			340	4.4	0.30
	Completely water exposed			350	5.4	0.20

Another characteristic that contributes to the complex emission of Trp is the existence of two excited states: ^1L_a and ^1L_b, which present different excitation/emission, anisotropy, and also different sensitivity to changes in local environment (Figure 1.8). In a completely non-polar environment, ^1L_b has the lowest energy and it is thus the emitting state. In contrast to ^1L_b, the L_a transition is sensitive

to the local environment and becomes more stable in more polar surroundings, becoming the major emitter. The reason for this difference between ^1L_a and ^1L_b is believed to be the dipole moments of the two excited states: ^1L_a transitions involve the nitrogen of the -NH group which has a stronger moment and interacts with polar solvents via hydrogen bonding. (Figure 1.9B) Both ^1L_a and ^1L_b transition states are excited at shorter wavelengths, but the environment sensitive ^1L_a is selectively excited at λ_{ex}>295 nm.[h] [104, 105, 108]

pH and temperature can also affect the emission properties. For instance, Trp has different lifetimes depending on the ionization state: τ =1.2 , 3.0, and 8.8 for the cationic, neutral, and anionic forms respectively.[109] And Tyr emission is quenched (*vide infra*) by its ionized form, Tyrosinate, at high pH (pH>10.7).[104, 110] Also, an increase in temperature is usually accompanied by a decrease in quantum yield because collisions with solvent molecules and intramolecular vibration/ rotation occur more efficiently at higher temperatures.[106]

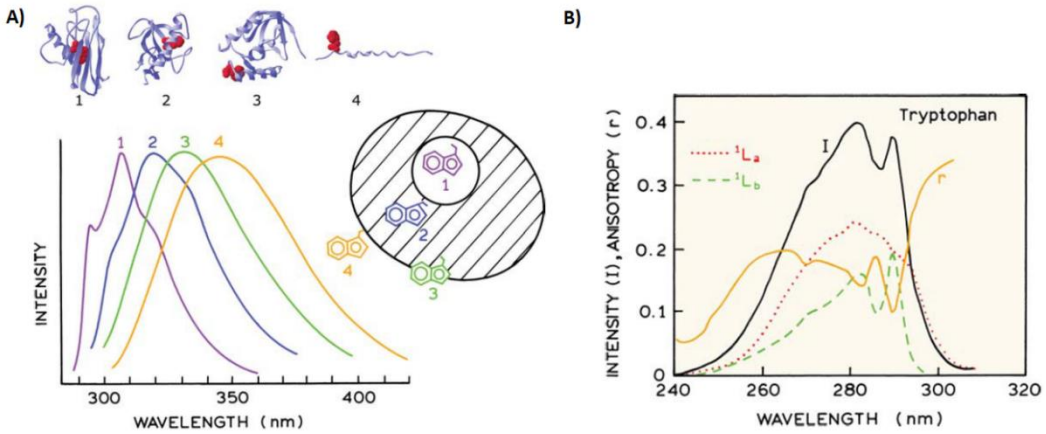


Figure 1.9-(A) Effect of tryptophan environment on the emission spectra, being 1,2,3 and 4: Apoazurin Pfl, ribonuclease T_1, staphylococcal nuclease, and glucagon, respectively. Excitation anisotropy spectra of tryptophan in propylene glycol at –50°C. Also shown are the anisotropy-resolved spectra of the ^1L_a (dotted) and ^1L_b (dashed) transitions. Content reproduced from Principles of Fluorescence [105] with permission from Springer Nature.

Static and dynamic quenching

Fluorescence quenching is characterized by a decrease in fluorescence quantum yield and/or lifetime as a result of interaction of fluorophores with a quencher (*Q)*

[h] The excitation maxima of the ^1L_a is around ~280 nm while, ~290 nm for the ^1L_b state.

present in the surroundings. The process of quenching can occur via two mechanisms: dynamic and static quenching.[105]

Dynamic (or collisional) quenching is a type of non-radiative transition during the lifetime. It occurs via the collision of the quencher in solution with the fluorophore in the excited state, resulting in the deactivation of the fluorophore upon contact (return to the S_0 without emission of a photon), but no chemical modification. (Figure 1.10A-left) Molecules like oxygen, amines, halogens, and nitro groups can act as collisional quenchers via mechanisms like energy transfer, electron transfer, etc.[22] The Stern-Volmer equation (Equation 1.2) gives the ratio between fluorescence intensity in the absence (F_0) and presence (F) of the quencher and is usually used to describe the dynamic quenching process:

$$\frac{F_0}{F} = 1 + k_q \tau_0 [Q] = 1 + K_D [Q] \qquad \text{Equation 1.2}$$

where k_q is the biomolecular quenching constant τ_0 is the lifetime of the fluorophore in the absence of quencher, and [Q] is the concentration of quencher. The Stern-Volmer or F_0/F plot (or τ_0/τ for dynamic quenching) is the most common way to present quenching data. (Figure 1.10A-bottom plots)

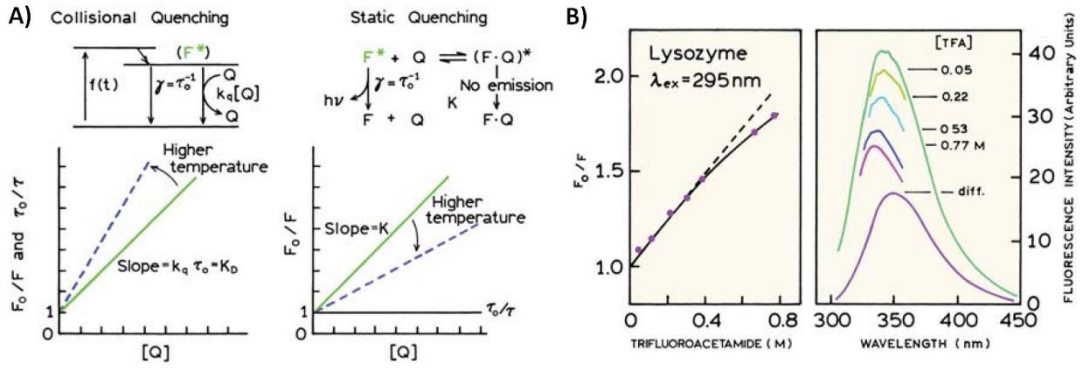


Figure 1.10-(A) Representation of the differences between dynamic and static quenching, Steiner-Volmer plots and temperature effect in each case. In dynamic quenching the change in fluorescence lifetime results from the depopulation of excited state via collision with the quencher, which will occur at a faster rate at higher temperatures. Static quenching is a result of the inhibition of the excited state formation. In the plot, F* refers to the excited fluorophore and γ is the decay rate in the absence of quencher. (B) Quenching of lysozyme by trifluoroacetamide (TFA). Left: Stern-Volmer plot. Right: Emission spectra with increasing concentrations of TFA. Also shown is the difference spectrum (diff), 0.0–0.77 M TFA. Reproduced from Principles of Fluorescence[105] with permission from publisher.

Static quenching, on the other hand, occurs in the ground state when fluorophores form nonfluorescent complexes with quenchers, inhibiting the formation of the excited state.[22] This does not involve molecular collisions.[104-106]

$$\frac{F_0}{F} = 1 + K_S[Q], \text{ where } K_S = \frac{[F-Q]}{[F][Q]} \qquad \text{Equation 1.3}$$

As per Equation 1.3, static quenching does not depend on fluorescence lifetime (as opposed to dynamic quenching), thus a τ_0/τ versus [Q] plot can be used to differentiate the two mechanisms of quenching as $\tau_0/\tau=1$ for static quenching and $\tau_0/\tau= F_0/F$ for dynamic quenching.[106] (Figure 1.10A)

Fluorescence quenching has been applied in a number of studies to provide information about fluorophores location within a molecule, structural changes, or binding.[111, 112] Figure 1.10B illustrates the effect of quenching in fluorescence emission and its use for resolving accessible and inaccessible Trp residues in LZ. With increasing concentrations of the quencher trifluoroacetamide (TFA), the fluorescence intensity decreases and shifts towards shorter wavelengths (Figure 1.10B- right plot) as a consequence of collisional quenching of, firstly, more externally located Trp residues. F_0/F and [Q] are usually linearly dependant for systems with a single population of fluorophores, but because of the presence of different populations of fluorophores in LZ, the Stern-Volmer plot presents a characteristic downward curve. (Figure 1.10B- left plot, full line in comparison to dotted line).[105, 113]

Förster Resonance Energy Transfer (FRET)

Most proteins have multiple fluorophores, many in close proximity (within 10 Å) which produces extensive FRET and affects the emission parameters. FRET is the non-radiative transfer of energy from an excited donor molecule (either fluorescent or non-fluorescent) to an acceptor which can occur if there is spectral overlap between the emission of the donor with the excitation of the acceptor.(Figure 1.11) There are two types of FRET: homo -FRET which occurs if donor and acceptor are chemically identical and hetero-FRET if they are different.[104, 105, 114]

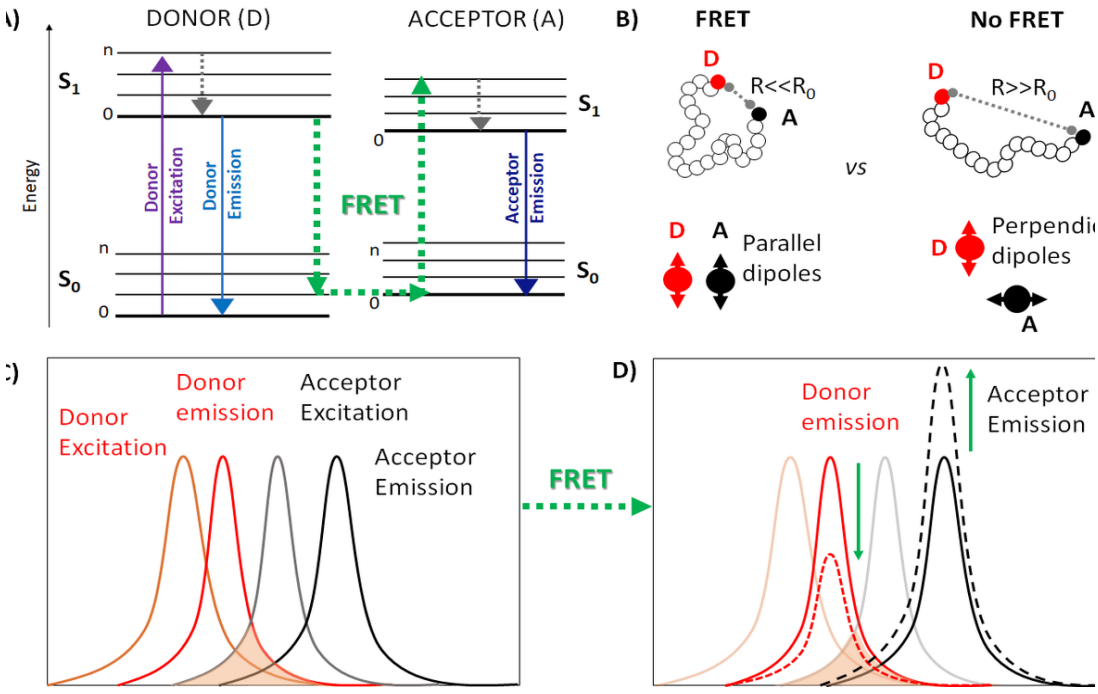


Figure 1.11- Schematic representation of FRET. (A) Jablonski diagram showing the donor (red) excitation and emission, FRET between the D-A pair and acceptor (black) emission. The FRET dependence on (B) distance between D-A pair and dipoles orientation and (C) on the spectral overlap (spectral region highlighted in orange) between the emission spectra of the donor and excitation spectra of the acceptor. (D) Effect of FRET (here, heterotransfer) on the emission spectra of donor and acceptor.

The efficiency of energy transfer (FRET efficiency, E), described by Equation 1.4, depends on the extent of the spectral overlap, the physical distance between the two involved molecules (R), and the alignment of their dipoles. R_0 is the Forster distance, which represents the distance at which the FRET efficiency (E) is 50%, and is defined according to Equation 1.5.[105]

$$E = \frac{1}{1+\left(\frac{R}{R_0}\right)^6} \qquad \text{Equation 1.4 [105]}$$

$$R_0 = 2.8 \times 10^{17} . \kappa^2 . \Phi_D . \varepsilon_A . J(\lambda) \qquad \text{Equation 1.5[114]}$$

Where κ^2 is the angle between donor and acceptor dipoles, Φ_D is the donor quantum yield in the absence of the acceptor, ε_A is the molar absorptivity of the acceptor and J(λ) is the area of spectral overlap between the two spectra. Förster distance varies from the donor-acceptor (D-A) pair but the values usually range from 20-60 Å, [105] and if R_0 is known for a specific D-A pair, one can calculate their separation (R) within a molecule.[104, 105] As a consequence, FRET is very sensitive to structural changes which makes it a potentially diagnostic tool for assessing protein structure, stability and binding.[107, 115]

Inner Filter Effect (IFE)

Fluorescence emission shows a linear correlation with sample concentration at low optical density (OD) conditions (Absorbance<0.1). However, if IFE is present, an apparent decrease in quantum yield and/or distortion in spectral shape are observed, leading to deviations from this linear dependence.[22, 106] (Figure 1.12B)

Most fluorescence spectrometers use right-angle geometry, which means that the light is incident on one face and the emission light is collected at 90 degrees to this axis.[i] The main drawback of this set-up is the light attenuation that can occur as light travels through the cuvette (and solution), as shown in Figure 1.12A. When using highly absorbing solutions, a significant part of the incident light is absorbed by chromophores before reaching the central part of the cuvette, which is called the primary IFE. The secondary IFE is characterized by the attenuation of emitted light. Thus, in general terms, IFE can be defined as the re-absorption of light by any molecule in the sample that can absorb light, which is an effect dependant on the pathlength and the number of molecules in solution.[105]

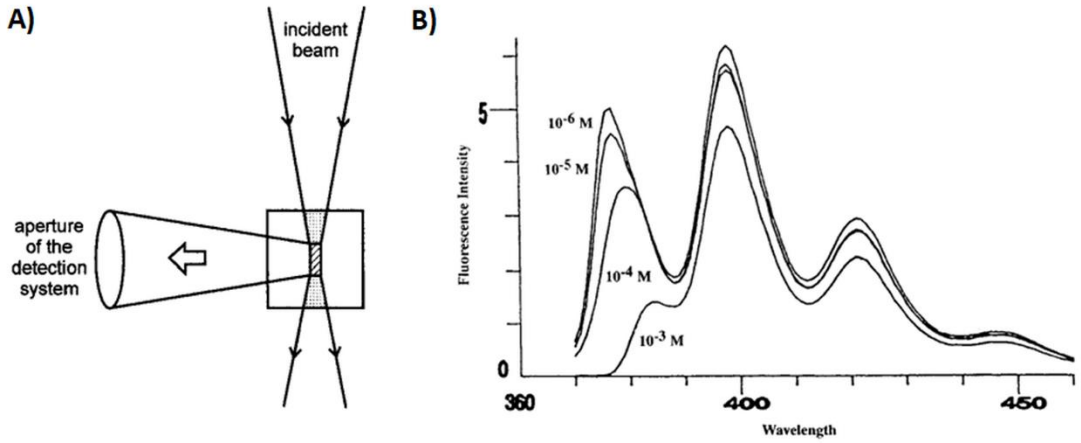


Figure 1.12- (A) Geometric arrangement of a conventional right-angle spectrometer. Figure reproduced from [106]with permission from John Wiley and Sons. (B) Example of effect of IFE in the ability of the inner filter correction procedure to correct for non-linearities between the fluorescence and absorbance of an analyte. Content reproduced from[116] with permission from Taylor & Francis.

There are four common strategies for correcting the fluorescence response for IFE. First, one can dilute the sample to a concentration with negligible IFE, but this option might not be always feasible because sample handling can change the sample,

[i] For 1×1cm cuvette, for example, this means that the incident light travels 0.5cm to the centre of the cuvette, where the collection lens is focused, and then 0.5 cm in the emission pathway, making a total of 1cm.

introduce errors and decrease signal-to-noise ratio. The second method is by changing cuvette pathlength, which is also relatively simple to implement as there are many types of cuvettes available commercially, but this might not solve the problem for very high concentration solutions. The mathematical correction method is one of most widely used procedures. The absorbance-based approach for instance, takes into consideration the measured absorbance at each $\lambda_{ex}/\lambda_{em}$ pair.[105, 106, 117] (Equation 1.6)

$$F_{corrected} = F_{observed} \times 10^{(0.5\times(A_{\lambda Ex}+A_{\lambda Em})} \quad \text{Equation 1.6}$$

For samples with OD>1.5, the mathematical correction method is not an option, and in some cases, changing sample concentration is not a practical solution either (e.g. industrial applications). In this case, IFE can be considered as part of the sample fingerprint to provide information about the sample as reported by some authors. One can also change the geometry in the instrument (e.g. use front surface excitation, FSE) but this is a much less reproducible approach and often leads to errors.[117]

1.6.4. Multidimensional Fluorescence measurements (MDF)

Fluorescence measurements are conventionally acquired using a single excitation wavelength and measuring the emission at a single or multiple wavelength, generating a 2D spectrum (λ *vs* I_F). However, to obtain more information about the sample system, one can use multidimensional measurements like Excitation Emission Matrix (EEM) or Total Synchronous Fluorescence Spectroscopy (TSFS). These 3D measurements should be better for accurately characterising multi-fluorophore systems like proteins because the topography of the spectral map will change in response to changes in many factors such as molecular composition, FRET, quenching, fluorophores local environment, concentration and IFE.

The major difference between EEM and TSFS lies in the acquisition mode. EEM spectra are obtained by collecting emission spectra for various excitation wavelengths, thus the multidimensional matrix can be considered to be a combination of excitation/ emission spectra.(Figure 1.13A) A TSFS spectra in contrast is collected by simultaneously scanning excitation and emission while increasing the offset, which corresponds to the interval between emission and excitation. This acquisition mode makes it possible to avoid the 1st order Rayleigh scatter (RS) region, which is present

in EEM spectra.[118] (Figure 1.13A). Because the objective here is to extract the most information out of the fluorescence signal, the EEM spectra was the choice as it provides in one goal two pieces of information that can be used for assessing the sample: the RS band and the pure fluorescence signal.

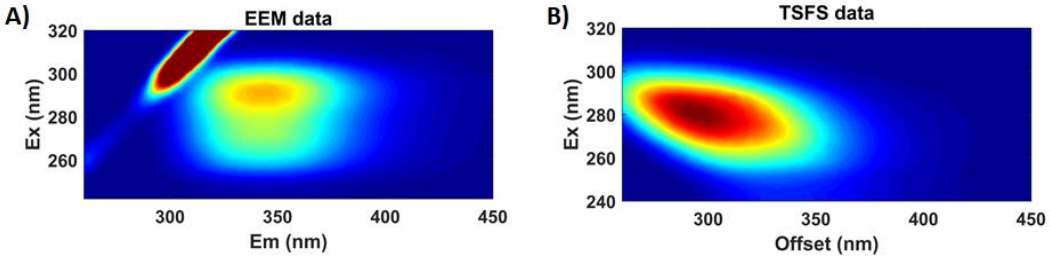


Figure 1.13- Differences between EEM (A) and TSFS (B) data. EEM plots have the fluorescence signal and the diagonal band of the first order RS, whereas only the fluorescence signal appears in the TSFS plot. Both spectra correspond to rIgG ~1 g/L, but TSFS data was acquired by a colleague (Marina Steiner-Browne) in a different instrument. The two spectra however are shown here only to highlight the differences between the two types of MDF.

1.6.5. Polarized Excitation Emission matrix (pEEM) and *Aniso*-EEM

MDF measurements are a tool for analyzing protein-based samples, but its combination with fluorescence polarization and anisotropy to generate pEEM and a*niso*-EEM maps, provides an extra level of information to assess molecular size, local viscosity and fluorophores mobility.[119-121]

Fluorescence polarization and anisotropy

The principle of polarization or anisotropy is related to photo selection and the fact that fluorophores preferentially absorb photons of light with electric vectors with close orientation to their transition moment[j] (the closer the difference in orientation is to zero, the higher the probability of absorption). In other words, in an isotropic solution, molecules are found in all orientations, but upon excitation using polarized light, fluorophores with their dipoles oriented in parallel to the incident light will be preferentially excited. [105, 106, 122] However, when occupying the excited state, the transition moment of fluorophores can change, causing depolarization of emission (partial or total) and consequent decrease in anisotropy. (Figure 1.14) Rotational

[j] The transition moment of a fluorophore has a specific orientation with respect to the molecular axis.

diffusion is one factor that can cause this change in orientation, and is associated with molecular mobility and size, viscosity of the medium, among others.[105, 106, 123]

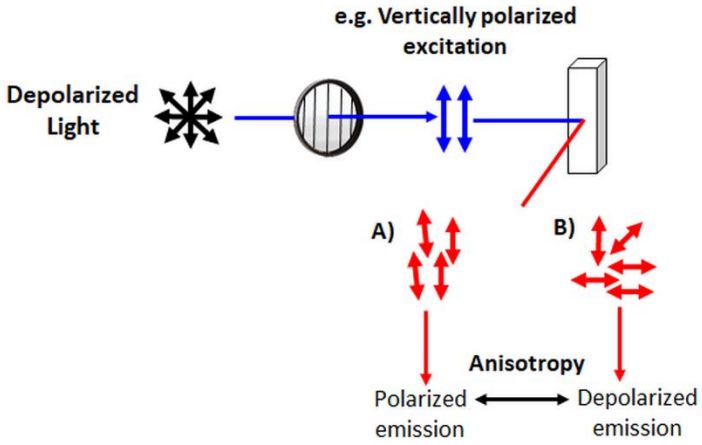


Figure 1.14- Representation showing the photo selection after polarization of incident using polarizers and the effect of depolarization after excitation.

The fluorescence anisotropy (r) and polarization (P) can be calculated as per Equation 1.7 and Equation 1.8, respectively.

$$r = \frac{I_{||} - I_{\perp}}{I_{||} - 2I_{\perp}} \qquad \text{Equation 1.7}$$

$$P = \frac{I_{||} - I_{\perp}}{I_{||} - I_{\perp}} \qquad \text{Equation 1.8}$$

where $I_{||}$ is the intensity of parallel emission and $I_{\perp}$ the intensity of perpendicular emission when sample is excited with parallel light. Both r and P measure the degree of depolarization after excitation and can be interconverted as described by Equation 1.9 and Equation 1.10 respectively.

$$P = \frac{3r}{2+r} \qquad \text{Equation 1.9}$$

$$r = \frac{2P}{3-P} \qquad \text{Equation 1.10}$$

By introducing polarizers in the excitation and emission pathways in a standard spectrometer one can collect four sets of polarized EEM data with different configurations: EEM_{VV} (vertical/vertical), EEM_{VH} (vertical/horizontal), EEM_{HH}

(horizontal/horizontal), HV (horizontal/vertical).[k] Because of the different sensitivity of the instrument to parallel and perpendicular polarized light a correction factor (G-factor) has to be calculated using EEM_{HH} and EEM_{HV}.

$$Aniso - \mathrm{EEM} = \left(\frac{EEM_{VV}\text{-}G \times EEM_{VH}}{EEM_{VV}+2 \times G \times EEM_{VH}} \right) \qquad \text{Equation 1.11}$$

$$G = \frac{EEM_{HV}}{EEM_{HH}} \qquad \text{Equation 1.12}$$

Typical *r* values range from -0.2 to 0.4, depending on the angles between excitation and emission dipole (e.g. *r*=-0.2 if angle is 90° and *r*=0.4 for 0°) but most fluorophores have an *r* lower than 0.4 because of depolarization.[105, 124] As previously mentioned, rotational diffusion is one cause of depolarization of emission. For instance, bigger proteins or proteins in a restricted environment rotate slower in solution. This translates to less depolarization and higher anisotropy values compared to smaller proteins (or molecules in unrestricted environment), which will tend to rotate faster in solution and show lower *r* values.[105] It is important to note that the total anisotropy of the sample will be a response to various factors affecting all fluorophores in solution. Apart from the rotation of fluorophores in solution, transfer of excitation energy to a molecule with different orientation can also cause depolarization of emission,[106] and IFE and contamination of fluorescence signal by scatter or polarizers misalignment can also affect anisotropy.[123]

Anisotropy resolved multidimensional emission spectroscopy, ARMES, is a methodology largely used in the group,[119-121] which refers to the combination of MDF and chemometrics to resolve individual components, of which the anisotropy can be calculated. However, this study is more focused towards industrial applications and chemometrics will only be used here for assessing variance in pEEM and *aniso*-EEM data and correlation with known process or quality parameters.

[k] The subscripts indicate the direction of the polarizers in the excitation and emission pathways respectively.

1.7. Chemometric data analysis

In order to facilitate the analysis of the complex multidimensional data, one can make use of chemometrics or multivariate data analysis, which are simply a combination of statistics and mathematics to better extract the useful information contained in the spectra. Chemometrics has been an important tool for the implementation of PAT.[125] There are a number of chemometric tools that can be used for different purposes. For instance, while standard tools like Principal components analysis (PCA) and Partial Least Squares (PLS) regression are mostly used for exploratory analysis and correlation (i.e. quantification), curve resolution methods like Parallel Factor Analysis (PARAFAC) and Multivariate Curve Resolution (MCR) are mostly applied for signal decomposition into individual contributors.

The use of PARAFAC and MCR for the analysis of EEM data is however, beyond the scope of this work as it has been studied by others in the group. [119-121, 126] In addition, we understand that the use of less complex and more widely used tools like PCA facilitates the adoption of these new analytical methods being developed here as a routine tool within an industrial context. Here we describe the basics of PCA, PLS, and two classification algorithms which were used in this thesis.

1.7.1. PCA and ROBust PCA (ROBPCA)

The principle of PCA lies on the reduction of data dimensionality while retaining the variation present in the data (as much as possible). This is done by projecting the original variables into smaller groups of new uncorrelated variables or new coordinates which are called the principal components (PCs). In PCA the aim is to find a meaningful way to approximate a large data matrix through patterns by focusing on the differences between the data points, in a way that a few PCs explain most of the variability of the dataset (dimensionality reduction, Figure 1.15B). Thus, PCA can be considered a bilinear model that emphasizes the variance of the principal components, which are orthogonal and linear combinations of the original ones.[127, 128]

In PCA a matrix is decomposed as a sum of products, which consist of a pair of loading + score vector and residuals. The scores provide information about the relationship between the various samples while the loadings describe how the variables

contribute to the variance in the original data set. Thus, scores and loadings plots are the visual results used for the interpretation of a PCA model.[127-129] (Figure 1.15B)

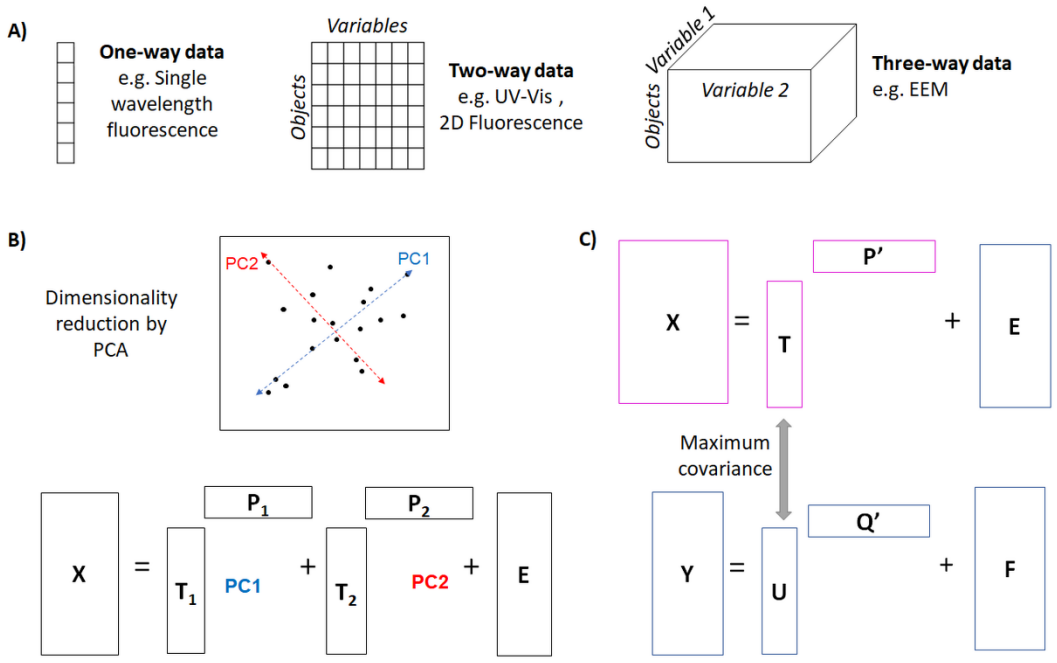


Figure 1.15- Representation of (A) one, two- and three-way arrays; (B) dimensionality reduction and decomposition of two-way arrays into vector products by (B) PCA and (C) PLS-R.

A variation of the classical approach, called ROBust PCA (ROBPCA), is a multivariate chemometric technique that differs to the classical PCA approach by the minimization of outliers effect on the data due to the implementation of projection pursuit methods for the dimension reduction and the MCD (minimum covariance determinant) estimator.[130]

1.7.2. Partial Least Square-Regression (PLS-R)

PLS-R is a chemometric tool used for building predictive models based on the linear correlation between two datasets. As opposed to unsupervised methods like PCA models, this a type of supervised method which means that prior knowledge about the sample is required (e.g. process or quality parameters).

For implementation, two datasets are provided for calibration: the first being the measurement data X (here, the fluorescence spectra) and the second dataset the known parameter Y. These will be used to build a model for prediction of an unknow parameter (y) from the measurement data x in a next step called validation. For doing

so, PLS find factors (referred to as latent variables) that compromise between the fitting of X data and prediction of Y capturing the greatest amount of variance of the X variables. In simple words, PLS decomposes both the X and Y matrices into scores and loadings structures, and the X structures are iteratively changed until the solution that leads to the maximum covariance between the scores is achieved.[128, 131] (Figure 1.15C)

1.7.3. SIMCA and SVM Classification

Classification algorithms are often referred to as pattern recognition or cluster analysis methods because they make use of pre-determined classes to allocate unknown samples into groups, being this a classical type of supervised method. [128] Some common examples are Soft Independent Modelling of Class Analogies (SIMCA) and Support Vector Machine (SVM).

SIMCA is a widespread supervised classification method which fundamental is the use of a training dataset to calculate a PCA model for each predefined group to statistically predict the class of the test samples.[132] For doing so, the resulting PCA models are used to calculate a distance to the model which takes into account the distance of a sample to the centre of the scores space (scores distance) and the residuals (orthogonal distance), which are used to define boundaries (according to a set threshold, usually 95%) for each class and assign the test samples into the different "groups". (Figure 1.16A) In the case of SIMCA, the sample can be classified as belonging to either of the classes, none of the classes, or both of the classes. [133-135]

SVM is another largely used classification method that define optimal boundaries to separate two classes. Its combination with Kernell functions (e.g. radial basis function, RBF) allows the classification of non-linearly[l] separable classes. In linear SVM classification, the extreme points from each class are used to define the optimal decision boundaries to separate the classes. These extreme points are called the support vectors. When classes are not linearly separable (e.g. Figure 1.16B), the definition of the boundaries can be achieved using Kernell functions, which basically

[l] Linear classifiers are used when the classes are linearly separable (a line can be drawn to separate samples from the different classes), while non-linear classifiers are used if the separation is not possible only by drawing a line.

transforms non-linearly into linearly separable data points by remapping the original space into a higher dimension where the separation is supposed to be easier.[132, 136, 137]

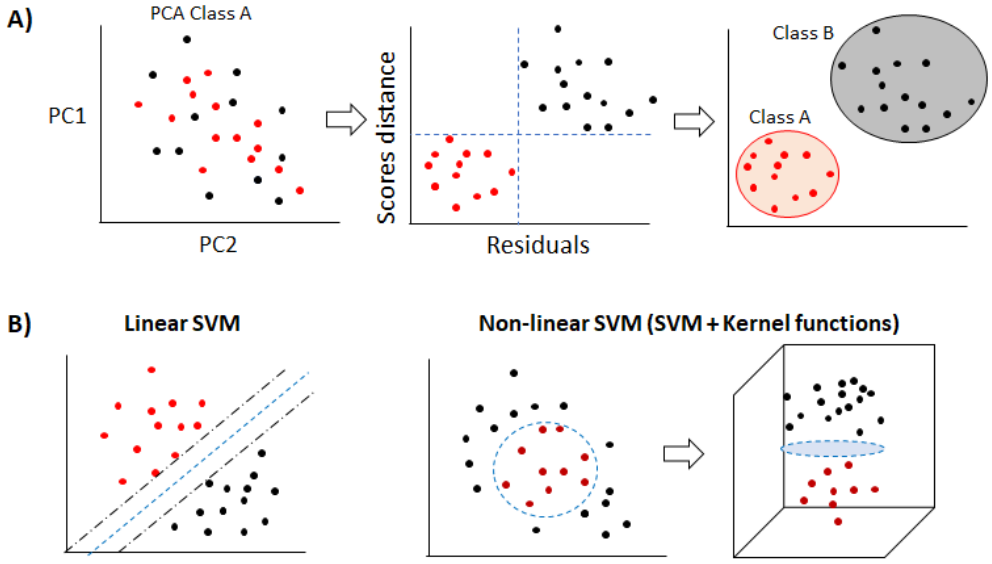


Figure 1.16- Schematic representation of classification using (A) SIMCA and (B, left) linear and (B, right) non-linear SVM.

The classification results are usually given as a confusion table which shows the number of samples predicted as each class *vs* actual class and a confusion matrix which summarizes the main results for each class in terms of, for instance, true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), false negative rate (FNR) and error.

Table 1.5 summarizes some studies where PCA, PLS, SIMCA and SVM were successfully applied to EEM data.

Table 1.6- Examples of use of PCA, PLS, SVM, and SIMCA in combination with EEM for different applications.

Method	Application	Reference
PCA	Monitoring antigen manufacturing process	Zavatti *et al.*[138]
	In-line bioprocess monitoring of cell culture processes	Claßen *et al.*[139]
	Tracking fluorescent components of dissolved organic matter	Pan *et al.*[140]
PLS-R	In-line bioprocess monitoring of cell culture processes	Claßen *et al.*[139]
	Monitoring of monoclonal antibody aggregation	Ohadi *et al.*[141]
	Monitoring bioprocess productivity	Li *et al.* [142]
SVM	Classification of breast tissue for diagnostics	Dramićanin *et al.*[143]
	Classification of Spanish vinegars according to region	Ríos-Reina *et al.*[144]
SIMCA	Classification of white wines	Goicoechea *et al.* [145]
	Classification of water samples	Hall *et al.* [146]

1.8. Thesis outline

The chemical modification of proteins is a widespread strategy for enhancing therapeutic potential and has been used in many marketed products. A typical protein modification workflow usually starts from the preparation of starting materials (proteins or peptides, linkers, small molecule surrogate), followed by one or two-step conjugation reaction, and finally a purification step, and ideally this process should be assessed at every step, in order to ensure product quality. Thus, each chapter of this thesis focus on the use of pEEM and/or *aniso*-EEM for assessing a different stage of chemical modification: the preparation of starting material, monitoring of the reaction mixture, and the characterization of final purified products.

In the first discussion chapter (Chapter 3) we use lysozyme PEGylation as a model to show the potential of pEEM to assess the variance in protein with conjugation, focusing on the characterization of final purified products with different levels of conjugation. Despite the fact that characterization of purified products is usually the last step on the modification workflow, here it was discussed at first because it represents the case of a simpler and smaller protein, before we progressed to the study of rIgG in Chapters 4 and 5.

Chapter 4 focus on assessing the quality of proteins (here rIgG) during the preparation steps prior to its final use (e.g. preparation to be used as starting material for a chemical modification or before formulation if being used for therapeutic purposes as a naked antibody). In this chapter we compare the use of pEEM with the standard technique, SEC, for assessing changes after the protein is subjected to stressful conditions like mechanical stress, improper storage and variations in sample preparation, which simulate stresses during various steps of manufacturing. In Chapter 5 we discuss an example of in-situ assessment of chemical modification, which was done using a two-step conjugation of PEG to rIgG as a model. Here, the goal was to evaluate if pEEM would be able to detect variations even when only subtle changes are expected (here, because of the low level of conjugation achieved).

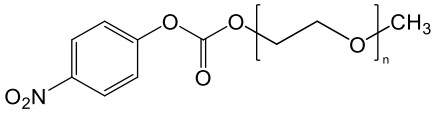
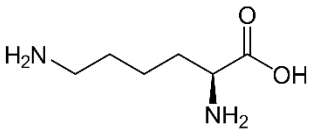
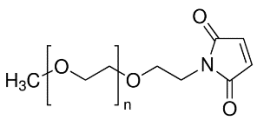
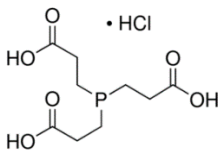
Chapter 2. Materials and Methods

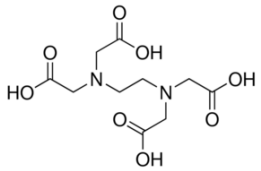
This section details the reagents, sample preparation, instrumental parameters and data analysis methods used in the study.

2.1. Reagents and Materials

All reagents were used as received, without further purification. Buffer reagents (Sodium phosphate monobasic, sodium phosphate dibasic heptahydrate, and sodium chloride) and reagents for HPLC mobile phase preparation were all purchased from Sigma Aldrich.

Table 2.1- Description of the main reagents used in the thesis.

Reagent	Details/Structure	Source
Lysozyme from chicken egg white	Lyophilized powder, protein ≥90 %, ≥40,000 units/mg protein	Sigma-Aldrich
Methoxypolyethylene glycol p-nitrophenyl carbonate (mPEG-pNPC)		Sigma-Aldrich
L-Lysine.HCl		Sigma-Aldrich
IgG from rabbit serum	≥95%, essentially salt-free, lyophilized powder	Sigma-Aldrich
L-Tryptophan	≥98%, reagent grade	Sigma-Aldrich
Methoxypolyethileneglycol maleimide (MAL-PEG 5kDa) Ω-end : maleimide / α-end: CH3	 Average 5.000 g.mol-1 CAS: 99126-64-4 ≥90% (NMR)	Sigma-Aldrich
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP.HCl) Powder	 $C_9H_{15}O_6P \cdot HCl$ MW=286.65 g.mol^{-1} CAS: 51805-45-9	Sigma-Aldrich

Ethylenediaminetetraacetic acid (EDTA)	 $(HO_2CCH_2)_2NCH_2CH_2N(CH_2CO_2H)_2$ MW= 292.24 g.mol^{-1} CAS: 60-00-4	Sigma-Aldrich
High Purity water (HPW)	Chromatography grade	Lenox / Honeywell

2.2. Sample preparation

2.2.1. Sample preparation for lysozyme PEGylation study

Phosphate buffer pH 8.0 was prepared each day using non-sterilized high purity water (HPW) and were membrane filtered using 0.2 µm polyethersulfone (PES) Captiva Premium syringe filters (Agilent Technologies) and BD Plastik^{TM} syringes before use.

Lysozyme starting material (Certificate of analysis available in Section 7.2) solutions were prepared in amber volumetric flasks 10 mL using 11 mg of the lyophilized powder and the pre-filtered buffer. The solutions were re-filtered using the same Captiva filters and transferred into 1×1cm pathlength quartz cuvettes (Lightpath Optical, UK). To generate products with different PEGylation profiles the protein was mixed with varying PEG:protein molar ratios (1, 2, 4, 6, 12, 24) and incubated with gentle stirring at 20.0 or 30°C in the fluorimeter for 2, 24, or 72 hours. mPEG-pNPC was the PEGylating agent selected because of the slower reaction rate compared to other agents which enabled us to easily vary the degree of PEGylation.[147]

The reaction was stopped by the addition of excess of L-lysine hydrochloride prepared in the reaction buffer and the final mixture was purified using Amicon ultracentrifuge filters (10 kDa MW cut-off)[a] in order to remove p-nitro phenol (pNP)

[a] For the purification, 500µL of the reaction mixture was transferred to the centrifugal filters, spun at 14,000 × g for 5 minutes and the filter was reloaded with aliquots of 450µL at a time (because of the low capacity of the centrifugal filters) until all the sample had been loaded. The retained sample was washed five times with the reaction buffer (with spinning at 14,000×g for 5 minutes) and the purified sample was collected by reverse spinning at 1,000×g for 2 minutes and diluted with the reaction buffer to an approximate concentration of 1 g/L.

groups that were released during the conjugation causing an increasing yellowing in the solution (*vide infra*). All reactions used a single source lot of lysozyme[b], PEG reagent, L-Lysine hydrochloride, and buffer reagents. UV-Vis and EEM spectra were collected from all starting materials and purified products in the same cuvette and aliquots of each were stored at -70 °C (in LoBind Eppendorf tubes) for SEC, SDS-PAGE, and DLS analysis. Some reactions were also analysed hourly using only EEM‖ measurement for a time course experiment.

Table 2.2- Summary of reaction conditions used for each reaction and concentration of SM and Purified products calculated from A_{280} corrected for scatter effect as described in Section 7.5.

PEG-pNPC molar excess	Reactions	Temperature (°C)	Reaction time (h)	Conc. g/L SM	Conc. g/L Pur, PEG-LZ
0 PEG/LZ	**R28-30**	20	24	1.00±0.00	0.93±0.00
1.0 PEG/LZ	**R13-15**	20	2	1.00±0.00	0.95±0.01
2.0 PEG/LZ	**R22-24**	20	24	1.00±0.00	0.97±0.01
4.0 PEG/LZ	**R31-33**	20	24	1.00±0.00	0.96±0.01
6.0 PEG/LZ	**R1-3**	20	24	0.99±0.00	0.99±0.01
6.0 PEG/LZ	**R4-6**	20	24	1.00±0.00	0.95±0.01
6.0 PEG/LZ	**R16-18**	20	24	0.98±0.00	0.94±0.02
12.0 PEG/LZ	**R7-9**	20	24	0.99±0.00	0.93±0.01
12.0 PEG/LZ	**R10-12**	30	24	0.99±0.00	0.93±0.06
24.0 PEG/LZ	**R19-21**	20	24	0.99±0.00	0.96±0.01
24.0 PEG/LZ	**R25-27**	20	72	0.99±0.00	0.93±0.01

For the chemical unfolding study, stock solutions of 1 g/L of lysozyme were prepared by mixing phosphate buffer and 8 M guanidine hydrochloride (in phosphate buffer) to obtain final concentrations of 0 M, 2 M, 4 M, 6 M, and 8 M Gd.HCl.

For pNP control experiments, LZ solutions were spiked with 4-Nitrophenol (Reagent Plus®, ≥99% from Sigma) to build a calibration curve with concentrations ranging from 0- 10 µM of pNP and absorbance measured at 400 nm. A test sample was prepared containing LZ and 0.5 µM of pNP. For studying the impact of free PEG on the fluorescence of LZ, 1 g/L LZ solutions were spiked with 2,4,16, and 24 molar excess of inactive ~6 kDa polyethileneglycol (from Sigma) and analysed by EEM.

[b] All the reactions were carried out using lyophilized powder from the same container, which was parafilm sealed after used and kept at -20°C. The certificate of analysis of the lysozyme lot used is available in Section 7.2.

2.2.2. Sample preparation for rIgG quality screening studies:

PBS/EDTA buffer pH=7.0 was prepared by mixing a freshly prepared PBS buffer pH 6.5 with EDTA (in 2 M NaOH) to a final concentration of 10 mM. The pH was then adjusted to 7.0 with 1 M HCl. This buffer was the choice because some of the samples were used as starting materials for the PEGylation reactions used in Chapter 5.

Unstressed samples were prepared by adding 10mL of buffer to the original sample container[c] to an approximate concentration of 1 g/L (Table 2.3), using three different rIgG source lots from Sigma: SLBP7449V (Lot 1, n=15), SLBW8687 (Lot 2, n=5), and SLBZ5214V (Lot4, n=6). The different stock solutions for each lot were prepared on different days using the same bulk lots but different source vials and were then filtered before being split into aliquots for analysis.[d]

A set of stressed rIgG samples was prepared to simulate aggregate containing solutions that could be formed during manufacturing, or by improper storage, shipping, and handling. For this, Lot2 aliquots were transferred to 5mL Lobind Eppendorf tubes, mechanically stressed using a vortex shaker under different conditions and transferred back to the cuvettes.[e] Lot3 samples were prepared from a batch of the powder which had been unintentionally stored at room temperature for one month, opened, and then stored at 2–8°C for 18 months prior to use. (Figure 2.1)

[c] The rIgG lyophilized powder was reconstituted on the original container instead of a volumetric flask because of the sensitivity of the powder to static electricity and the foaming tendency of the solution, which would make harder the recovery and transferring.

[d] The certificate of analysis (Section 7.2.2) shows different water (from 2-4%) and sodium (from 0.2-0.5%) content for the different lots, which might suggest that they were prepared using slightly different processes and possibly different.

[e] For soln.6 from Lot2, the shaking was done with 3mL of solution in 5mL tubes whereas for soln.7 the sample volume was 2 mL in a 5 mL tube, producing a bigger headspace (and thus increased the chance of aggregation at the air-water interface).

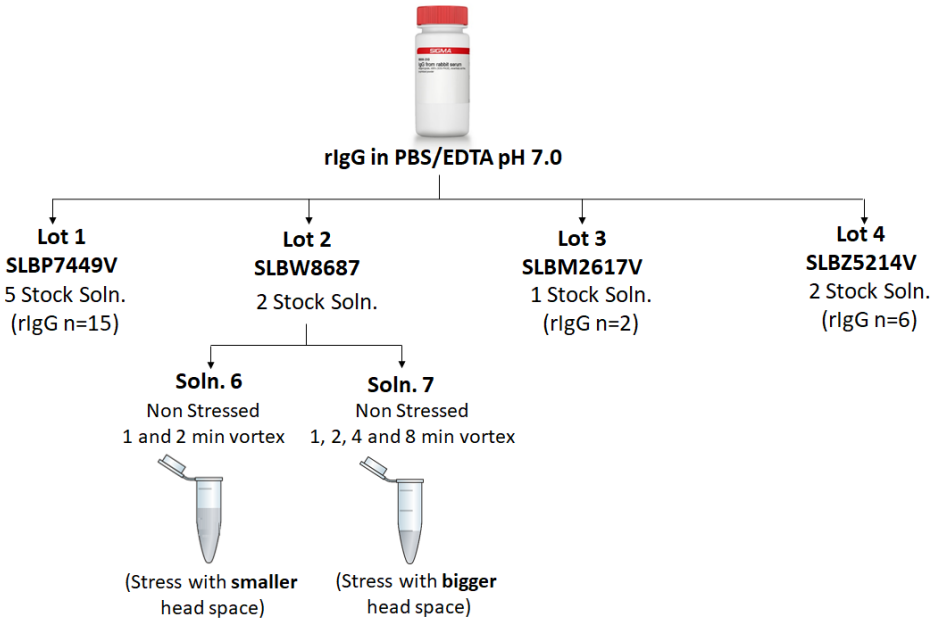


Figure 2.1- Schematic showing procedure for rIgG starting material sample preparation, different lots used, and mechanical stress applied.

Table 2.3- Calculated concentrations for each rIgG solution used in Chapter 4.

Lot	rIgG Stock Solution	$A_{apparent,280}$	$A_{protein,280}$	Protein Conc.($g.L^{-1}$)
Lot1	soln. 1 NS (n=3)	1.40	1.38	1.02
	soln. 2 NS (n=3)	1.39	1.31	0.97
	soln. NS 3 (n=3)	1.50	1.38	1.02
	soln. 4 NS(n=3)	1.33	1.29	0.96
	soln. 5 NS (n=3)	1.32	1.30	0.96
	Overall Lot 1(n=15)	1.32	1.29	0.95
Lot 2	soln. 6 NS (n=1)	1.34	1.29	0.96
	soln.6 1min vortex (n=1)	1.34	1.30	0.96
	soln.6 2 min vortex(n=1)	1.34	1.30	0.96
	soln.7 NS (n=4)	1.38	1.35	1.00
	soln.7 1min vortex (n=1)	1.38	1.34	0.99
	soln.7 2min vortex (n=1)	1.38	1.35	1.00
	soln.7 4min vortex (n=1)	1.36	1.32	0.98
	soln.7 8min vortex (n=1)	1.35	1.32	0.98
Lot 3	soln.8 (Poor storage) (n=2)	1.37	1.32	0.98
Lot4	soln.9 (n=3)	1.37±0.04	1.32±0.03	0.98±0.02
	soln.10 (n=3)	1.36	1.32	0.97

All freshly prepared starting material samples were analysed by absorbance and fluorescence spectroscopy and aliquots of each sample were stored at -70°C in Protein LoBind Tubes (Eppendorf) for SEC analysis.

2.2.3. Sample preparation for rIgG PEGylation:

Three independent sets of reactions were analysed, using three different rIgG source lots from Sigma: SLBP7449V (Lot 1), SLBW8687 (Lot 2), and SLBM2617V (Lot 3). The first set of well-controlled reactions (n=14, R1–R14) used Lot 1, unstressed rIgG, while the poorly-controlled reactions (R15–R17) were the second set which used mechanically stressed (vortex mixer) rIgG, and a third set (R18–R19) used a degraded lot of rIgG. Lot3 (SLBM2617V) had been inadvertently stored at room temperature for one month, opened, and then stored at 2-8°C for 18 months prior to use. (Table 7.2 in Section 7.3 summarizes the starting material and reaction condition used in each reaction).

The reaction buffer PBS/EDTA buffer pH=7.0 and starting materials were prepared as detailed in Section 2.2.2. rIgG-PEG conjugates were prepared multiple times by reduction of IgG disulphide bonds with 1.5 and 3.0 molar excesses of TCEP.HCl at 20 °C for 2:15 hours, followed by alkylation with 10 molar excess 5kDa MAL-PEG for 2 hours, at 20 °C. Figure 2.2 shows the reaction scheme used for PEGylation reaction. TCEP.HCl stock solutions were prepared in the reaction buffer and MAL-PEG prepared in DMSO immediately before use.

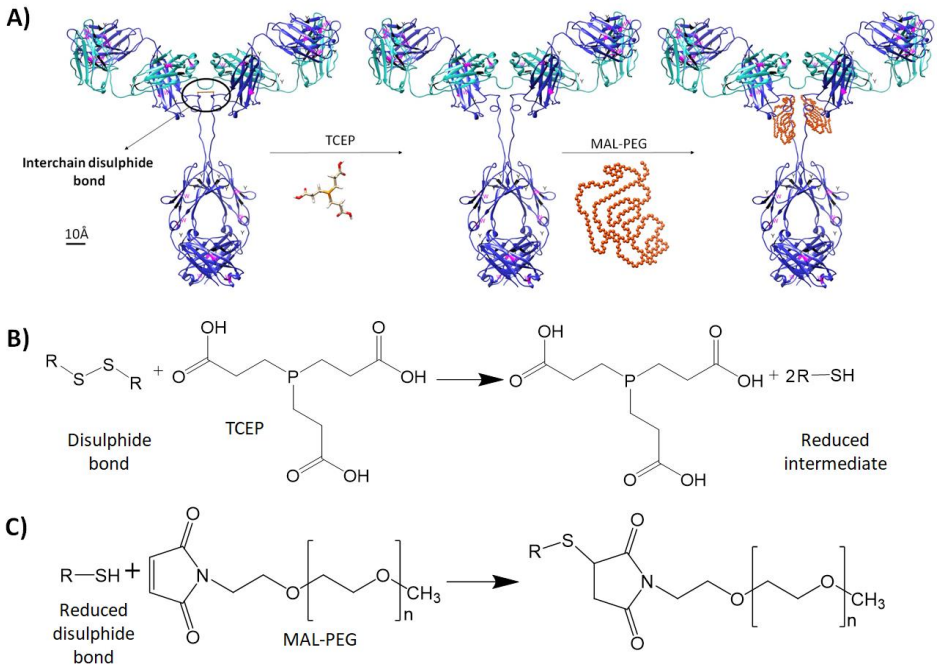


Figure 2.2- (A) IgG PEGylation reaction scheme showing IgG structure. IgG crystal structures are a partly theoretical model created from a ribbon diagram of IgG (PDB entry 1IGT) with computational models. [1, 2]. Chemical reaction of the (B) reduction of disulphide bonds by TCEP and (C) addition of PEG to the free thiol.[148]

The conjugation reaction was conducted in 1×1cm path length quartz cuvettes with gentle stirring and temperature control in the fluorimeter. The reaction in each of the three cuvettes was begun at different times in order to allow for the three reactions to be run and analysed simultaneously, as shown by Figure 2.3.

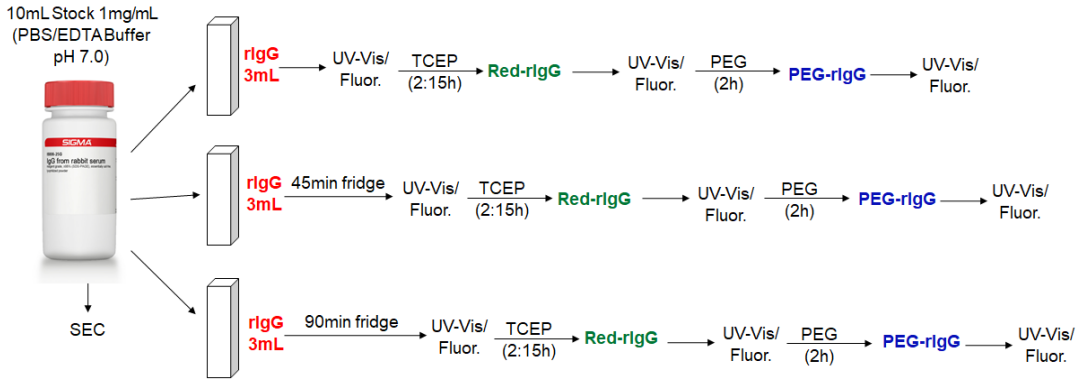


Figure 2.3- Scheme for sample preparation, starting material incubation time prior to reaction and data collection used for stock solutions prepared.

Thus, it was possible to keep the same reaction time, however, the incubation times in the cuvette (prior to TCEP addition) varied. For R1–R14, rIgG stock solutions (~1 g/L) in PBS buffer pH7.0±0.1, were prepared fresh daily, and dispensed into three different reaction cuvettes (for stock solution 3 and Lot3, only two reactions were possible). Spectral measurements were made at different points during the reaction and samples were aliquoted to Protein LoBind tubes (Eppendorf) and frozen at -70°C for further analysis.

R15 to R19 were carried out 18 months after reactions R1-14 and a correction factor had to be used to ensure that the data was comparable as instrument throughout had changed after maintenance leading to weaker intensities. The intensity correction factors for $EEM_{\parallel}$ (1.80), $EEM_{\perp}$ (1.89), and EEM_{T} (1.84) were calculated from measurements of p-terphenyl in PMMA (Agilent) a polymer block (Refer to Section 7.8.3. for details)

Quantification of free sulfhydryl groups using Ellman's reagent

The Elman's method was used for the quantification of free thiols in IgG.[149] The DTNB reagent (5,5'-dithiobis-(2-nitrobenzoic acid) and Cysteine*HCl*H_2O (used for the calibration) were prepared in PBS buffer as used for the reactions. Since TCEP can reduce the Ellman's reagent,[150] the reaction mixture had to be purified before carrying

out the test, which was done by spinning the mixture through an Amicon Ultracentrifuge filter of 10kDa cut-off. For the test, the Ellman's solution was mixed with the test samples and the solutions incubated at room temperature for 15 min. The concentration of free sulfhydryl in the sample was then estimated using absorbance at 412nm by comparison to a standard curve of known concentrations of cysteine (Section 7.8.2).

2.3. Instrumentation and data collection:

2.3.1. UV-Visible absorbance spectroscopy

Absorption spectra were recorded at 20ºC using a Cary 60 spectrometer (Agilent) over the range 200–600 nm with 2nm resolution and scan rate of 1200nm/min using the corresponding buffers as reference. Phosphate buffer pH=8.0 was the reference buffer for LZ studies, PBS/EDTA buffer pH 7.0 was used for rIgG starting materials and for Reduced-rIgG (Red-rIgG) intermediate and PEG-rIgG final products the reference was PBS/EDTA buffer pH 7.0 with TCEP and TCEP+MAL-PEG, respectively.

2.3.2. Fluorescence measurements

EEM were collected using a Cary Eclipse fluorescence spectrophotometer (Agilent Technologies) fitted with a bespoke polarizer accessory [119] and equipped with pairs of broadband wire grid polarizers (WGP) and a temperature controlled multi-cell holder from Agilent Technologies.

pEEM spectra were collected from non-degassed solutions held in 1×1cm quartz cuvettes (Lightpath Optical, UK) over an excitation range of λ_{ex} = 240–320 nm and an emission range of λ_{em}= 260–450 nm (2 nm increments in each case) with 10 nm excitation/emission slit widths. The scan rate was set at 1200 nm/min and photomultiplier voltage at 650V for experiments with lysozyme (Chapter 3) and 600V for experiments with rIgG (Chapter 4 and 5). All the samples were measured using four different polarization configurations: HH (horizontal-horizontal), HV (horizontal-vertical), VH (vertical-horizontal), and VV (vertical-vertical), unless otherwise stated.

2.3.3. Size Exclusion Chromatography (SEC)

SEC of all samples was performed in a 300×7.8 mm mAb PAC-SEC1 column from ThermoFisher with a particle size of 5µm in an Agilent 1260 HPLC system equipped with a DAD detector. 10µL of sample were injected in triplicate at 30 ºC with 50mM Sodium Phosphate pH 6.8+300 mM NaCl buffer as the mobile phase, and a flow rate of 0.76 mL/min.[f]

Mobile phase was membrane filtered (0.45 µM) and sonicated for 15 minutes prior to use. Column and system were pre-washed with water using gradient flow and then pre equilibrated using 20 column volumes of buffer prior to injection. The detector was turned on for the last two hours of column conditioning for lamp equilibration. After use the column was washed with water and stored in a solution of Acetonitrile/Water 20%. Autosampler temperature was set to 25ºC during all analysis.

2.3.4. Sodium Dodecyl Sulphate-Polyacrylamide gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed according to the BioRad TGX Precast Gels® specifications. 7.5 µL of rIgG sample were mixed with 2.5 µL of 4X Laemmli loading buffer from BioRad, heated at 90 °C for 5min, cooled down and then loaded into the precast 4–15% Mini-PROTEAN® TGX™ protein gels from BioRad.

10 µL of Pre-stained SDS-PAGE Standards (Bio-Rad) were used as the molecular marker. The gels were run (using a Mini-PROTEAN Tetra Cell from BioRad) connected to a power supply at a constant voltage of 200 V in a 1× solution of Tris/Glycine/SDS SDS running buffer at room temperature for approximately 45 min (or until the dye front reached the end the gel).[g] Gels were fixed by microwaving for 50 seconds with a solution of ethanol/acetic acid and stained with Coomassie blue solution.[151] Gels of PEGylated proteins were fixed, washed with a 0.1M perchloric acid solution for 20 minutes and stained by adding 10 mL of a 5 % barium chloride

[f] This method was obtained from ThermoFisher MAbPac SEC-1 column manual. This can be accessed in: https://assets.thermofisher.com/TFS-Assets/CMD/manuals/Man-065402-LC-MAbPac-SEC-1-Columns-Man065402-EN.pdf

[g] These are the conditions indicated in the Mini-PROTEAN® Precast Gels Instruction Manual and Application Guide from BioRad: (http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_1658100.pdf)

solution followed by 10 mL of a 5% Iodine solution.[152] The gel was destained in distilled water for up to two hours and then scanned.

2.3.5. Dynamic Light Scattering (DLS)

DLS data were collected at 20 °C, after filtration (0.20μm PES filter), using a Malvern Zetasizer Nano ZS (173°detection angle). Each sample was measured 5 times (each measurement was an average of 10 runs of 10 second duration) in plastic disposable cuvettes. The hydrodynamic radius (R_h) was computed from the diffusion coefficients using Zetasizer Software version 7.13 from Malvern Panalytical.[h]

2.4. Data analysis and Chemometric tools:

Chemometric analysis were performed using PLS_Toolbox8.2.1® working in MATLAB (ver. 9.1.0) environment and in-house written codes (FluorS).

2.4.1. Univariate data analysis

Aggregation indices:

Two simple spectroscopic measurements were evaluated here as qualitative parameters of protein aggregation. UV-AI was calculated using the absorption spectra defined as follows:

$$\text{UV-AI}\ (\%) = \left(\frac{A_{350}}{A_{280}\text{-}A_{350}} \right) \times 100 \qquad \text{Equation 2.1}$$

where A_{280} and A_{350} are the absorbance values at wavelengths of 280 and 350 nm, respectively. [153, 154]

A second measure of aggregation, the fluorescence aggregation index (Fl-AI), [155] was defined as the percentage ratio of the intensity of the Rayleigh scattered light at λ_{ex} = 280 nm (I_{280}) with the fluorescence emission intensity from the protein measured at 340 nm (I_{340}) using 280 nm excitation:

[h] The DLS instrument was only acquired after studies with IgG had been finished. Thus, DLS was only available for Chapter 3 experiments. For particle size analysis, it is important that the sample is preserved, which made it impractical to freeze samples before analysis or send to a different facility for measurements.

$$\text{Fl} - \text{AI}\ (\%) = \left(\frac{I_{280}}{I_{340}}\right) \times 100 \qquad \text{Equation 2.2}$$

Similarity Index (SimI):

The similarity index parameter is a type of correlation function used to rapidly assess similarities between EEM data. *SimI* between two EEM matrices X1 and X2 (dimensions I × J) was calculated using a penalty parameter (λ) used to set a detectable limit of variance between X1 and X2.[156]

$$\text{SimI} = 1 - \lambda \frac{\sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{x1-x2}^2}}{\sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{x1+x2}^2}} \qquad \text{Equation 2.3}$$

Here it was set to 4 which corresponds to 5% variance and was proven to be adequate. px1-x2 and px1+x2 were elements of (X1-X2) and (X1 + X2) respectively. The closer *SimI* value was to one, the more alike the two matrices X1 and X2 were.

2.4.2. EEM data pre-processing

The raw pEEM spectra have to be pre-processed to make the data more suitable for analysis. The first step of the pre-processing is the blank subtraction which was done in the four spectra (HH, HV, VV, VH) using the correspondent buffers for each sample.

Instrumental correction factor

For PEG-LZ studies, $EEM_{\parallel}$, $EEM_{\perp}$ and EEM_T spectra were corrected for instrumental response prior to G-factor correction using a factor calculated using a Spectral Fluorescence Standard Kit (Sigma, product No. 69336), certified by BAM (Federal Institute for Standards and Materials, Germany).[157] For this, the blank subtracted data was selected to the range $\lambda_{em=}$302-450 [i] and the correction factor was applied.

[i] This was necessary because the standards provided in the kit only covered the emission range 300-700nm.

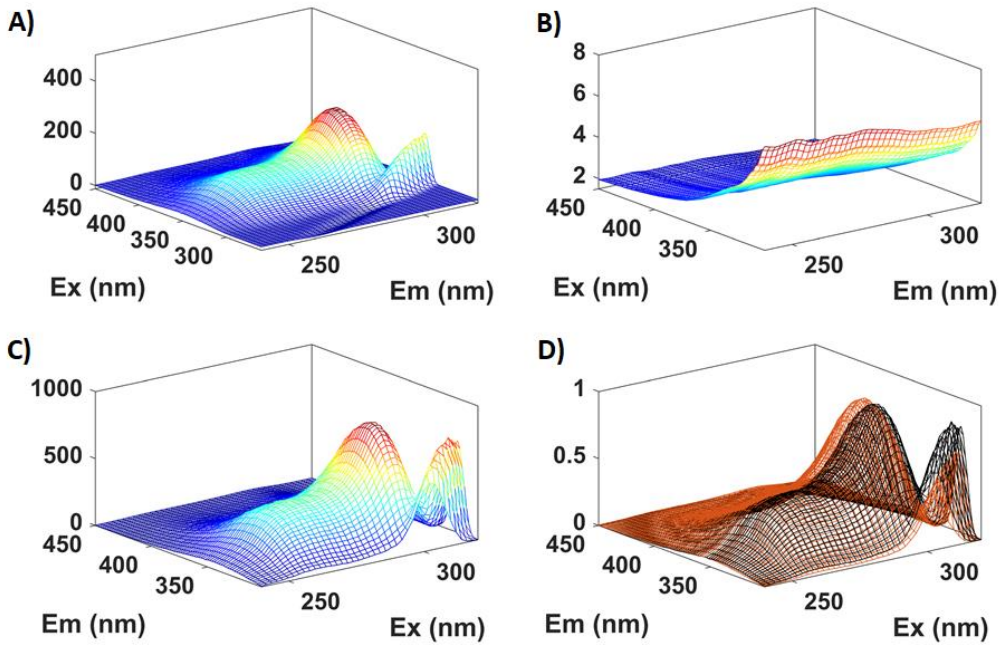


Figure 2.4- (A) Blank subtracted EEM∥ spectra of lysozyme, (B) Correction factor calculated using Spectral fluorescence Standard Kit, (C) instrumental factor corrected EEM∥ and (D) normalized EEM∥ before (orange) and after (black) instrumental correction showing changes in the spectral shape with correction.

The instrumental correction factor was only used for the LZ study because the other studies had been completed when the instrumental correction was implemented in the laboratory. For the other studies, the step after blank subtraction was G-factor correction (when necessary). (Figure 2.4)

G-factor correction

EEM_{HH} and EEM_{HV} data were used to calculate the G factor (Equation 2.4), which is a correction for the different instrument sensitivity for vertical and horizontal polarized light. EEM_{VH} data were G factor corrected [158] as per Equation 2.5 and the corrected spectra will be de referred to here as perpendicular polarized ($EEM_{\perp}$), whereas the EEM_{VV} will be called the parallel polarized ($EEM_{\parallel}$).

$$G = \frac{EEM_{HV}}{EEM_{HH}} \qquad \text{Equation 2.4}$$

$$EEM_{\perp} = EEM_{VH} \times G \qquad \text{Equation 2.5}$$

Unpolarized EEM (EEM_T) were not collected but generated from the $EEM_{\parallel}$ and $EEM_{\perp}$ measurements using the following equation: [105]

$$EEM_T = EEM_{\parallel} + 2 \text{ x } EEM_{\perp} \qquad \text{Equation 2.6}$$

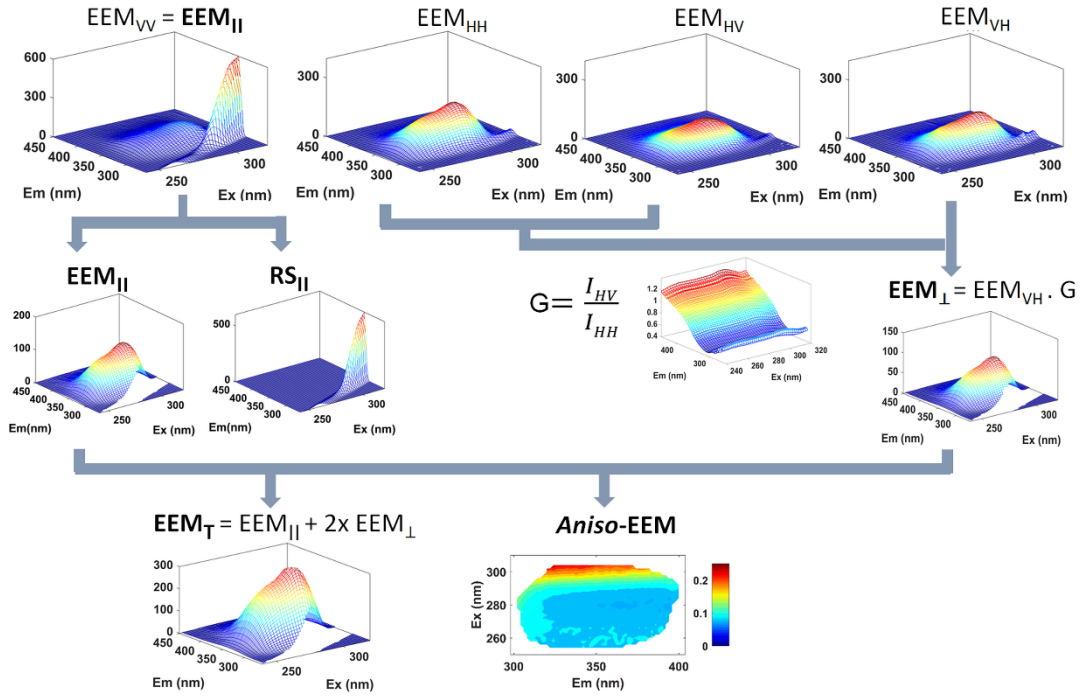


Figure 2.5- Scheme showing the four raw data spectra collected and used to generate the four outcomes of the method: EEM$_{\parallel}$ EEM$_{\perp}$, EEM$_T$ and *Aniso*-EEM.

Scatter Removal

The next step of pre-processing is scatter removal. The RS and Raman scatter can be a source of bias in chemometric data analysis because of their non-linear behaviour and the fact that scatter is generally unrelated to absorbance or emission properties. Various methods are described in the literature for removing Rayleigh and Raman scatter.[159, 160] Here, Raman scatter was minimized by using a buffer blank subtraction as previously mentioned[158] while the RS of each sample was estimated as a separate bilinear component and modelled by PARAFAC, resulting in a structure that was then reshaped, subtracted from the original matrices and its area replaced with missing values.[119, 161] This method allows the complete subtraction of the RS, which is key for chemometric data analysis, despite removing part of the fluorescence information.

While some sources infer that the scatter provides no meaningful information about the sample, this is not always true, as the RS can be related to the physicochemical properties of the sample, of which the most important is the presence

of particles. [162] Here, the reshaped RS[j] component (Figure 2.6B) was evaluated separately to calculate univariate scattering values, for either the whole scatter volume or at several discrete excitation and emission wavelengths and used as a qualitative measurement of HMWS formation.

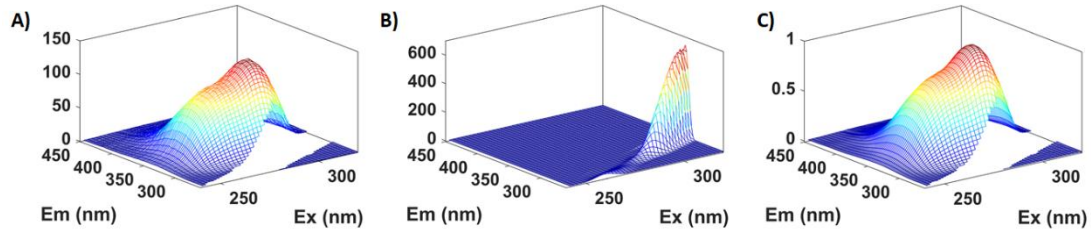


Figure 2.6-(A) Blank subtracted/missing data corrected, (B) extracted RS and (C) normalized and smoothed EEM‖ data of a rIgG sample (here taken as an example).

After RS removal, the matrices were normalized to the maximum peak, which was done to underline the separation based on spectral shape changes rather than intensity variations, which could be caused by intrinsic measurement errors and/or sample preparation variation. Absorbance[k] and pEEM spectra were smoothed using a Savitzky–Golay filter (with a second-order polynomial with a 15-point window size) and then normalized to maximum value.[163]

2.4.3. Aniso-EEM plots:

Aniso-EEM maps were constructed from anisotropy values calculated at each $\lambda_{ex/em}$ combination. Fluorescence anisotropy (r) was calculated using I_{VV}, $I_{VH,}$ and the G factor (from I_{HH}, I_{HV}), as described in Equation 1.11 and 1.12. (Figure 2.5)

2.4.4. Chemometric data analysis

Multivariate analysis was performed on the normalized or non-normalized data after unfolding and pre-processing by mean centering or auto scaling. EEM data is a 3-way array, thus, in order to be analysed using ROBPCA (a bilinear model), the dataset has to be unfolded by multiplying the two variables corresponding to the excitation and emission. (Figure 2.7B) Mean center is defined as the subtraction of

[j] The RS was extracted from the blank subtracted data.
[k] Absorbance data were selected to 250-600nm prior to any pre-processing to minimize the amount of noise being modelled.

the mean spectra of the data set from every individual spectra. It is a typical pre-processing method used prior to ROBPCA analysis to highlight the differences between samples ensuring that the main component is not related to the average signal. [128] The use of mean centering pre-processing in the context of ROBPCA calculations is referred to as robust centering. (Figure 2.7C) Auto-scaling is simply applying mean-centering on the data followed by division of each variable by its respective standard deviation. (Figure 2.7D)

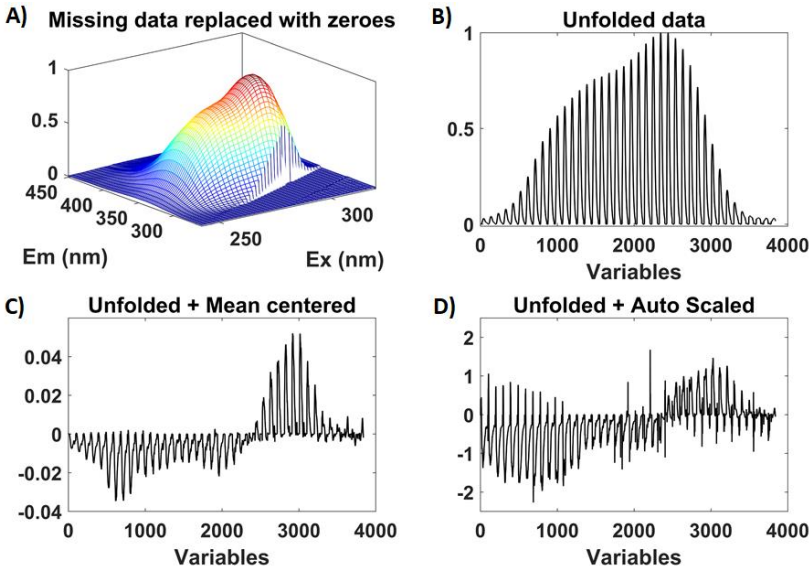


Figure 2.7- Steps of pEEM data processing for multivariate analysis: The missing data area is first replaced with zeroes (A), (B) unfolded and finally mean centered (C) or auto scaled (D) depending on the application.

Exploratory data analysis

Exploratory data analysis was carried out using ROBPCA, which is a more suitable method for screening applications. ROBPCA loadings plot were used to identify the significant spectral features which differentiated samples in the Scores plots. The Venetian blind method (4 splits) was used for ROBPCA cross-validation and Root Mean Square Error (RMSE) was used to select the optimum number of PCs.

Sample classification:

SIMCA and SVM were used to compare the ability of UV-Visible and EEM measurements to discriminate reaction state based on significant spectral changes. Statistical classification methods apply mathematical models to identify the class that an unknown sample belongs to, based on spectral features of a classification dataset.

Sensitivity, specificity, and misclassification error (Err) statistical parameters were used to assess the efficacy of each algorithm for sample discrimination.

$$\text{Sensitivity (TPR)}(\%) = \left[\frac{\text{TP}}{(\text{TP+FN})}\right] \cdot 100 \qquad \text{Equation 2.7}$$

TNR was the proportion of negatives that were correctly classified:

$$\text{Specificity (TNR)}(\%) = \left[\frac{\text{TP}}{(\text{TN+FP})}\right] \cdot 100 \qquad \text{Equation 2.8}$$

Err was the proportion of samples which were incorrectly classified:

$$\text{Err}(\%) = \left[\frac{(\text{FP+FN})}{(\text{TP+TN+FP+FN})}\right] \cdot 100 \qquad \text{Equation 2.9}$$

Regression and quantification:

Quantitative modelling for quantification of concentration, conjugation levels and % of aggregates was implemented using unfolded PLS (u-PLS).[130] The model complexity (number of latent variables) was determined using standard cross-validation methods (Venetian blind) and the F-ratio criterion, which was also used to detect outliers in the calibration dataset. [164] u-PLS model performance was assessed by the Coefficient of determinations (R^2), the RMSE (Equation 2.10), and the relative error of prediction (REP, Equation 2.11), and the elliptical joint confidence region (EJCR) test were used to compare the accuracy and precision of different models at a 95% confidence interval. [165]

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \qquad \text{Equation 2.10}$$

where N is the number of samples, $\hat{y}_i$ was the predicted value of the i^{th} prediction object, and y was the i^{th} measured value.

$$\%REP = \frac{RMSEP}{Mean\ PRED}\ x\ 100\% \qquad \text{Equation 2.11}$$

Different pre-processing methods (normalization and mean-centering/auto scaling), as well as variable selection algorithms such as variable importance in projection, VIP, [166] and interval partial least squares, iPLS, [167] were evaluated with the

objective of improving model accuracy by reducing the influence of noise and non-informative data. While VIP selects spectral regions that contributes the most to the models, iPLS compares the results with and without each variable, and then selects those that return lower cross validation errors.

Chapter 3. Characterization of lysozyme PEGylation products

3.1. Introduction

In this first chapter we use PEGylation of lysosome as a model system to compare the efficacy of pEEM spectroscopy with conventional techniques and standard spectroscopic methods as a rapid tool for characterizing changes in a protein with chemical modification. This work was published in *Biotechnology & Bioengineering* in July 2020.[168]

Lysozyme is a well characterized enzyme[6] and has been widely employed in research for studying protein structure, stability, function and chemical modification, being thus, an ideal model system for studying PEGylation. Its small size makes it particularly useful for this study, which could facilitate the characterization of the conjugates as the attachment of a 5kDa PEG would likely cause measurable changes in hydrodynamic radius by the available orthogonal techniques.

Characterization of PEGylation reactions and products can be challenging because of the relatively small impact on protein structure, the lack of an accessible chromophore in the PEG ligand, and the complexity of the heterogeneous final product mixtures.[63] The first objective was to assess the efficacy of absorption and IPF spectroscopy for monitoring PEGylation reactions with a 5 KDa PEG ligand. Second, to investigate the use of pEEM and RS to measure protein structural changes and other physicochemical changes caused by PEGylation. Finally, we investigate the use of various spectroscopic data to develop an accurate method for predicting the degree of PEGylation.

3.2. The conjugation reaction

The reaction conditions used here were expected to result in the attachment of between one and six PEG residues to each lysozyme molecule, with the PEGs being attached following the order of lysine reactivity: Lys33>Lys97>Lys116.[169] The reaction takes place between the amino reactive groups in the protein (6 ε-amino from lysine residues and 1 N-terminal NH_2) yielding a stable urethane linkage and releasing a p-nitro phenol (pNP) leaving group in the reaction mixture (Figure 3.1). Despite the

reasonably well controlled conditions, the final products should be a heterogeneous mixture of proteins because of the different number of PEG conjugated to each molecule (PEGylation status), the potential presence of positional isomers, and the heterogeneity arising from the polydisperse PEG ligand.

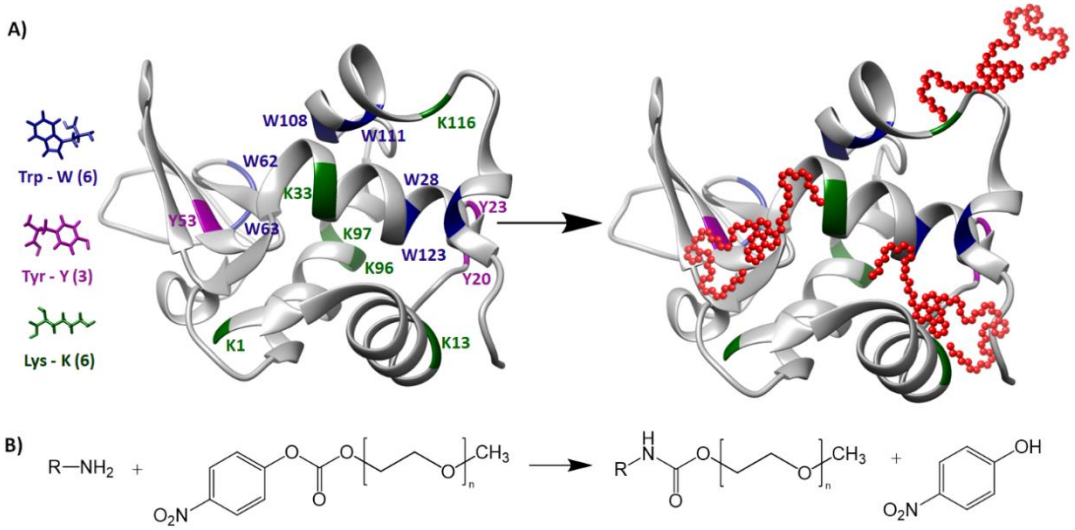


Figure 3.1- (A) Crystal structure of hen egg-white lysozyme (PDB entry: 1DPX Chimera software) showing location of Trp and Tyr fluorophores and the lysine sites for conjugation. (B) Chemical reaction scheme for the conjugation of pNPC- PEG to lysine.

The first change observed during the reaction was an increasing yellowing of the reaction mixture caused by release of the pNP group during conjugation, which immediately caused fluorescence quenching *(vide infra)*. In order to remove the small molecule reagents, the reaction mixture was purified and the retained sample was recovered using the original buffer (Phosphate pH=8.0) to an approximate concentration of 1.0 g/L (yield=75%).

The main change observed in the absorption spectra after purification (compared to the starting materials), was a 4% decrease in A_{280} (LZ = 0.99±0.01 and PEG-LZ = 0.95±0.02 g/L, Table 2.2), which agreed with the 4% decrease in fluorescence intensity. This was consistent for all products and is probably caused by some protein loss during the purification and reconstitution steps[a]. The purification strategy adopted

[a] The purification step involved extensive sample handling and was a long process (~3-4hours for the purification of 3 reaction mixtures) because the centrifugal filters used were of low capacity, requiring the splitting of the unpurified reaction mixture into a few aliquots until all the sample had been purified (Details in Section 2.2). This increased the probability of protein loss due to handling errors and/or adherence of the protein to the container walls (despite the fact we were using LoBind containers and pipette tips to reduce the loss). Also, the reconstitution step was very prone to errors because the approximation of final concentration was done by checking the absorbance and adding more buffer when necessary, which is not a very reproducible strategy and involved extensive handling.

here allowed for removal of the pNP quencher as shown by the control experiment (Figure 7.4) but did not remove any free PEG residues. However, the free PEG did not cause any significant change in emission properties even at the highest concentration (24 molar excess) used for these reactions (Figure 7.5).

3.3. Characterization by conventional techniques:

SEC was used to evaluate the degree of PEGylation and the purified product chromatograms (Figure 2A) showed five peaks: Peak1 (non-PEGylated lysozyme, Rt=14.6 min), Peak 2 (1PEG–LZ, Rt=12.8 min), Peak 3(2PEG–LZ, Rt=11.8 min), Peak 4 (3PEG–LZ, Rt=11.1 min) and Peak 5 (Multi-PEG–LZ, Rt=10.8 min). All reactions produced mixtures of pegylated species and the areas of peaks 2-5 increased with higher PEGylating agent concentration and reaction times (Table 2.2 and Table 3.1). It was observed a shift in the chromatogram of some samples run after an attempt of column recovery, which probably caused small changes in column packing and retention times. However, this did not seem to impact in the column recovery and peak areas. There were some small measurement variations (between injections) and the average chromatograms of three injections were therefore used for calculating the % area under the curve (AUC) of each peak. (Figure 3.2A)

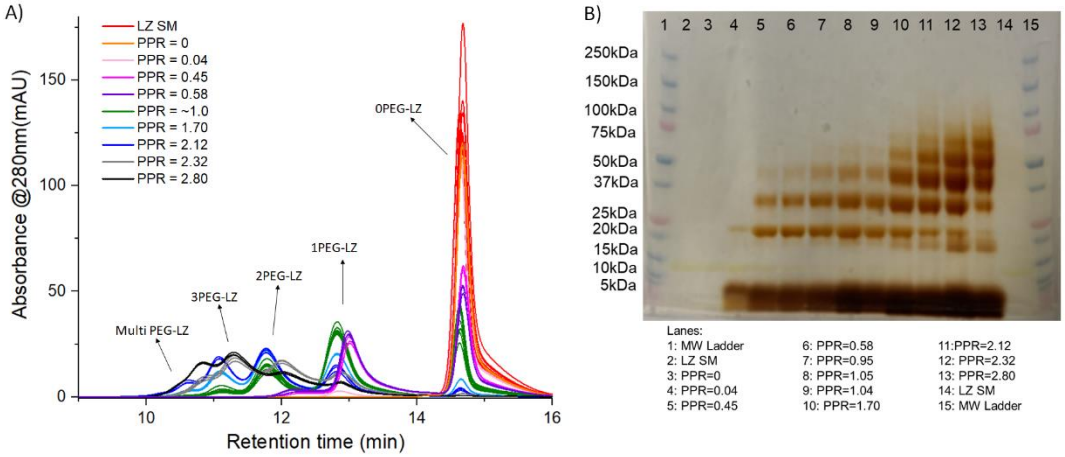


Figure 3.2- Analysis of LZ starting material and purified reaction products: (A) Unprocessed SEC chromatograms; (B) SDS-PAGE gels with Barium Iodide staining.

The reaction conditions chosen here should result in a heterogenous population of PEGylated species in the final reaction mixture, and thus the degree of PEGylation

is a critical quality attribute. Here, we used two parameters to define the degree of PEGylation: PEG-LZ$_{\%}$ and PPR. (Table 3.1)

The total PEG-LZ$_{\%}$ (Equation 3.1) parameter was defined as the sum of the % AUC of the four resolved PEGylated species peaks. A PEG to protein ratio, PPR, (Equation 3.2) was also defined which should be a more useful univariate value as it provides an estimate of the total number of PEG molecules attached to the protein sample.

$$\text{Total PEG LZ}_{\%} = \left(\frac{(\text{AUC}_{\text{Peak2}} + \text{AUC}_{\text{Peak3}} + \text{AUC}_{\text{Peak4}} + \text{AUC}_{\text{Peak5}})}{\text{AUC}_{\text{Total}}}\right) \times 100 \quad \text{Equation 3.1}$$

$$\text{PPR} = ((\%\ 1\ \text{PEG} - \text{LZ} \times 1) + (\%\ 2\ \text{PEG} - \text{LZ} \times 2) + (\%\ 3\ \text{PEG} - \text{LZ} \times 3) + (\%\ \text{Multi PEG} - \text{LZ} \times 4))/100 \quad \text{Equation 3.2}$$

Table 3.1-SEC PEGylation profile calculated for starting materials (SM) and PEG-LZ products. Values are the mean ± stdev of three replicates for each reaction condition

Reaction	%LZ SM	% 1PEG-LZ	% 2PEG-LZ	% 3PEG-LZ	% Multi PEG-LZ	Total PEG-LZ %	PPR
R28-30	100	-	-	-	-	-	-
R13-15	96.0± 0.1	3.9±0.1	0.1±0.0	-	-	4.07±0.13	0.04±0.00
R22-24	57.0±2.2	40.6±1.6	2.4±0.6	-	-	43.00±1.90	0.45±0.02
R31-33	48.1±1.3	46.3±0.4	5.6±1.4	-	-	51.90±1.23	0.58±0.02
R1-3	33.0± 3.5	43.6±0.4	19.5± 2.2	3.5±0.7	0.4±0.1	66.58±3.99	0.95±0.88
R4-6	28.4± 5.4	43.6±0.4	22.7±3.8	4.7±1.5	0.6±0.3	71.15±6.18	1.05±0.16
R16-18	28.3± 1.2	44.9± 0.3	22.2± 0.9	4.1±0.4	0.5±0.1	72.22±1.17	1.04±0.03
R7-9	9.4±0.3	33.7±1.0	37.2± 0.5	16.2±1.1	3.4±0.5	90.72±0.52	1.70±0.03
R10-12	4.7±0.5	22.5±1.9	37.4± 1.2	26.5±1.8	9.0±1.4	95.32±0.64	2.12±0.08
R19-21	1.0±0.0	21.2±1.1	35.1±0.6	32.5±1.1	10.9±0.9	96.05±5.51	2.32±0.04
R25-27	-	11.1±1.0	23.0±0.6	40.7±0.3	25.1±0.8	99.97±0.06	2.80±0.01

SDS-PAGE is widely used for the qualitative assessment of protein PEGylation. The band profiles in the gels (Figure 3.2B, lanes 10-13) with iodine staining indicated that more than four species were formed for some of the reactions which used higher excesses of the PEG reagent (12 and 24 molar excess) but unfortunately the conditions used for SEC did not allow for resolution of species with more than four added PEG. Further, the approximation of molecular weight (MW) based on the molecular ladder applied to the SDS-PAGE gels does not correlate well with the expected MW after

attachment of up to 6 PEGs[b] (maximum 39.3kDa). This MW overestimation (up to 75kDa for the highest pegylated species) is likely related to the steric hindrance caused by the attached polymers, which prevents the even binding of SDS molecules[c], causing different migration of unpegylated and differently pegylated LZ through the gels.[99, 171]

DLS of the purified PEG-LZ showed an increase in hydrodynamic radius (2.19 nm for LZ– 4.83 nm for the highest PEGylated products, Table 3.2) consistent with PEG attachment,[70, 77, 172] and also relatively small changes in polydispersity related to very low levels of larger particles (Figure 3.3A). These results suggest that PEG assumes a random coil conformation upon conjugation.

Table 3.2- UV aggregation index (UV-AI), Fluorescence intensity maxima (I_{MAX}), Total RS volume (RS), and hydrodynamic radius (R_h) for starting materials (SM) and PEG-LZ products. Values are the mean ± stdev of three replicates for each reaction condition.

Parameters/ Reactions		**UV-AI (%)**	**EEM$_{\parallel}$ I_{MAX} (a.u.)**	**RS (a.u.)**	**EEM$_{\perp}$ I_{MAX} (a.u.)**	**EEM$_T$ I_{MAX} (a.u.)**	**DLS R_h (nm)**
R28-30 (Control)	SM	0.07±0.02	871±6	11406±436	629±12	2121±29	2.00±0.06
	Prod	0.12±0.01	856±5	6024±158	629±8	2111±18	2.19±0.91
R13-15	SM	0.07±0.01	879±11	7197±236	637±5	2145±22	2.11±0.03
	Prod	0.47±0.20	859±21	6797±1494	631±12	2117±41	2.75±0.16
R22-24	SM	0.08±0.03	869±7	7272±59	636±4	2135±15	2.07±0.04
	Prod	0.17±0.03	857±10	7383±315	619±9	2091±22	3.00±0.72
R31-33	SM	0.20±0.09	878±18	8414±175	645±1	2157±15	2.03±0.01
	Prod	0.12±0.04	852±16	9712±1421	618±12	2086±37	3.27±0.08
R1-3	SM	0.08±0.03	913±12	6984±404	667±26	2244±26	2.01±0.04
	Prod	0.55±0.24	897±11	11942±1402	646±12	2178±33	3.33±0.12
R4-6	SM	2.35±0.03	891±10	11453±1699	665±12	2211±39	2.01±0.04
	Prod	0.29±0.09	889±22	10354±1359	651±13	2181±49	3.25±0.71
R16-18	SM	0.13±0.03	878±14	15056±156	635±10	2139±41	2.08±0.03
	Prod	0.25±0.03	853±12	9497±739	616±8	2084±26	3.45±0.12
R7-9	SM	0.11±0.08	885±5	11797±211	645±7	2170±24	2.05±0.01
	Prod	0.34±0.10	854±2	16481±1902	612±5	2076±6	3.77±0.08
R10-12	SM	0.16±0.13	879±19	10143±258	639±12	2146±33	2.07±0.02
	Prod	0.35± .06	825±13	14910±1143	605±19	2030±47	4.08±0.14
R19-21	SM	0.06±0.01	875±5	6835±117	640±8	2149±23	2.09±0.01
	Prod	0.24±0.06	857±9	13092±465	607±6	2066±20	4.30±0.25
R25-27	SM	0.09±0.03	864±19	7583±59	632±11	2126±39	2.03±0.07
	Prod	0.31±0.07	854±13	15003±615	601±15	2052±44	4.83±0.11

[b] Calculated as the sum of the individual diameters of a single lysozyme plus a single 5kDa PEG. Thus, the expected MW for the PEGylated species are: 1PEG-LZ: 19.3kDa, 2PEG-LZ: 24.3kDa, 3PEG-LZ: 29.3kDa, 4PEG-LZ: 33.4kDa, 3PEG-LZ: 39.3kDa and 5PEG-LZ: 43.4kDa.

[c] SDS is a surfactant commonly used in electrophoresis which allows the migration through the gel primarily based on mass by binding to the proteins and creating a uniform negatively charged surface. In the absence of SDS, proteins with similar MW could migrate differently due to differences in mass-charge ratio. [170]

The normalized absorbance spectra (Figure 3.3B) showed no shifts or spectral changes after the PEGylation and ROBPCA did not show any trends that could be used to monitor the reaction, discriminate products from starting materials or assess PEGylation degree. PC1 (47.54% of the variance) separated starting materials from R4-6 based on small differences in the 250–270 nm region and in an increased absorbance at 350 nm (Figure 3.3C/D).

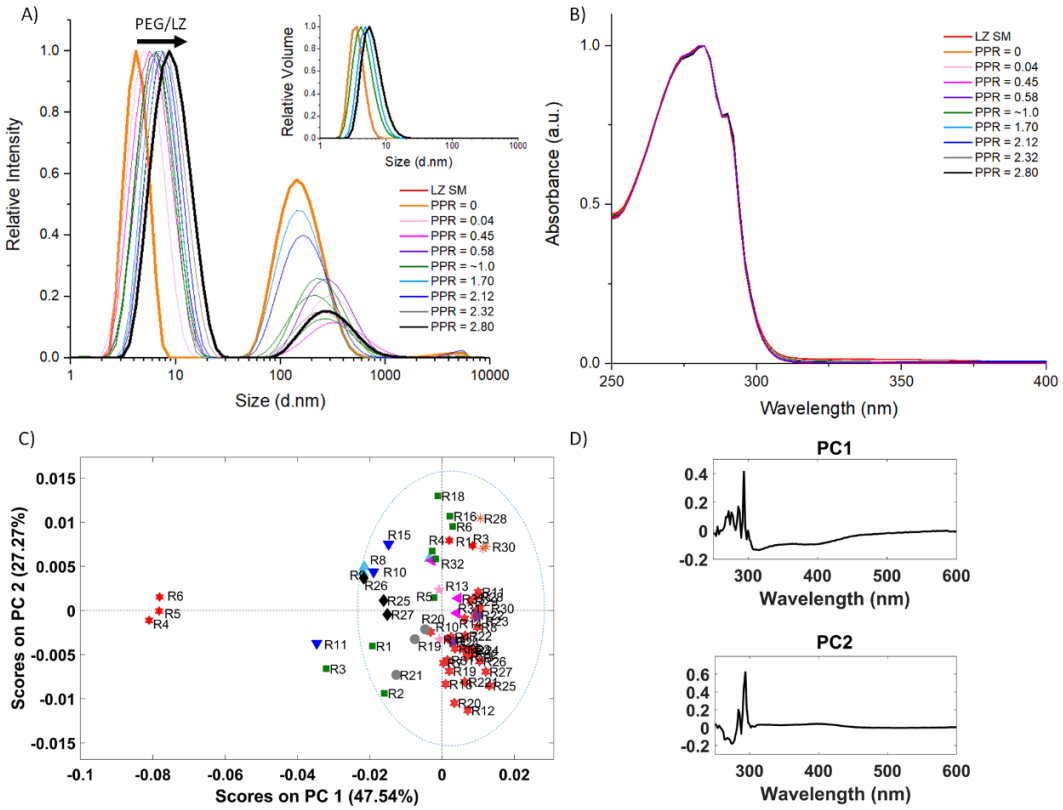


Figure 3.3- (A) Size distribution by intensity and volume from DLS measurements of all purified products (B) Normalized absorbance spectra of all starting materials and purified PEG-LZ; (C) ROBPCA scores plot and (D) loadings from the UV data showing a lack of sensitivity for discriminating PEGylated products and assessing conjugation degree. DLS and absorbance data shown are the average of replicates for each reaction condition.

PC2 had a significative contribution to the total variance (27.27%) but overall, it did not correlate with the degree of PEGylation in the purified products. When the outliers (SM from R04-06) were excluded from the model, it also showed no trend that could separate samples based on the PEGylation level. The UV-AI was also examined but did not provide reliable information about HMWS formation in this case. (Table 3.2) The UV-AI calculation (Equation 2.1) uses absorbance at 350 nm, which is usually very low and largely affected by noise. This makes it an unreliable tool for assessing physical status of the sample, as discussed in details in Chapter 4.[173]

Here the absorption spectra was used only to quickly evaluate the reproducibility of sample preparation in terms of protein concentration. The variability of both the starting materials (RSD[d]= <2%) and the purified products (RSD=<3%) was low, indicating that sample preparation and handling was well controlled.

3.4. Reaction Monitoring by fluorescence

In order to investigate spectral changes over the 24-hour reaction time, we carried out time course experiments using only EEM‖ measurements (Figure 3.4). The first significant change noticed was the dramatic decrease (~6×) in fluorescence intensity due to quenching, which was probably caused by the nitrophenol group of the PEGylating agent itself, and the para-nitrophenol leaving group.[105] Despite the weak and noisy signal caused by the strong fluorescence quenching after addition of the PEG regent it was possible to see an increasing blue-shift in emission peak and quenching over the course of the reaction, which was related to the amount of PEG reagent used.

As discussed in the introductory chapter, fluorescence emission is determined by the excitation wavelength, with λ_{ex} = 280 nm exciting both Tyr and Trp, whereas at λ_{ex}> 290 nm, mostly Trp is excited[105] and Trp 62 and Trp108 are the major contributors to lysozyme intrinsic fluorescence.[8] Thus, the blue shift observed in the time course experiments (from 355 to ~338 nm) is most likely due to quenching of these Trp by the nitro containing leaving group, and potentially a small contribution due to an increasing hydrophobic character in the environment around these Trp residues as more PEG ligands were attached. It is probable that the quenching effect is dominant because of large decrease in emission intensity observed and by recovery of emission intensity for the purified products (I_{MAX}, Table 3.2). Therefore, any fluorescence-based reaction monitoring will be mostly based on the quenching effect rather than direct change in structure and/or fluorophores environment induced by conjugation.

[d] RSD (Relative Standard deviation): ratio of the standard deviation to the mean value (here expressed in %) and was used here to assess variability within samples from a dataset.

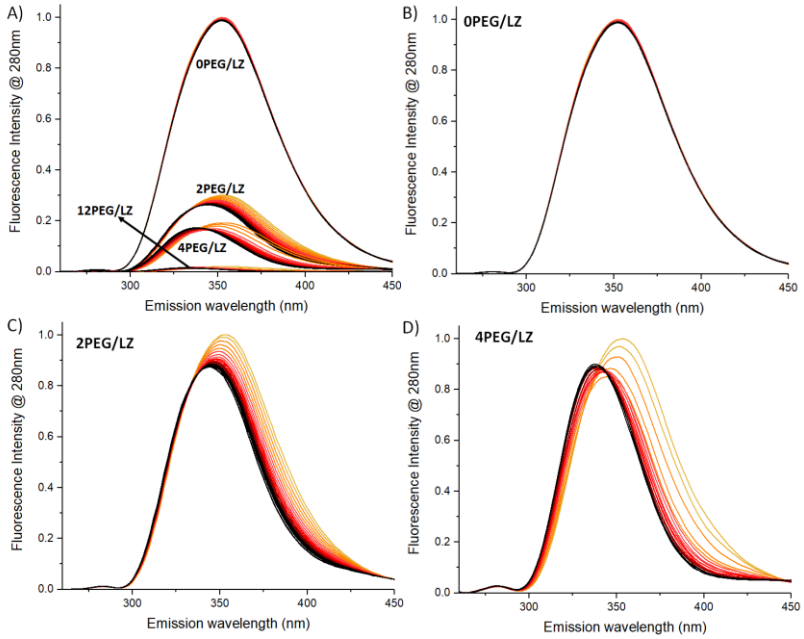


Figure 3.4- Fluorescence 2D spectra at excitation 280 of time course experiments for a control reaction where no PEG was added and reactions with various molar ratios of PEG reagent per lysozyme. The colour scale indicates the time course of the reaction: 0h (Orange) → 24h (Black). The intensity was normalized to the maximum peak of the respective starting materials for each reaction.

3.5. Products characterization by pEEM

Once the products had been purified and the small molecule quenchers removed, the fluorescence intensity was recovered as seen in the pEEM measurements (I_{MAX}, Table 3.2). The 2D spectra (λ_{ex} = 280 nm) showed very small spectral changes, which indicates that it is insensitive for a conclusive assessment of purified products. Simple univariate indexes can be extracted from the 2D fluorescence plot to screen for trends in sample spectra. (Figure 3.5B). For instance, the I_{350}/I_{330} (λ_{ex} = 296 nm) ratio showed a small increase with increasing degree of PEGylation which could be explained as being due to a progressive drop in the intensity of the directly excited, solvent exposed Trp emission concomitant with an increase in the blue shifted Trp emission.

This suggested an increased relative emission from Trp in more hydrophobic environments. It is also clear that the degree of change increases as the degree of PEGylation increases. We can discount quenching-based effects here because the nitro containing species have been removed in the purification, as confirmed by the control experiment with p-nitrophenol.

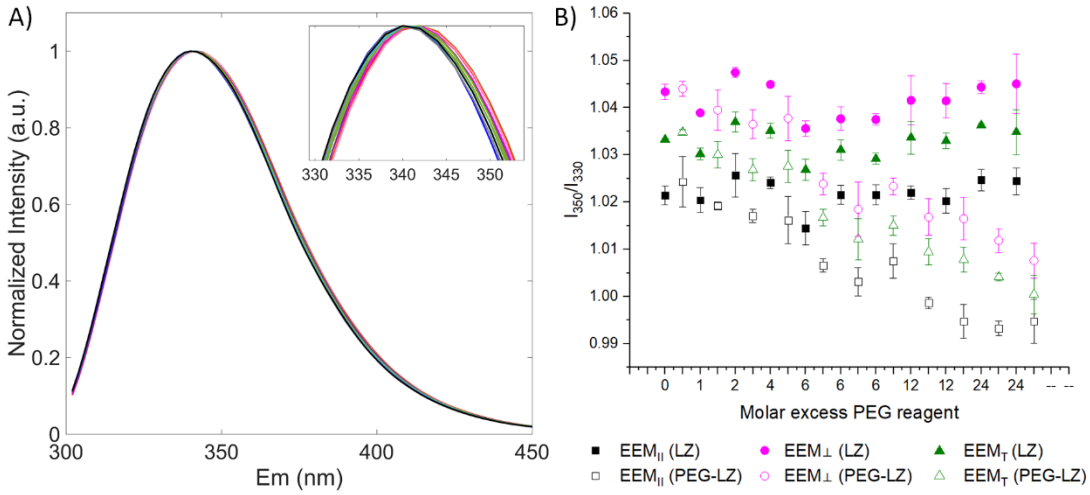


Figure 3.5- (A) Normalized parallel polarized 2D Fluorescence spectra at λ_{ex}=280 nm and (B) Ratio I_{350}/I_{330} (λ_{ex} = 296 nm) parameter calculated using $EEM_{\parallel}$, $EEM_{\perp}$ and EEM_{T} from the various starting materials and purified products. Full symbols are the starting materials (SM) and open symbols are the respective purified products (PEG-LZ). All the data is shown as the average of three replicates and the error bars indicate the standard deviation.

To make better use of fluorescence one needs to examine the complete emission using pEEM (or pTSFS) measurements. The normalized EEM difference spectra showed significant, but relatively small spectral changes.(Figure 3.6) The structure of PEG residues attached to proteins in solution has been shown by SANS and NMR to be essentially separate from the protein core, and not wrapped around the protein.[70, 174] If this is the case here, the conjugation should have only a small effect on Trp emission, which agrees with the maximum changes of ~4% observed in the difference spectra. As more PEG residues were added we saw a progressive emission change, which was dependent on the proximity of the attachment site with Trp62 and 108 (the major contributors to LZ intrinsic fluorescence).

Analysis of the distances involved, calculated using structural analysis tool in UCSF Chimera v1.14, showed that, of the three potential attachment sites (K33>K97>K116), K97 and K116 were significantly closer (by ~9 and 3 nm on average) to these Trp compared to K33. (Figure 3.7) This suggested that the biggest emission changes might arise after attachment of the second PEG at one of these sites. This correlated with the u-PLS regression models for the individual species (*vide infra*) which showed that model for the 2 and 3 PEG-LZ had the best linear correlations.

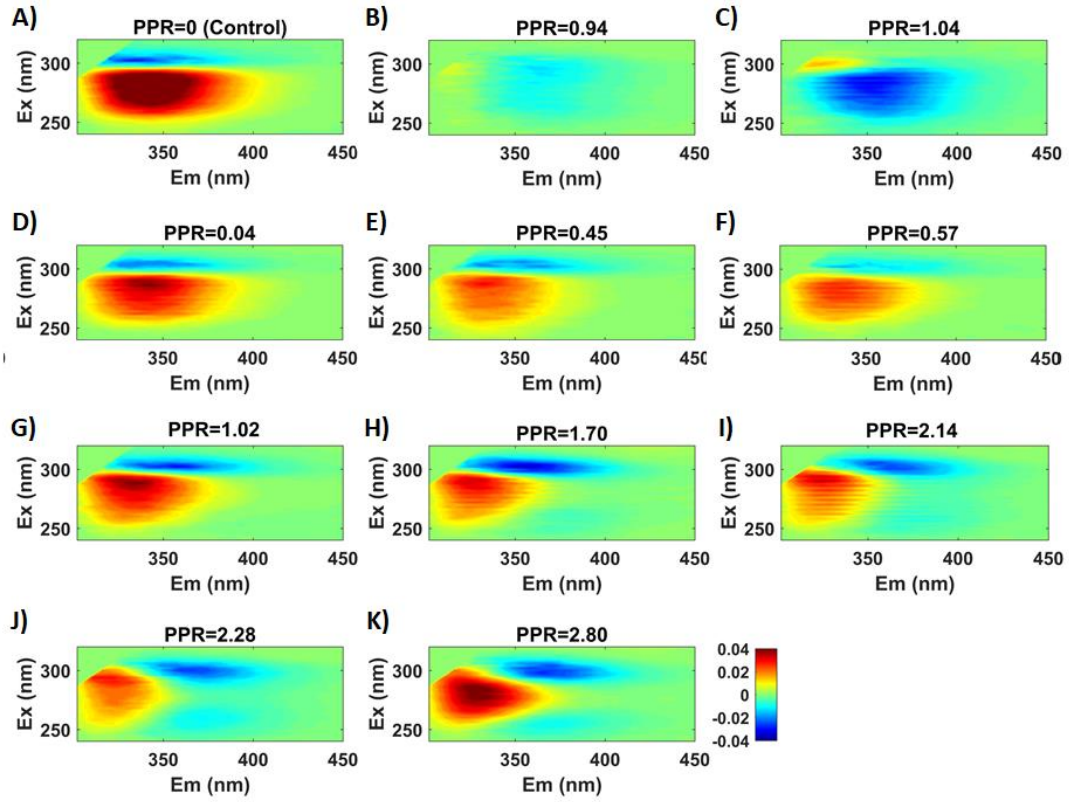
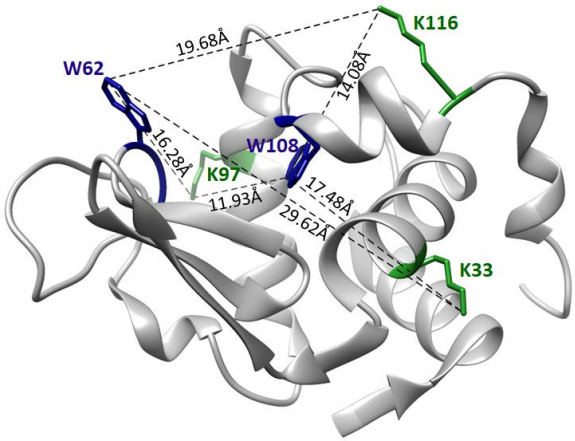


Figure 3.6 - Difference spectra calculated using normalized EEM‖ data from the purified product material and the lysozyme SM for samples with different PPR. Difference spectra was calculated using the average spectra of staring materials and products for each condition. ($EEM_{Dif.} = EEM_{Mean\ PROD} - EEM_{Mean\ SM}$)



Atom1	Atom2	Distance(Å)
W 62.A CH2	K33.A NZ	29.62
W 62.A CH2	K97.A NZ	16.28
W 62.A CH2	K116.A NZ	19.68
W 108.A CZ2	K33.A NZ	17.48
W 108.A CZ2	K97.A NZ	11.93
W 108.A CZ2	K116.A NZ	14.08

Figure 3.7- Crystal structure of hen egg-white lysozyme (PDB entry: 1DPX Chimera software) showing the distance between residues Trp (W) 62 and 108 and Lys (K) 33,97,116 calculated using structure analysis tool from Chimera.

For the control reactions (R28-30), there were differences in the Trp and Tyr spectral regions indicating structural changes caused by solution stirring, sample handling, and the purification processes. In addition there was an overall increase in

light scatter (i.e. RS volume extracted from EEM‖) observed for most reactions, which indicated HMWS formation,[175] probably of weakly bound species. (Figure 3.8A). This agrees with the DLS measurements (Figure 3.8B) of a control reaction (with no PEG reagent added), which indicates an increase in sample polydispersity (PdI=0.49 to 0.81) with the stirring but no overall increase in the hydrodynamic radius was observed at times 0, 6 and 24 hours (R_h=2.08 to 2.07 nm). In terms of size changes due to PEGylation, the PPR ratio from SEC (Table 3.1) correlated linearly with the hydrodynamic radius values from DLS measurements (r^2= 0.95) but the correlation with RS from the purified products was less good (r^2 = 0.72). One of the reasons for this poorer correlation is the low reproducibility associated with RS measurements, particularly due to shot noise. Ideally, RS measurements need to be averaged over several acquisitions to reduce the measurement variance which here were >10% in some cases and is much more significant than the emission component. (The use of RS will be discussed in details in the next chapter)

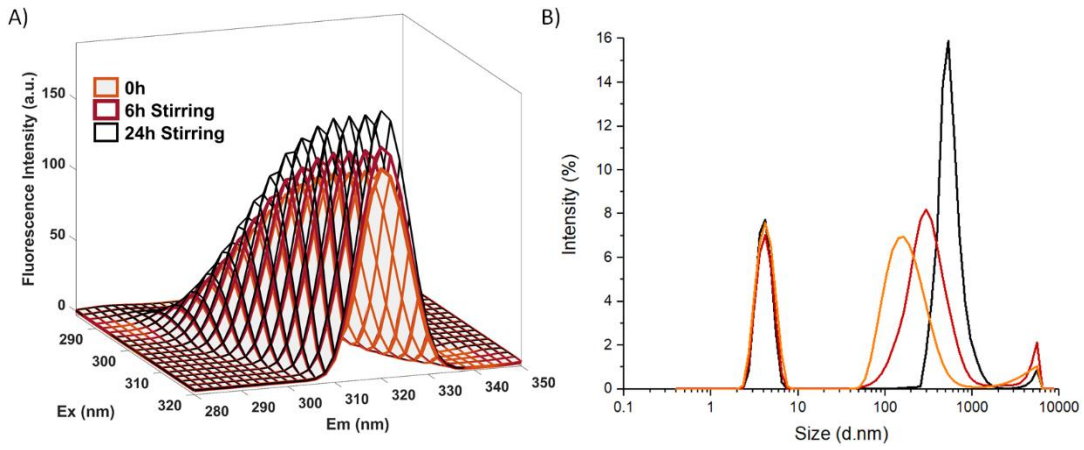


Figure 3.8- RS‖ band extracted from EEM (A) indicates a progressive increase in particle size over time which corroborates with the DLS data (B) collected at time 0,6,24h showing no change in protein radius (first peak) but the increase in sample polydispersity after the 24h stirring.

The difference spectra of R1-6 were significantly different compared to all other reactions (Figure 3.6B/C) and this might have been caused by the different conditions of the starting materials used for this reactions. This variance can be explained by the fact that R1-3 were the first reactions carried out and probably the least controlled samples in terms of sample handling. The starting materials used for R4-6 had been used for a 24hours stirring control experiment and were thus stirred for much longer

than the other reactions. So, despite the fact that a single lot of material was used for all reactions, sample handling and mechanical stress induced by stirring can cause significant structural changes in lysozyme which is manifested as spectral changes in emission.

A separate unfolding experiment using guanidine indicated that the R1-R6 samples had some spectral similarities to the unfolded protein, which supports the hypothesis that the stirring/different sample handling of the R1-6 samples caused changes in the protein and was the source of variation. (Figure 3.9) This further demonstrates the utility of fluorescence EEM measurements for quality assurance testing of protein starting material variance as discussed in Chapter 4.[173]

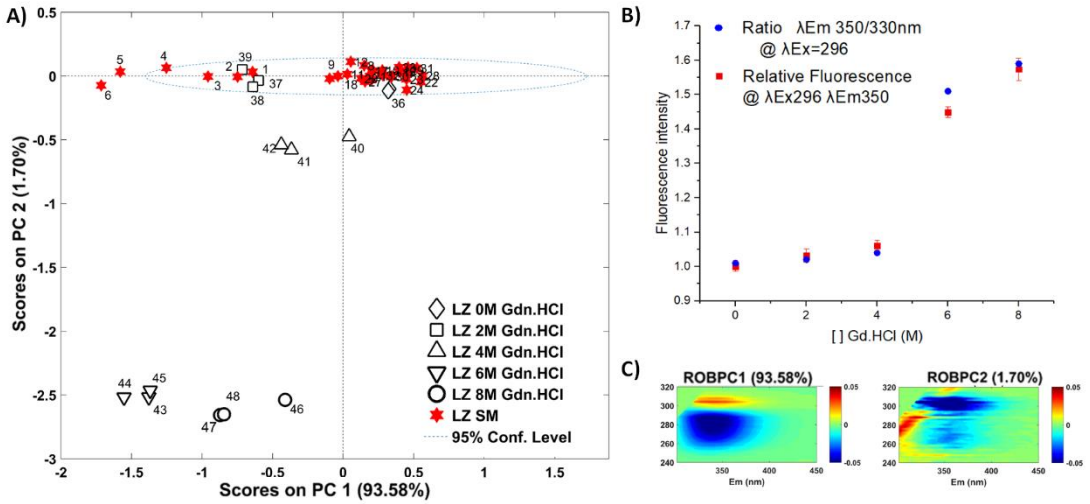


Figure 3.9- ROBPCA (A) scores plot and (C) loadings shows the clustering of staring materials from reactions 1-6 with unfolded LZ using 2M of Gdn.HCl. This suggests that the difference of these starting materials with the other SM could have arisen from partial unfolding before the reaction (probably caused by poorly controlled sample preparation and/or handling). (B) Ratio I_{350}/I_{330} (λ_{ex} = 296 nm) and Relative I_{350} (λ_{ex} = 296 nm) used to confirm LZ unfolding.

3.6. *Aniso*-EEM for product variance

The anisotropy maps of lysozyme samples show regions of high anisotropy (~0.20) around the excitation wavelength of 300 nm, which is due to tryptophan residues.

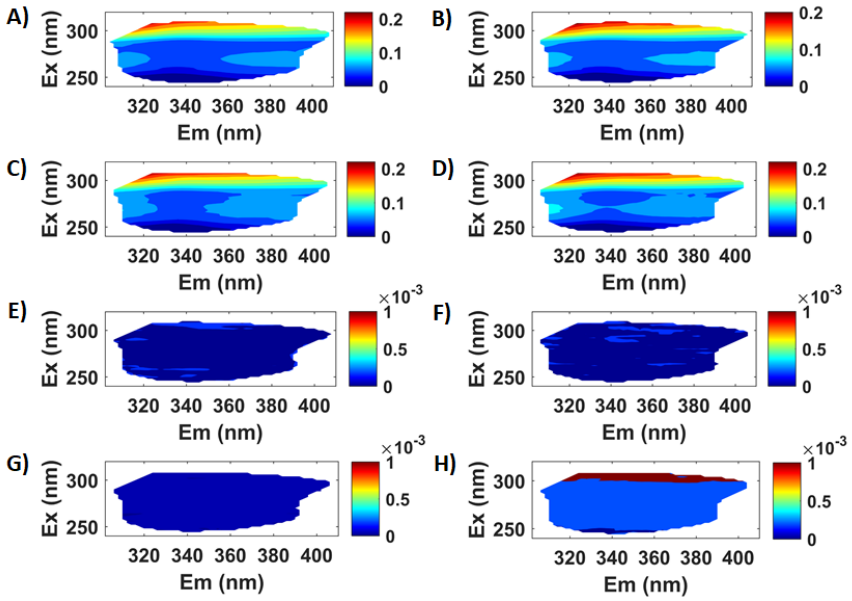


Figure 3.10- *Aniso*-EEM plots of (A) lysozyme (mean of all SM used in the reactions), (B) a low (PPR=0.04), (C) medium (PPR=1.02) and high (PPR=2.80) purified PEG-LZ products. Difference spectra calculated for (E) the control reaction where no PEG reagent was added, (F) the low, (G) medium and (H) high PEGylated product. Difference spectra was defined as *Aniso*-EEM$_{Dif.}$ = *Aniso*-EEM$_{Mean\ PROD}$ – *Aniso*-EEM$_{Mean\ SM}$.

Anisotropy values varied more in the region λ_{ex}=270 to 290 nm, which is a region of excitation of both Tyr+Trp, but only small apparent variations were observed in the maps of purified products compared to starting materials. The difference spectra (Figure 3.10E-H) indicate larger changes in the maps of products with higher PPR, however, the overall low anisotropy values and low signal to noise ratio (SNR) limits the use of *Aniso*-EEM as a robust tool for assessing structural variance of PEG-LZ products.

3.7. Discriminating variance in starting materials and PEGylated products

Three PCs were required to account for 97% (EEM$_{\parallel}$), 99% (EEM$_{\perp}$), and 99% (EEM$_T$) of total variance and the scores plot easily discriminated starting materials from products, and also products with different degrees of PEGylation (Figure 3.11A,B,C and Table 3.3).

Table 3.3- Summary of ROBPCA results using normalized pEEM spectra for models containing all LZ and purified PEG-LZ samples, a reduced sample set (excluding R1-6) and only the purified products (excluding SM).

	All samples			Reduced sample set		Only products	
ROBPC	**EEM$_{\parallel}$**	**EEM$_{\perp}$**	**EEM$_{T}$**	**EEM$_{\parallel}$**	**EEM$_{\perp}$**	**EEM$_{\parallel}$**	**EEM$_{\perp}$**
1	89.65	95.68	94.75	83.10	91.50	68.22	78.19
2	6.64	2.15	3.20	8.44	1.70	24.29	18.58
3	0.79	0.69	0.55	2.19	2.21	-	-
Total variance	97.08	98.53	98.51	93.73	95.41	92.51	96.77

The different trends produced by the various measurements is due to the different levels of light scatter, noise, and fluorescence signal in each. The EEM$_T$, being the least noisy, usually produces a model with high percentage of variance explained, whereas EEM$_{\perp}$ data should have the purest emission but is noisier due to lower measurement intensities, and EEM$_{\parallel}$ should be more impacted by the scatter.

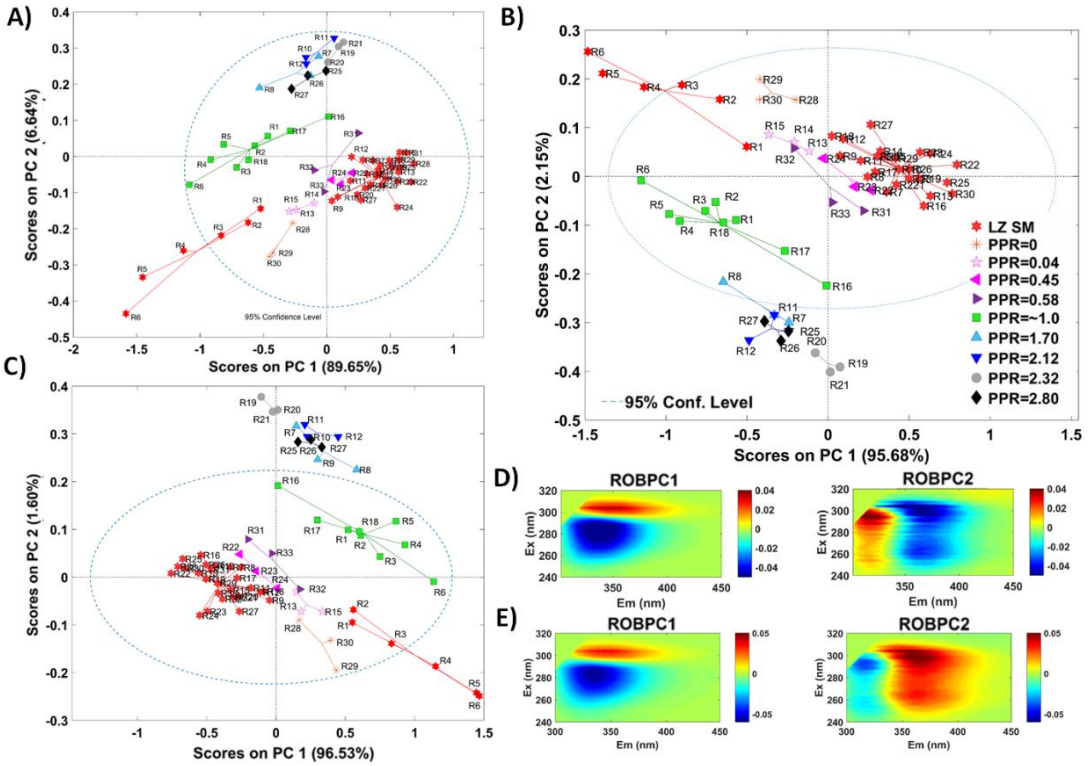


Figure 3.11- ROBPCA scores plots obtained using normalized: (A) EEM$_{\parallel}$; (B) EEM$_{\perp}$; and (C) EEM$_T$ data. Refolded ROBPC1 and 2 loadings from the (D) EEM$_{\parallel}$ and (E) EEM$_{\perp}$ normalized ROBPCA models.

The EEM$_{\parallel}$ model seemed to separate better samples with low level of conjugation, whereas EEM$_{\perp}$ better discriminated some samples with higher degree of PEGylation (R19,20,21), which may be due to the changing nature of the size and photophysical changes taking place. PC1 scores identified starting materials used in

the first reactions (R1-6) as being significantly different and their products were also discriminated because of spectral changes, which are probably linked to structure/composition differences. The refolded robust PC loadings plots (Figure 3.11D/E) showed the areas of highest spectral variance in the pEEM data. Starting material samples had the same size based on DLS measurements, which indicates that the size changes involved (if any) were too small to be discriminated by DLS. (Table 3.2) This suggests that the variability observed in pEEM is probably related to more subtle changes like partial unfolding where there is a minimal change in overall size.

When all samples were considered (n=66), the main spectral feature responsible for the separation of starting materials and products was the change in the contribution between the Tyr/Trp area (λ_{ex} ~280 nm) and Trp only area (λ_{ex} >302 nm) caused by reaction induced structural changes leading to variations in IFE and FRET rates. ROBPC2 seemed to be more related to differences between the various products and represent changes in Trp emission caused by changes in the environment around Trp 62 and Trp108 (increasing hydrophobic character). There was a positive contribution from more exposed Trp (λ_{em}>346) to the lower PEGylated species and a negative contribution for the higher PEGylated species. The later PC's had less discriminating power as the spectral changes are relatively small. After samples R1-6 were excluded from the model a similar trend was observed. ROBPC1 was more related to the difference between the SM and products clusters, and the discrimination between different PEG-LZ seems to be related to a combination of PC1 and 2. This was done to evaluate if it was possible to extract more information about the different products when these outliers were not present. (Figure 3.12)

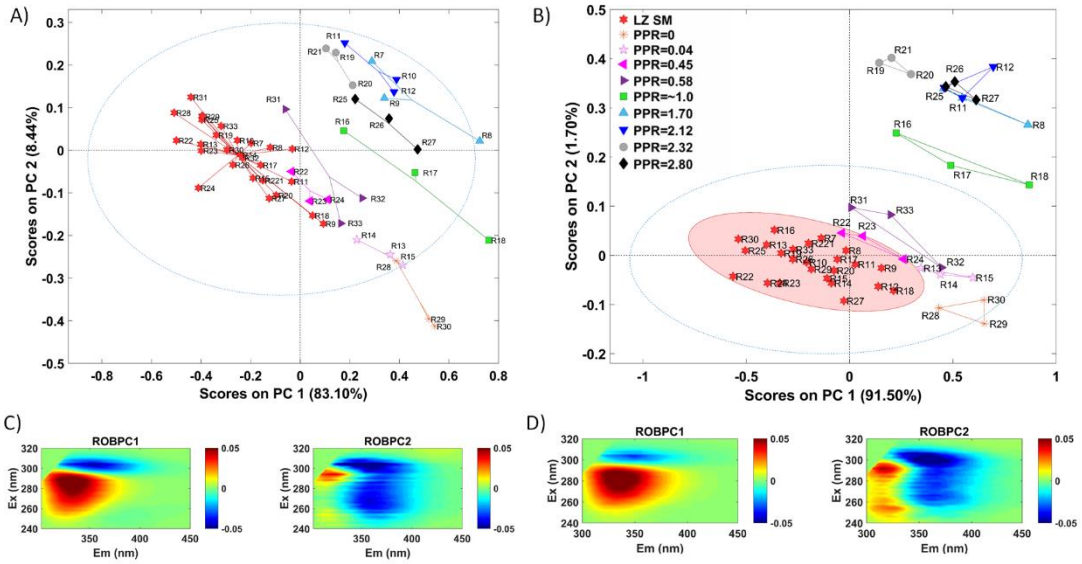


Figure 3.12- ROBPCA scores and refolded loadings plots of starting materials and purified products from reactions R07-33 (n=54) using normalized (A,C) EEM∥ and (B,D) EEM⊥ spectra.

When only the purified products were considered (thus not including the large starting material variation), the model better discriminated the structural changes induced by PEGylation and ROBPC1 (Figure 3.13C/D), seemed to be related to the emission of a more solvent exposed Trp population (λ_{ex}~280).

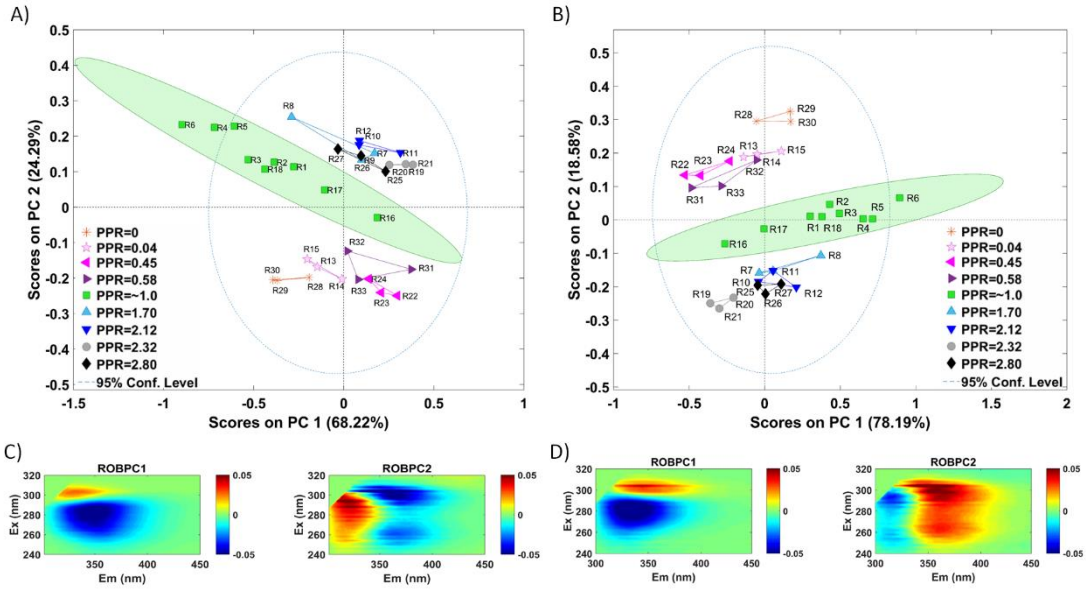


Figure 3.13 - ROBPCA scores and loadings plots of the purified products from all reactions (n=33) using normalized (A/C) EEM∥ and (B/D) EEM⊥ spectra. Here we excluded starting materials.

For both models, it was observed a big change in the % of variance explained by each ROBPC and the $EEM_{\parallel}$ and $EEM_{\perp}$ scores plot (Figure 3.13A/B) showed similar discrimination trends (inverted PC2 orientation) with the $EEM_{\perp}$ model showing slightly better discrimination. Comparison of the PC1 loadings showed significant spectral changes, with the $EEM_{\parallel}$ having weaker, blue shifted emission when excited at 300 nm, but more red shifted emission for λ_{ex} <300 nm compared to $EEM_{\perp}$ (Figure 3.13D). Overall, the ROBPCA results indicated that the observed pEEM changes, while relatively small, were significant and there is sufficient spectral change induced by PEGylation to consider developing a predictive model for PEGylation quantification.

3.8. Predicting degree of PEGylation using pEEM:

Here, we first assessed the different combinations of measurements (UV-Vis, $EEM_{\perp}$, $EEM_{\parallel}$, and EEM_T) and parameters (Total PEG-LZ%, PPR and % of individual PEG-LZ species[e]) to determine which combination would be the best for determining the PEGylation status of each sample. The total PEG-LZ% (Equation 3.1) was the simplest parameter used to assess the gross degree of PEGylation and it was defined as the sum of the %AUC of the peaks for the four resolved PEGylated species. The PPR (Equation 3.2) however should be a more useful univariate value as it takes into account the actual PEG contribution of each species resolvable by SEC.

For building the predictive models, samples of purified products (n=33) were split into calibration (n=25) and validation (n=8) datasets using the Kennard-stone algorithm [176] and the various parameters obtained from SEC were used as the nominal values. Despite a reasonably good linear correlation between fluorescence and individual % of some species (e.g. 2 and 3 PEG-LZ, as seen by the R^2 values, Table 7.4, and Figure 3.14), most models had high relative errors of prediction (>30%) which suggests, as expected, that the fluorescence response is related to a combination of all species contained in the sample. (Section 7.6.3.)

[e] % of individual refers to the distribution of each non-PEGylated LZ, 1PEG-LZ, 2PEG-LZ and 3PEG-LZ in the purifed products. It was defined as a percentage of the area in the SEC data, e.g. %2PEG =(AUCpeak3/AUCtotal)×100).

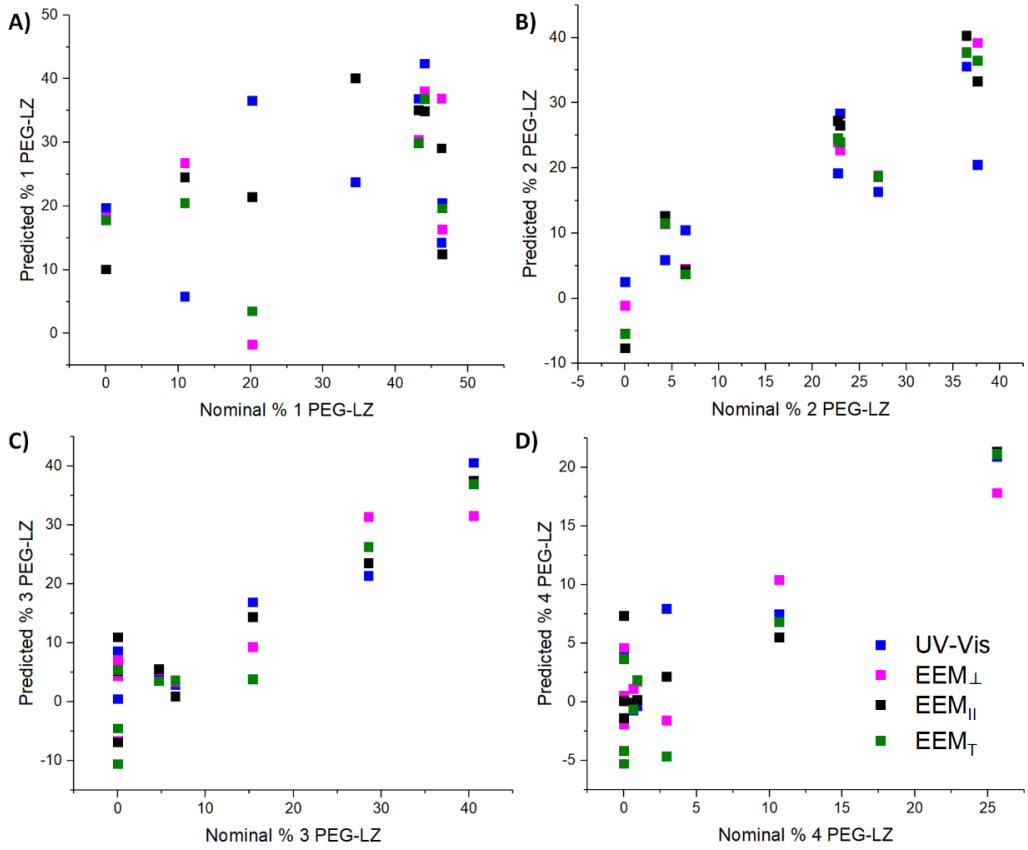


Figure 3.14- Plots of SEC nominal vs predicted % of individual species containing 1 (A), 2 (B), 3 (C), and 4 (D) PEG per LZ. The % of individual species was calculated using the AUC obtained from the SEC chromatograms.

As previously discussed, the significantly better correlations for 2 and 3PEG species content, validates our hypothesis that the biggest emission changes will occur after attachment to the Lys residue closest to Trp68 and 102. There was also a very poor correlation between absorbance spectra and any of the parameters evaluated here (as shown by EJCR in Figure 3.15A/B and RMSEP, R^2 values, in Table 3.4) and thus absorbance spectroscopy is not a good candidate for the analysis of PEGylation. The lowest prediction errors and highest R^2 were observed when using $EEM_{\perp}$ and $EEM_{\parallel}$ (REP values that were ~30% lower) as shown in Table 3.4, thus further optimization was carried out using the selected parameters $PEG\text{-}LZ_{\%}$ and PPR and the two polarized measurements.

Optimization involved comparing models using non pre-processed data [f](NoPrep), normalized to the maximum (Norm), and mean centered (MC). The results (Table 3.4) indicated that both $EEM_{\parallel}$ and $EEM_{\perp}$ data yielded statistically similar

[f] Pre-processing here means the processing of the data for the modelling only (besides blank subtraction + missing data + smoothing + unfolding)

prediction results, and that pre-processing did not have much of an impact on model accuracy (REP). The error in the SEC based reference method (Appendix 7.6.2) was estimated at ~7.2% for PPR and 4.4% for total-PEG and the PPR error is comparable to the RMSEC values (~10%) obtained for pEEM based models and the higher %REP errors. It should be noted that despite the higher error obtained with pEEM (REP $EEM_{\parallel}$=17%), the EJCR (Figure 3.15C/D) shows that all the models were accurate. Further, improvements in the prediction ability are not feasible using chemometric approaches on this data, because this will probably lead to non-robust solutions and over fitting. Thus, in order to improve predictive model capability, one needs to first improve SNR, and potentially use a better parameter for assessing PEGylation.

Table 3.4 -Comparison of u-PLS model results using UV-Visible, EEM_T, $EEM^{\perp}$ and $EEM_{\parallel}$. Models using $EEM_{\perp}$ and $EEM_{\parallel}$ were optimized by using different pre-processing methods: Norm (normalization to the maximum) and MC (mean centering). NoPrep. means non-preprocessed data. The values highlighted in yellow represent the best prediction performance, and these were the models chosen for optimization.

Measurem.		**UV-Vis**	**EEM_T**	**$EEM^{\perp}$**				**$EEM_{\parallel}$**			
Pre-processing		**No Prep**	**No Prep**	**No Prep**	**MC**	**Norm**	**Norm + MC**	**No Prep**	**MC**	**Norm**	**Norm + MC**
PEG-LZ%	**RMSE Cal**	8.2	10.0	**9.9**	9.7	10.3	3.4	**8.9**	8.1	9.2	9.1
	RMSE CV	20.3	10.6	**10.6**	11.1	11	10.2	**10.7**	10.2	11	11.2
	RMSE Pred	15.7	11.4	**11.1**	11.2	11.2	17.9	**10.7**	17.1	16.0	15.4
	REP (%)	**23**	**17**	**17**	**17**	**17**	**27**	**16**	**19**	**18**	**17**
	R^2 Cal	0.94	0.92	**0.92**	0.92	0.91	0.99	**0.93**	0.95	0.93	0.93
	R^2 CV	0.72	0.91	**0.91**	0.90	0.90	0.92	**0.91**	0.91	0.90	0.90
	R^2 Pred	0.79	0.93	**0.93**	0.93	0.92	0.74	**0.93**	0.95	0.93	0.92
PPR	**RMSE Cal**	0.24	0.15	0.13	0.13	**0.12**	0.10	0.06	0.06	0.06	**0.06**
	RMSE CV	0.31	0.19	0.21	0.19	**0.19**	0.19	0.17	0.16	0.16	**0.16**
	RMSE Pred	0.37	0.25	0.22	0.23	**0.21**	0.24	0.22	0.24	0.22	**0.22**
	REP (%)	**29**	**20**	**18**	**19**	**17**	**19**	**17**	**19**	**18**	**17**
	R^2 Cal	0.93	0.97	0.98	0.98	**0.98**	0.99	1.00	1.00	1.00	**1.00**
	R^2 CV	0.88	0.96	0.95	0.95	**0.95**	0.96	0.97	0.97	0.97	**0.97**
	R^2 Pred	0.84	0.95	0.95	0.94	**0.96**	0.94	0.95	0.94	0.95	**0.96**

As high SNR pEEM measurements might not always be feasible for reaction monitoring using scanning based spectrometers, we next investigated if it was possible to reduce the data acquisition time by evaluating the use of variable selection [177] to remove any non-essential information from the pEEM. The variable selection methods reduced the numbers of variables required significantly. The newly selected variables could be collected using fewer excitation wavelengths which could translate into much shorter acquisition times. We estimate that this could be 30 or 60 seconds instead of the 410 seconds needed here. (Table 3.5)

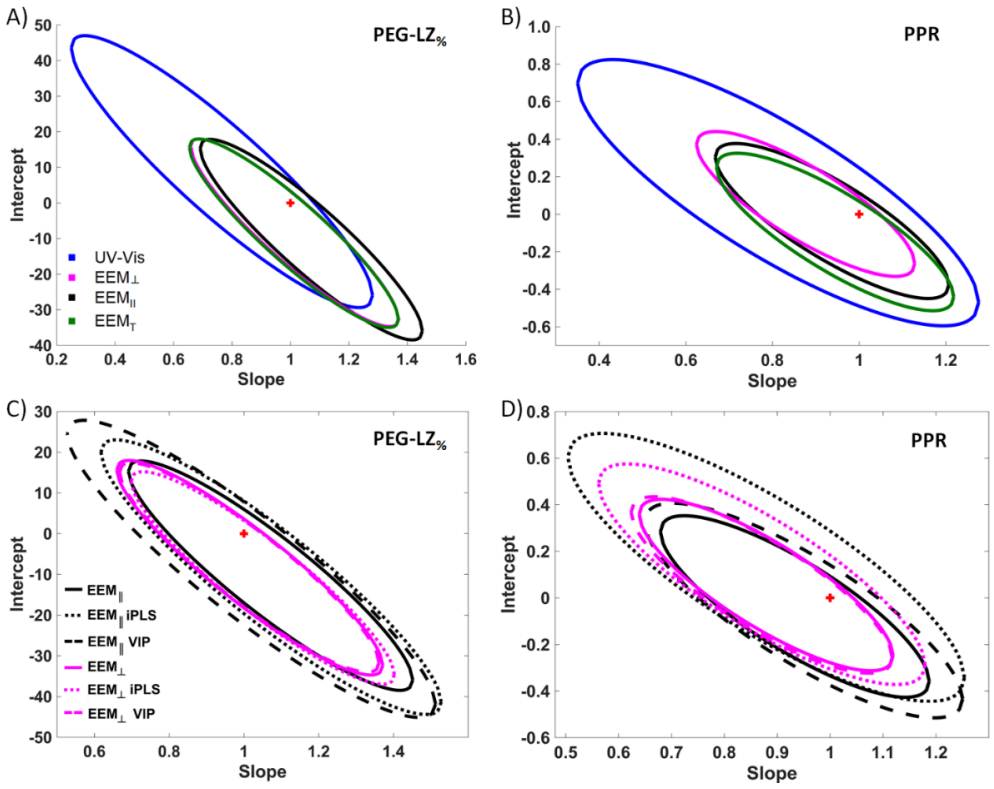


Figure 3.15- EJCR at 95% confidence level comparing the performance of the u-PLS models for the quantification of (A) PEG-LZ% and (B) PPR using different methods (without pre-processing). EJCR comparison using variable selection methods for the quantification of (C) PEG-LZ% using EEM⊥ and EEM‖ methods (without pre-processing) and (D) PPR using EEM⊥ (Normalized), and EEM‖ (Normalized and Mean centering). The red symbol represents the ideal point (1,0).

Both iPLS and VIP facilitated a substantial decrease in the data collection time (from 410s to 30s or 60s) without compromising model accuracy. EJCR testing of the different models and data for the degree of PEGylation (Figure 3.15) showed that for all the models, the ideal point (0,1) lay inside the ellipse, implying that all methods were accurate. [196] When assessing model quality, the EEM⊥ (iPLS) model was best for PEG-LZ% whereas for the PPR parameter the iPLS models were significantly worse.

Overall, the PPR value was slightly better correlated with the emission properties, but more poorly predicted (REP and LOD) than PEG-LZ$_{\%}$, which was probably a consequence of the way the parameter was calculated.

Based on statistical parameters, EEM$_{\parallel}$ measurements with iPLS seemed to be the best option for predicting PPR, but with this data the differences were not statistically significant ($p>0.05$ based on a randomization test). In terms of data collection time, EEM$_{\parallel}$ may offer an advantage compared to EEM$_{\perp}$ as it can be implemented as a single measurement (there is no need for G factor correction and consequently the collection of EEM$_{HH}$ and EEM$_{HV}$ data). With regard to the selected variables (Figure 3.16 and Figure 3.17), VIP selected variables over a large emission range thus requiring EEM measurements which reduces the potential for online use, whereas iPLS selected variables are compatible with faster and simple 2D measurements, which are more suitable for online analysis. Thus, one could collect a selection of 3 or 6 2D spectra (λ_{ex}=244, 288, 300 nm and 244, 248, 258, 272, 284, 306 nm for PEG-LZ$_{\%}$ and PPR respectively) instead of the full EEM and obtain comparable results.

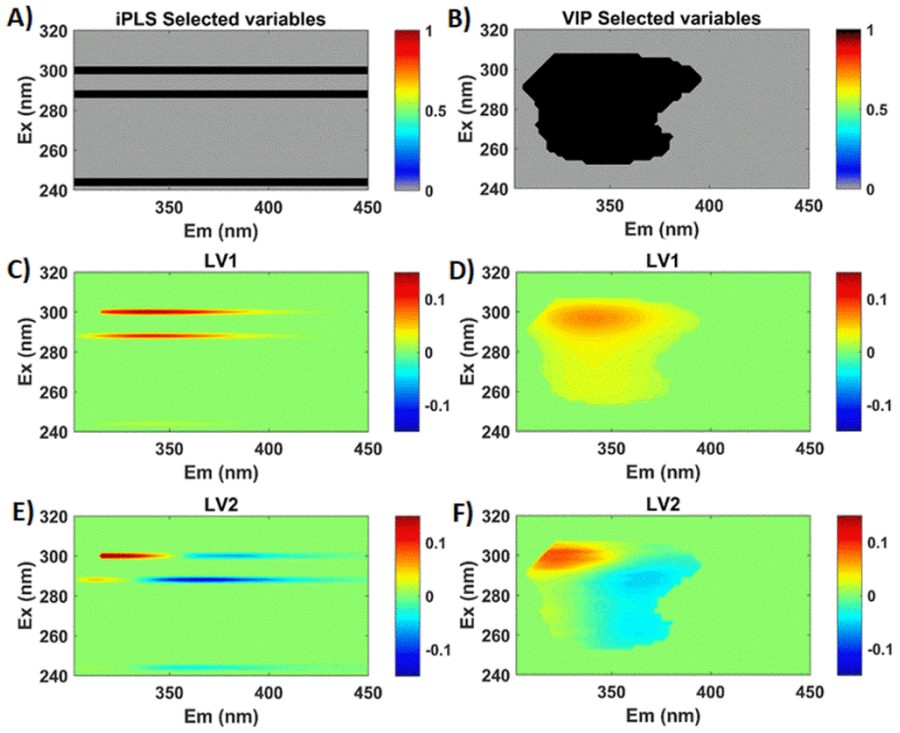


Figure 3.16- (A) iPLS and (B) VIP selected spectral variables (highlighted in black) from EEM$_{\parallel}$ data used for the prediction of: PEG-LZ$_{\%}$. (Second row) First (LV1) and (Third row) second (LV2) latent variables loadings (LVs) for PEG-LZ$_{\%}$ prediction using (C/E) iPLS and (D/F) VIP selected variables.

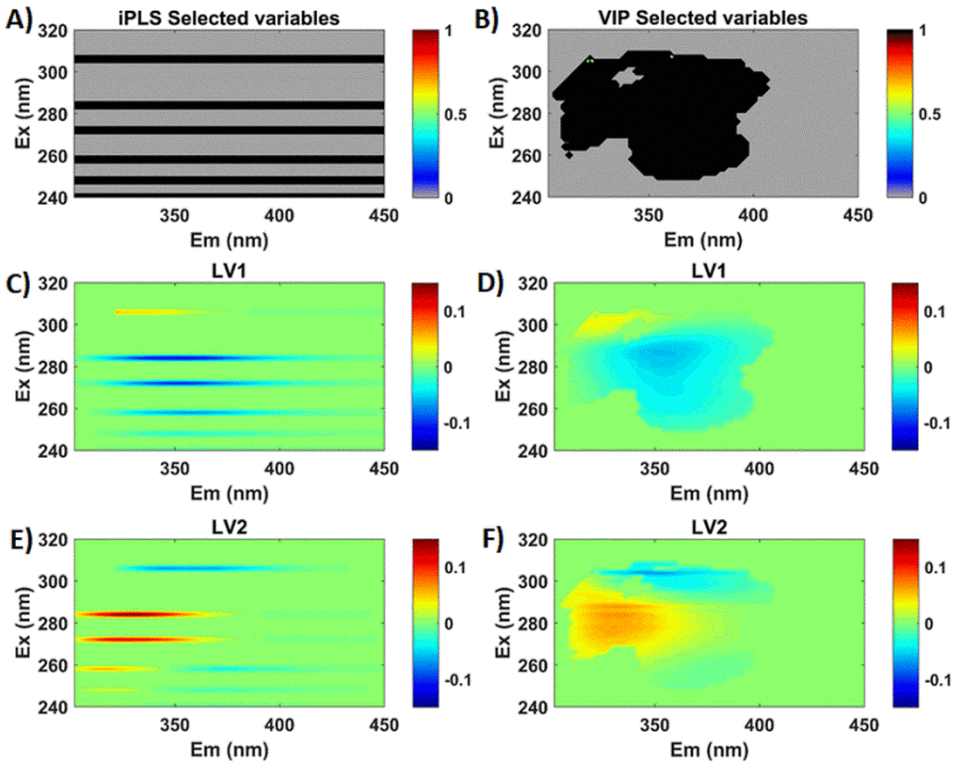


Figure 3.17- (A) iPLS and (B) VIP selected spectral variables (highlighted in black) from EEM‖ data used for PPR prediction. (Second row) First (LV1) and (Third row) second (LV2) latent variables loadings (LVs) for PPR prediction using (C/E) iPLS and (D/F) VIP selected variables.

Table 3.5- Statistical parameters of u-PLS models performance using selected datasets NoPrep-EEM⊥, NoPrep-EEM‖, for predicting PEG-LZ% and Norm-EEM⊥ and NormMC-EEM‖ for PPR. Sensitivity (slope of the calibration line) was 1 in all cases.

	PEG-LZ%						**PPR**					
Dataset	**No Prep-EEM⊥**			**No Prep-EEM‖**			**Norm-EEM⊥**			**NormMC-EEM‖**		
Var. selection	**-**	**VIP**	**iPLS**	**-**	**VIP**	**iPLS**	**-**	**VIP**	**iPLS**	**-**	**VIP**	**iPLS**
Selected Variables	**Ex: 240-320 Em: 302-450**	**Ex: 250-308 Em: 306-394**	**3 Ex Em: 302-450**	**Ex: 240-320 Em: 302-450**	**Ex: 252-308 Em: 306-396**	**3 Ex Em: 302-450**	**Ex: 240-320 Em: 302-450**	**Ex: 250-310 Em: 306-400**	**4 Ex Em: 302-450**	**Ex: 240-320 Em: 302-450**	**Ex: 248-310 Em: 306-408**	**6 Ex Em: 302-450**
~Acquisition time (sec)	**410**	**170**	**30**	**410**	**160**	**30**	**410**	**180**	**40**	**410**	**200**	**60**
RMSE Cal	**9.9%**	9.9%	9.4%	**8.9%**	10.6%	4.3%	0.12	0.13	0.11	0.06	0.04	0.08
RMSE CV	**10.6%**	10.5%	9.8%	**10.7%**	11.3%	8.2%	0.19	0.20	0.16	0.16	0.10	0.18
RMSE Pred	**11.1%**	10.98%	11.3%	**10.7%**	13.7%	12.4%	0.21	0.23	0.24	0.22	0.27	0.24
REP (%) Accuracy	**17**	16	17	**16**	20	**18**	**17**	18	19	**17**	22	**19**
R^2 Cal	**0.92**	0.92	0.93	**0.93**	0.91	0.98	**0.98**	0.98	0.99	**1.00**	1.00	0.99
R^2 CV	**0.91**	0.91	0.92	**0.91**	0.89	0.94	**0.95**	0.95	0.97	**0.97**	0.99	0.96
R^2 Pred	**0.93**	0.94	0.94	**0.93**	0.88	0.91	**0.96**	0.96	0.93	**0.96**	0.90	0.94
LOD[a]	14.3%	14.3%	13.5%	12.9%	15.5%	6.1%	0.14	0.15	0.12	0.06	0.09	0.04

[a] Limit of detection (LOD) was defined as the ratio between 3.3 times the standard error of the intercept and slope of the calibration line.

3.9. Conclusions

We have evaluated in this first chapter the use of pEEM spectroscopy for reaction monitoring and characterizing different degrees of PEGylation with a model lysozyme protein.

Ideally one should be able to assess this reaction in real time, but here the characteristic yellowing of reaction mixtures caused by the nitro containing leaving group of pNPC-PEG represented a big challenge. This caused the quenching of fluorescence, which significantly decreased the signal, and thus the robustness of the measurement. One possibility to overcome this drawback is by collecting data at a higher speed, which would enable the averaging of multiple measurements, and thus improvements in data quality. It is important to highlight that, for reactions using an UV active/quencher-containing surrogate or leaving group, any changes in fluorescence with the reaction could partly be caused by secondary effects like IFE and quenching and not only based in changes in protein emission induced by structure or environment changes. However, both IFE[178] and quenching[179] have been explored as tools for assessing proteins and other systems. To the best of our knowledge this would not be a problem if using other types of active PEG reagents like NHS and aldehyde derivatives.

For the case of reaction monitoring, absorbance spectroscopy might offer an advantage compared to fluorescence if one wants to rely on the signal of the leaving group (~400 nm) for assessing the changes.[180] In this case, the correct set up should be chosen, considering the usually high molar absorptivity of this compound compared to proteins (e.g. use of smaller pathlength cuvette to avoid saturation of signal).

UV-Vis spectroscopy has been used for the characterization of many protein conjugates and determination of conjugation degree.[60] However, for the case of small molecules which do not absorb in the UV-Vis region (like most PEG reagents), the absorption based measurement would only be possible if conjugation induces significant changes in the absorption spectra of the parent protein, which, as we demonstrated here, is not the case for PEG-LZ.

pEEM measurements by being both sensitive to protein structure and size changes offers a unique and potentially rapid, non-destructive method for product characterisation and quantifying the degree of PEGylation. The RS offers an alternative to DLS for in-situ assessment of the formation of HMWS during the conjugation/purification process and provided size change information for the product solutions with moderate correlation with the PPR from SEC (r^2=0.72) and hydrodynamic radius from DLS (r^2=0.65). The big limitation of using single measurement RS is that it is yet not sufficiently accurate because of measurement variation. Thus, the data needs to be made more reproducible by both averaging multiple scans per sample and implementing a calibration step. Unfortunately, *aniso*-EEM is not sensitive enough to assess changes in this conjugation, even with a relatively big change in size as demonstrated by DLS. This is probably because the anisotropy values were too small and largely affected by noise.

In contrast, the emission component of the signal detected the physicochemical changes induced by multi PEG residue attachment. ROBPCA of pEEM data indicated that changes were due mostly to Trp/Tyr emission fluctuations caused by changes in Trp environment and/or varying FRET rates induced by structural changes, suggesting thus that one could build predictive models for PEGylation degree using pEEM.

Furthermore, we have shown that the use of variable selection allows for the reduction of the data to only the excitation wavelengths containing the essential spectral information, enabling even further decrease in the acquisition time and facilitating spectral averaging. This would enable the use of pEEM measurements as a viable PAT solution for in- or on-line process monitoring of protein or peptide conjugation reactions like PEGylation or Antibody Drug Conjugate manufacture.

Chapter 4. rIgG multi-attribute quality screening

The chemical modification of proteins will be sensitive to a variety of factors related to the raw materials. The key goal in this chapter was to determine whether pEEM measurements could provide a more accurate, easier, or faster assessment of protein (rIgG) variance than conventional SEC and simple UV-visible and 2D fluorescence measurements. This work was published in *Analytica Chimica Acta* in December 2020.[173]

The gross quality[a] of the starting and reagent materials is critical in determining final product composition, purity, and quality, irrespective of the strategy chosen for protein conjugation. The small molecule raw materials (buffers, conjugates, linkers, solvents) can usually be easily assessed using standard chemical analysis techniques like HPLC. However, the protein substrate, is a more complex challenge because purity is not the only consideration. For protein starting materials, there is a need to verify both the concentration and the structural status of the protein in solution before use, as an element of the Quality by Design (QbD) principles being adopted by the (bio)pharmaceutical sector. [181] Three of the critical variables that need to be assessed, and thus controlled are the real concentration of native (active) rIgG forms, protein conformation in solution (tertiary), and the aggregation profile (quaternary). As these factors can be affected by a wide variety of processes including storage conditions and times, and changes in these structures could influence the course of chemical reactions, one needs a quick, non-destructive, and effective method for assessing protein variability in solution. Another area in which IgG quality (concentration and structural assessment) needs rapid assessment involves the rehydration of lyophilized proteins. Lyophilisation offers many benefits for the transport and storage of therapeutic proteins,[182] however, the reconstitution step introduces a potential source of concentration and structural variability.[b] Other quality attributes such as charge and glycosylation require the use of ion exchange chromatography and mass spectrometry [183] for example, but these are not considered here.

[a] By gross quality we mean the quality status of the starting material in terms of concentration and structural composition (both tertiary and quaternary structure if the sample is a protein).

[b] These could be variations in final concentration caused by the use of wrong volumes of buffer/non-homogeneous re-hydration of powder, or changes in tertiary or even quaternary structure caused by shaking/ swirling during homogenization.

A polyclonal antibodies (pAb) was used here because it is relatively cheap, which is a critical factor when selecting a model system with which to determine the efficacy of a new measurement method. However, the use of pAb offers a challenge as it is comprised of a mixture of closely related structures, with different epitope recognition abilities, produced by a large number of B cell clones.[184] One might then expect that the polyclonal rIgG should show significant batch-to-batch variability (because they have been produced in different animals at different times) and to be less pure compared to mAbs.[184, 185]

4.1. Comparison of methods

The first step was to assess the raw data generated by the various techniques and evaluate ability of each technique to discriminate rIgG samples that were known to be different because of stress, different source lots, etc (total sample set, n=34).

SEC chromatograms (Figure 4.1A) showed three peaks with retention times (tR) of 9.28min (oligomers, peak 1), 10.05min (dimers/trimers, peak 2), and 11.36min (monomer, peak 3). The separation of polyclonal rIgG with its variable composition leads to relatively broad peaks, however, resolution values (R)[c] indicated a good separation between dimer/trimer and monomer peaks (R>2) despite the poor resolution between dimer/trimer and oligomer peaks (R<1.5).

Here we used the total area (AUC between tR = 8-14 min), and the areas under peak 3 (monomer) and combined area of peaks 1 and 2 (aggregates), to calculate the percentages of monomers and aggregates respectively, which is referred to here as the aggregation profile. SEC of rIgG samples (Figure 4.1A) measured in triplicate were consistent (indicating good reproducibility between injection, Table 4.1) and overlaid almost perfectly once normalized (Figure 4.1B). The small variances in the raw data can be attributed to small concentration variations between injections, stock solutions, and the buffer components.

[c] The resolution factor (R) was used to evaluate the accuracy of chromatographic separation with R> 1.5 indicating good separation between adjacent peaks: $R=2\frac{tR2-tR1}{W1+W2}$

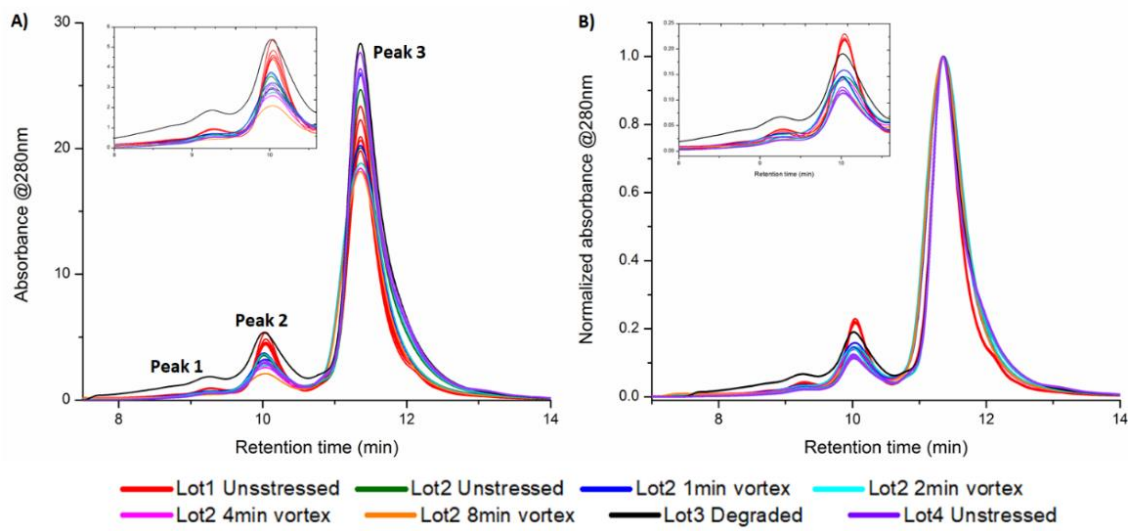


Figure 4.1- Unprocessed (A) and normalized (B) overlaid size exclusion chromatograms showing the concentration and % of oligomers as the main difference between rIgG solutions. Peaks 1-3 correspond to higher order aggregates, dimers/trimers and monomer, respectively.

Table 4.1- % Monomer (Mon), % Aggregates (Agg.), Mon./Agg. Ratio and % Loss[d] calculated using non-normalized SEC chromatograms of non-stressed (NS) and stressed rIgG solutions. The values in the table represent the mean ± stdev of three SEC runs of each stock solution (except for soln. 4 of which only two chromatograms were recorded).

Lot	Sample no.	rIgG Stock Solution	% Mon. area	% Agg. Area	Mon./Agg. Ratio	% Loss
Lot1	1/2/3	soln. 1 NS (n=3)	81.93±0.96	18.07±0.96	4.55± 0.29	--
	4/5/6	soln. 2 NS (n=3)	82.77± .56	17.93± .09	4.81± 0.19	--
	7/8/9	soln. NS 3 (n=3)	81.48± 0.63	18.52± 0.63	4.41± 0.18	--
	10/11/12	soln. 4 NS(n=3)	81.78±0.94	18.22±0.94	4.50±0.28	--
	13/14/15	soln. 5 NS (n=3)	81.54± 0.7	18.46±0.73	4.43±0.22	--
	--	Overall Lot 1(n=15)	81.86 ±0.93 (RSD 1.14%)	18.16 ±0.93 (RSD5.11%)	4.52±0.28 (RSD6.25%)	--
Lot 2	16	soln. 6 NS (n=1)	86.82±0.37	13.18±0.37	6.59±0.21	--
	17	soln.6 1min vortex (n=1)	87.03±0.64	12.97±0.64	6.73±0.37	--
	18	soln.6 2 min vortex(n=1)	86.58±0.73	13.42±0.73	6.47±0.40	--
	19/21/ 23/25	soln.7 NS (n=4)	87.21± 0.96	12.79 ± 0.96	6.84± 0.68	--
	20	soln.7 1min vortex (n=1)	85.66± 0.75	14.34± 0.75	5.99±0.38	1%
	22	soln.7 2min vortex (n=1)	86.27± 0.61	13.73± 0.61	6.30±0.32	8%
	24	soln.7 4min vortex (n=1)	87.81± 0.48	12.19± 0.48	7.22±0.32	10%
	26	soln.7 8min vortex (n=1)	90.91± 0.01	9.99± 1.14	9.14±1.09	14%
Lot 3	27/28	soln.8 (Poor storage) (n=2)	77.47±0.35	22.53±0.35	3.44±0.07	--
Lot4	29/30/31	soln.9 (n=3)	88.86±0.28	11.14±0.28	7.99±0.23	--
	32/33/34	soln.10 (n=3)	88.75±0.17	11.25±0.21	7.89±0.17	--

[d] % of loss was calculated comparing the total AUC of monomers peak before and after stress.

UV-Visible absorbance spectroscopy (Figure 4.2A/B) is an easy, fast and inexpensive method generally used for protein quantification. The rIgG absorption maxima is at 280 nm due largely to the aromatic amino acids. An increase in apparent absorbance at longer wavelengths (>300 nm) was also observed for some samples which is related to light scatter (*vide infra*). To increase sensitivity, one can use fluorescence spectroscopy since protein emission is very sensitive to structural and aggregation changes.[105] The simplest approach is to use a single excitation wavelength (i.e. 2D spectra), typically 280 nm. Figure 4.2 shows unpolarized[e] (C) and polarized (D) emission profiles at λ_{ex}= 280 nm. The spectra is characterized by two emission peaks, one strong main peak at λ_{MAX}~340 nm and the RS peak (λ_{MAX}=280 nm), which is increased for the parallel polarization. These two peaks were used to calculate Fl-AI values. (Equation 2.2)

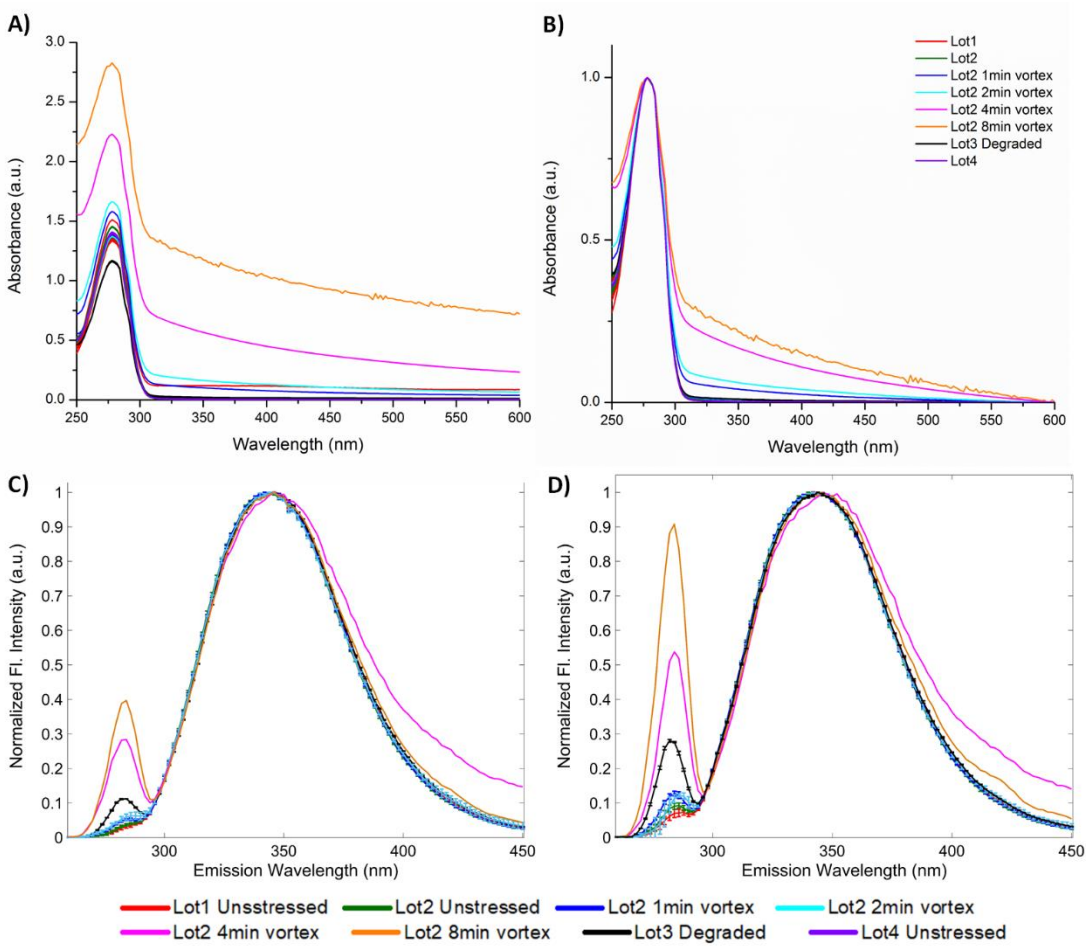


Figure 4.2 (A) Unprocessed and (B) normalized UV-Visible spectra show significant differences in light scattering for stressed samples; 2D-fluorescence spectra of rIgG measured at λ_{ex}=280 nm of normalized unpolarized (C) and parallel polarized (D) spectra.

[e] The unpolarized 2D spectra were calculated as per Equation 2.6.

Despite the higher sensitivity and specificity of intrinsic fluorescence, 2D measurements, may not be suitable for analysing complex proteins with many fluorophores like IgG.[67] MDF measurements, such as EEM are a more information rich alternative for subtle changes in these type of samples,[186] and its potential use for monitoring heat induced monoclonal antibody aggregation has been demonstrated.[141] rIgG has strong fluorescence emission over the $\lambda_{ex}/\lambda_{em}$ range: 270-300 / 300-380 nm, with the fluorescence maxima for all samples at $\lambda_{ex}/\lambda_{em}$ =292 / 344 nm which is mostly due to Trp emission. To assess and quantify the degree of rIgG variation we calculated the standard deviation (stdev) at every wavelength over whole EEM landscape for the different polarization settings ($EEM_{\parallel}$, $EEM_{\perp}$, and EEM_T), which shows the spectral regions with most variation (Figure 4.3). When considering all rIgG samples (n=34) the variability was highest in the region of directly excited Trp and also in the longer wavelength ($\lambda_{ex}/\lambda_{em}$>300/400 nm) region associated with "deep blue autofluorescence". [187, 188] Furthermore, this also indicated that $EEM_{\parallel}$ measurements were the most sensitive to sample variance and thus this data will be used for all further sample analysis.

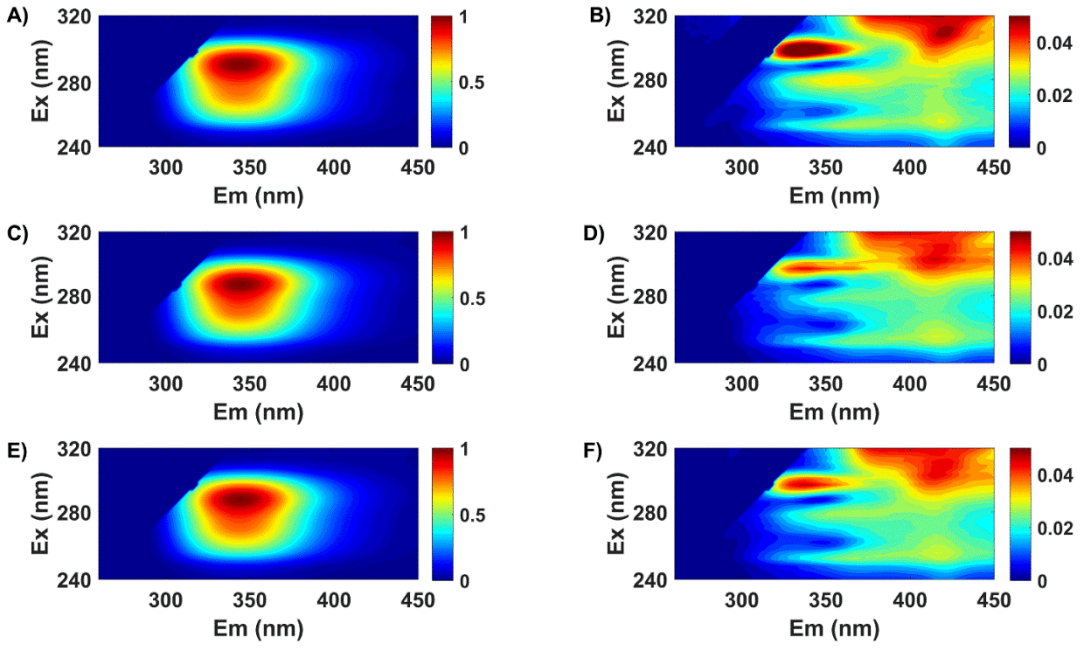


Figure 4.3- Mean (left) and stdev (right) plots calculated using $EEM_{\parallel}$ spectra (A/B), $EEM_{\perp}$ (C/D), and EEM_T (E/F) data from all rIgG samples (n=34).

The long wavelength emission plays an import role in discriminating lots with varying aggregation content (*vide infra*). The deep blue autofluorescence has been

associated with protein aggregation and fibrillation, however, others indicate that this can be induced by oxidation of Trp or Tyr, or is due to carbonyl emission. [187, 188] While the source of this emission is controversial, it is a real, but weak signal (typically 2-3% of the main band for unstressed samples), and also very sensitive to IFE.[178] A control study Figure 7.6 using a Trp solution confirmed that this variance was not due to instrumental or measurement factors. For the stressed samples, however, this long wavelength signal increased significantly to around 20% for the most mechanically stressed samples.(Figure 4.4) It is important to highlight that, if one needs to rely only on this spectral region, the data collection needs to be optimised to increase SNR. (This will be discussed in the conclusion)

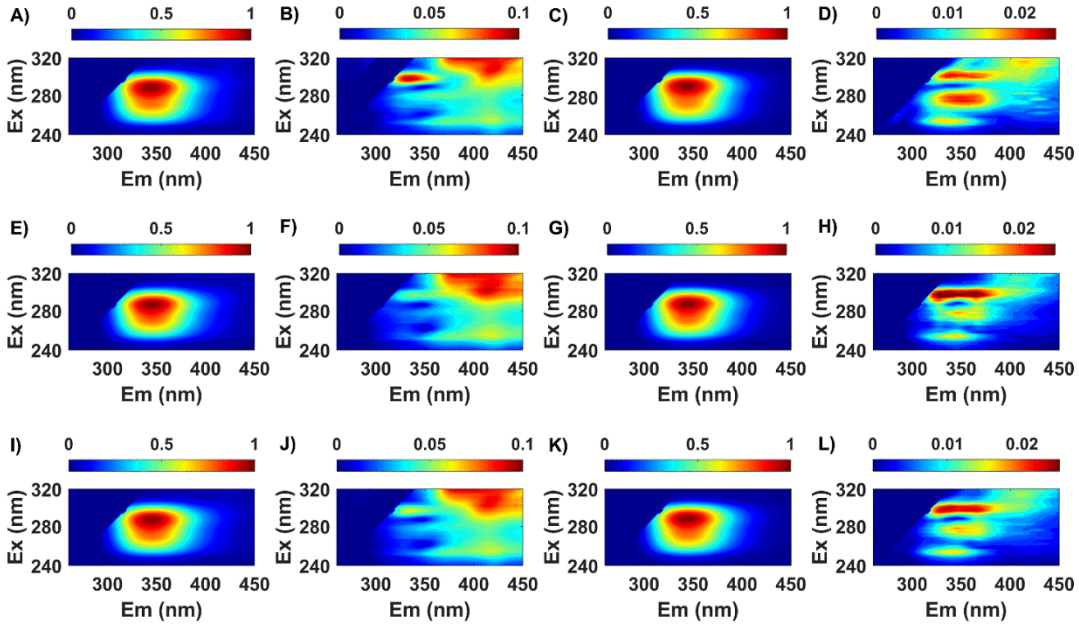


Figure 4.4 - Mean and standard deviation EEM-plots calculated for only stressed (A,B/ E,F/ I,J) and unstressed samples (C,D/ G,H/ K,L) using $EEM_{\parallel}$ (Top), $EEM_{\perp}$ (Middle), and EEM_{T} (Bottom) datasets

Another source of information from the EEM measurement is the Rayleigh scattered light. This is normally discarded during fluorescence analysis,[189] but can be very useful in providing information about changes in particle size distribution and particle concentration.[162] There is a complex relationship between Rayleigh (and Mie)[f]

[f] The light scattered by particles of much smaller size ($\alpha<2$) than the wavelength of incident light is usually described by the Rayleigh regime, whereas Mie scattering describes the scatter by particles of closer size (α= 2-10) to the wavelength of incident light.[22, 190] α is a size parameter defined as $\alpha = \pi D_P/\lambda$[175]

scattered light intensity and particle size distribution and concentration, which is also affected by refractive index, solution properties, and excitation wavelength.[175]

The Rayleigh scatter phenomena can be described by Equation 4.1, which relates the intensity of scattered light (*I*) and the diameter of the particle (D_P).

$$I = \frac{\pi^4 D_P^6}{8 r^2 \lambda^4} \times \left(\frac{m^2-1}{m^2+2}\right)^2 \times (1 + cos\theta^2)\, I_0 \qquad \text{Equation 4.1}^{175}$$

Where I_0 is the intensity of incident light at a distance *r*, θ is the direction of scatter from the particle, λ is the excitation wavelength, and *m* is the ratio between the refractive index of the medium and the refractive index of the absorbing material.[175]

Table 4.2- Simple fluorescence analysis of non-stressed (NS) and stressed rIgG solutions. The values represent the mean ± std of the replicates (except for soln. 6 samples and the mechanically stressed samples from Soln.7).

Lot	rIgG Stock Solution	UV-AI (%)	Fl-AI∥ (%)	RS_{296}/Max Intensity	RS_{316}/Max Intensity
Lot1	soln. 1 NS (n=3)	0.26±0.10	3.50 ± 0.70	1.49±0.02	4.49±0.08
	soln. 2 NS (n=3)	0.54±0.04	5.25±1.01	1.50±0.09	4.33±0.27
	soln. NS 3 (n=3)	0.20±0.03	5.82±1.22	1.60±0.12	4.66±0.40
	soln. 4 NS(n=3)	0.62±0.04	5.75±1.00	1.62±0.12	4.84±0.37
	soln. 5 NS (n=3)	0.85±0.27	6.19±0.71	1.61±0.09	4.71±0.32
	Overall Lot 1(n=15)	0.51±0.27 (RSD 53.19%)	0.51±0.27 (RSD 53.19%)	5.30±1.27 (RSD 23.94%)	1.56±0.10 (RSD 6.38%)
Lot 2	soln. 6 NS (n=1)	0.14	8.43	1.47	3.80
	soln.6 1min vortex (n=1)	0.37	9.95	1.65	4.45
	soln.6 2 min vortex(n=1)	0.86	10.93	1.65	4.30
	soln.7 NS (n=4)	0.50±0.09	6.12±0.35	1.59±0.06	4.46±0.23
	soln.7 1min vortex (n=1)	6.86	10.93	2.06	5.87
	soln.7 2min vortex (n=1)	10.91	3.11	1.66	7.22
	soln.7 4min vortex (n=1)	34.50	42.11	4.46	14.25
	soln.7 8min vortex (n=1)	72.27	79.63	9.07	16.95
Lot 3	soln.8 (Poor storage) (n=2)	1.58±0.53	27.25±1.11	2.47±0.06	6.19±0.17
Lot4	soln.9 (n=3)	0.32±0.07	6.75±0.48	1.44±0.05	3.99±0.15
	soln.10 (n=3)	0.36±0.07	6.11±0.13	1.42±0.01	3.85±0.05

The RS was, as expected, strongest in EEM∥ because of the polarized nature of scattered light and this was used for analysis. Absolute intensities are less robust in terms of reproducibility, thus we used here the ratio between RS at two λ_{ex} (296 nm and 316 nm) and fluorescence intensity maxima (I_{RS}/I_{Fmax}) to identify HMWS formation in solution. The variation in I_{RS}/I_{Fmax} ratio across the full sample set (Table 4.2) was larger at λ_{ex} =296 nm (RSD 73.86%) than 316 nm (51.93%), whereas for the

unstressed samples the values were 6.53 and 9.18% respectively. This was likely due to a combination of factors including changing particle size distribution (increased RSD) and resonant light scatter effects (difference between 296 and 316 nm). The RS peak extracted from the EEM‖ maps (Figure 4.5) by PARAFAC modelling provides a better measurement as it captures more of the particle/size distribution induced changes than a single wavelength.

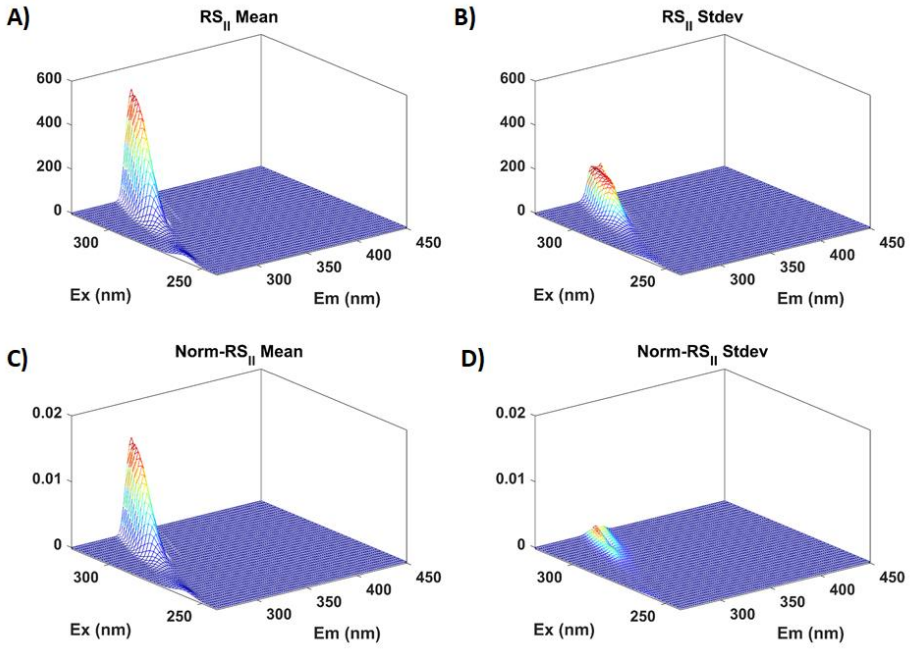


Figure 4.5- Plots showing the mean and stdev plots for the unprocessed (A,B) and normalized (C,D) RS spectra extracted from the EEM‖ measurements. .

4.2. Physical Homogeneity

When assessing the physical homogeneity for a single protein containing sample, one has to consider both size and the fold state (i.e. tertiary and quaternary structure), as well as information about reversible and non-reversible aggregates in solution. Some of the critical factors are the lot-to-lot and aliquot-to-aliquot (reproducibility) variance.

4.2.1. Aggregation profile

SEC is the pharmacopeial method for the characterization of reversible self-associated or non-reversible soluble HMWS in monoclonal antibodies.[191, 192] Thus, SEC analysis was used here as the reference technique for assessing the degree of

aggregation of samples from three different bulk lots (Lot1,2,4, n=26). The % of aggregates area calculated from the chromatograms indicated a different aggregation content for each lot (18.23±0.90, 12.98±0.68, and 11.20±0.31%) while the degraded samples (Soln.8) had the highest concentration of soluble dimers/trimers and higher order aggregates (22.5%). For the mechanically stressed samples (Soln.7) there was an increase in SEC detectable aggregates for the longer (>2 minutes) stress times, but this was also accompanied by very significant protein loss (by A_{280}) of 8, 10, and 14% for 2, 4, and 8 minutes of stress respectively, presumably due to either insoluble aggregate formation or adherence to the container walls. Smaller changes were noticed when Lot2 samples (soln.6) were stressed using smaller vial headspace, which is known to impact on the degree of structure disruption. [193]

Interestingly, the data from the simple spectroscopic methods did not correlate well with the SEC results. (Figure 4.6) The UV-AI data was very ambiguous showing a value of 0.47 ± 0.22% (unstressed Lots 1, 2, and 4, n=26), which suggested a very low aggregate content in each sample (<1%). Also, the RSD across these 26 samples was higher for UV-AI[g] (> 45%) than for the SEC[h] measurements (22%). The problem with the UV-AI measurement is the fact that it is based on an absorbance measurement at 350 nm which is very small (values are given below) for unstressed samples and may bias the results. For instance, the absorbance values obtained for the fifteen Lot1 IgG samples at this wavelength (A_{350}) was 0.014 ± 0.030 a.u., which has a high standard deviation and mean value lower than the limit of quantification [i] calculated using 10 blank solutions (LOQ=0.020 a.u.). In summary, the use of UV-AI is not suitable for QA/QC screening of protein starting materials because it is insensitive to soluble aggregates and typically only used to detect the presence of large particles (with a hydrodynamic radius greater than 200 nm) in protein preparations. [67]

Fl-$AI_{\parallel}$[j] values suggested a lower aggregation content, but these also did not correlate with the SEC measurements despite having a more similar RSD (21%) across the sample set. (Figure 4.6) A possible reason for this is that the level of aggregation

[g] All measured samples from Lot1, 2, and 4 were used for AI RSD calculation.
[h] All measured samples from Lot1, 2, and 4 (including measurement replicates) were used for SEC RSD calculation.
[i] LOQ (Limit of quantification) = 10 × stdev calculated using 10 blank solutions
[j] Fl-AI calculated from parallel polarized data were determined to be the most sensitive to particle content compared to other polarizations. (Details on Section 7.7.2)

produced here may not induce significant changes in the spectral profile of large molecules like IgG, particularly when the assessment is based in only two wavelengths as in the case of Fl-AI.

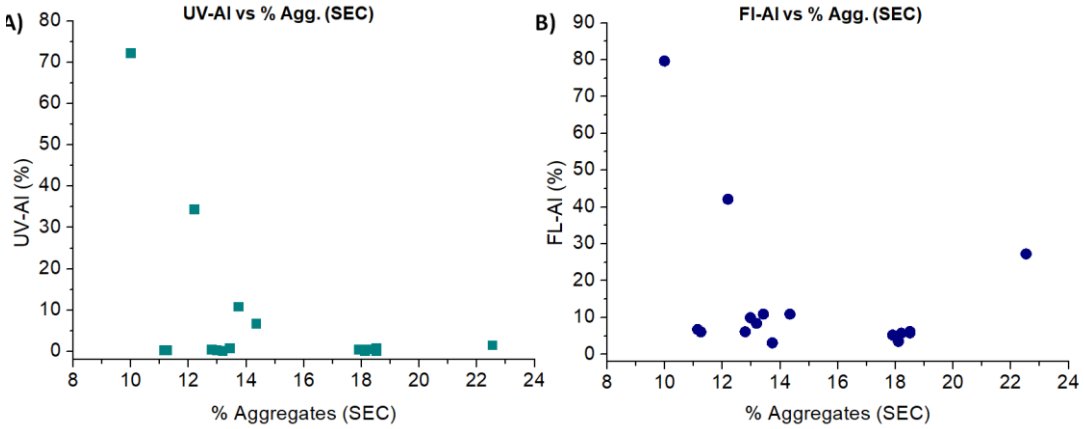


Figure 4.6- Scatter plots showing relationship between aggregation indices (A) UV-AI and (B) Fl-AI and the % Aggregates as measured by SEC for all samples (unstressed and stressed).

The next step was to use ROBPCA to try and extract more information from the absorption spectra and see if this would produce a better correlation with SEC measurements. However, since there was no significant change in the spectra apart from the increased background due to scatter (Figure 4.7 C/D), ROBPCA could not discriminate the samples according to their aggregate content as calculated using SEC. pEEM data provided greater spectral information (i.e. about size and structure changes) compared to UV-Vis, and the PC1-PC2 scores plot from ROBPCA modelling showed clear sample clustering according to SEC determined aggregation content. ROBPC1/2 loadings were refolded to produce EEM plots for clearer visualization of the spectral features which contributed the most to discrimination. (Figure 4.7 E/F) For PC1, the loadings indicate a negative contribution which corresponds mostly to Trp and the deep blue autofluorescence region, whereas the positive part of the loadings plots is related to tyrosine emission (λ_{MAX} ~320 nm) and possibly phenylalanine absorption (λ_{MAX} ~260 nm). This suggests that the spectral change which correlates with aggregation profile is FRET related. ROBPC2, on the other hand, is more likely to be related to sample preparation issues and small changes in concentration, which will be discussed in the next section. (Table 7.8 provides a summary of ROBPCA results)

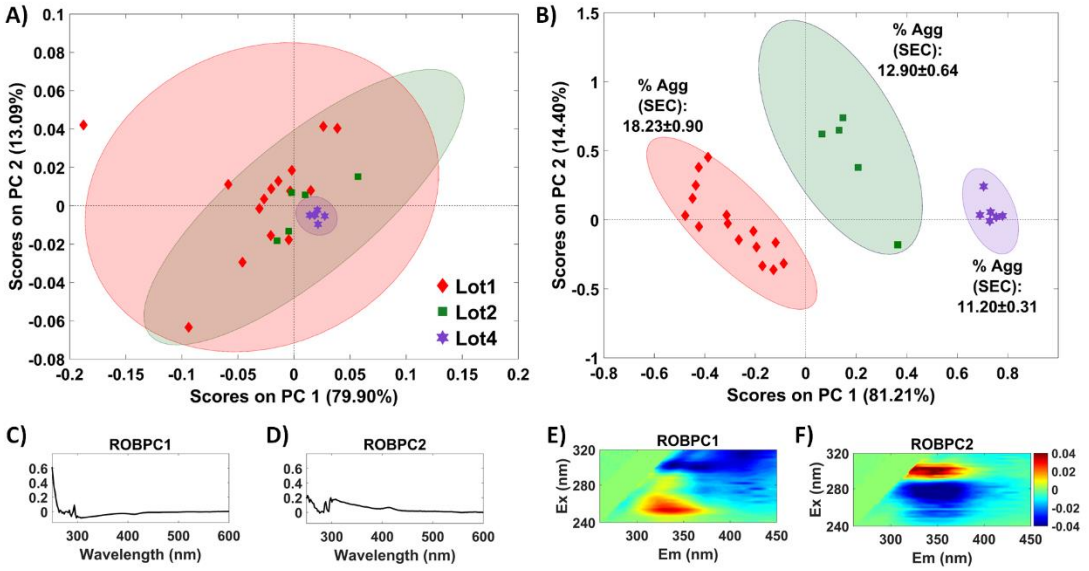


Figure 4.7- Scores plots for lot-to-lot rIgG discrimination analysis by: (A) UV-visible absorbance spectroscopy, and (B) EEM‖ measurements. Ellipses are the 95% confidence interval for each cluster. ROBPC1 and ROBPC2 loadings plots showing regions of most significant spectral variance in UV-Vis (C, D) and EEM‖ (E,F) data.

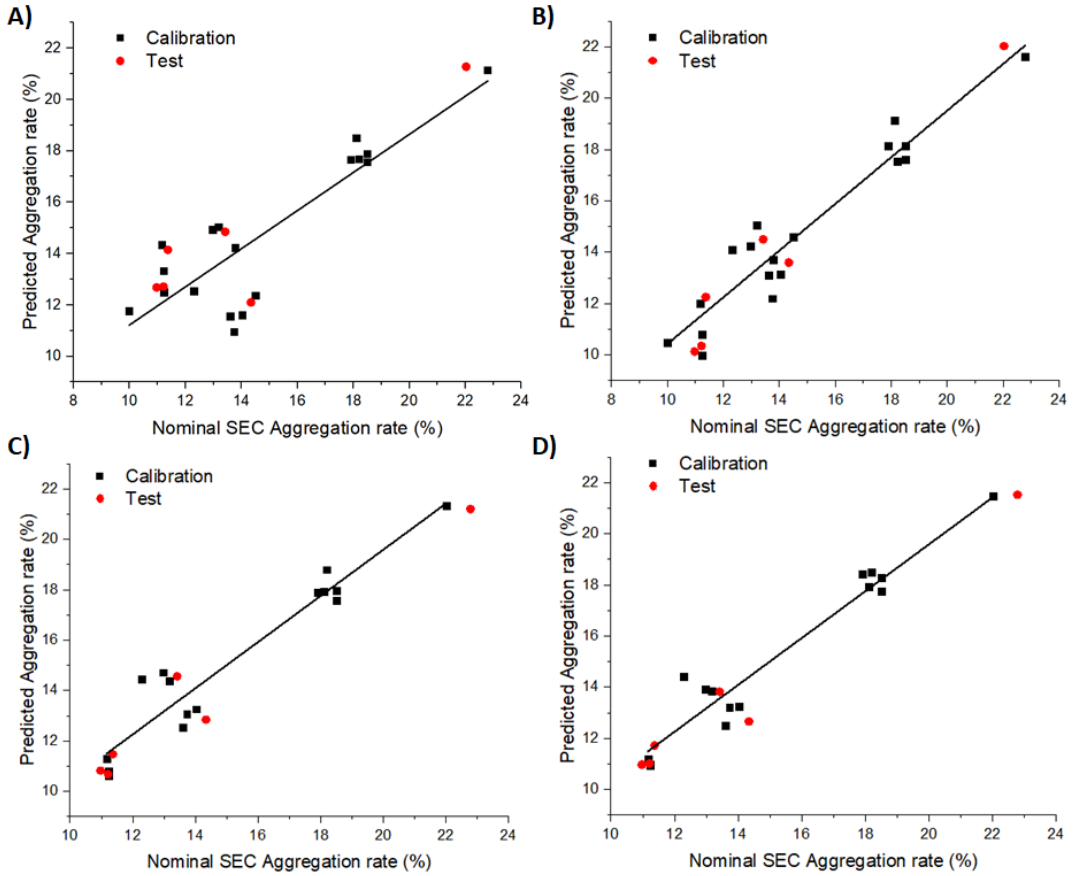


Figure 4.8- Plots of nominal versus predicted aggregate content (% area from SEC) from u-PLS modelling of all samples without (A) and with variable selection using VIP (B); and models using reduced sample set (excludes the 2-8 mechanically stressed samples) without (C) and with variable selection using VIP (D).

Based on these results we built quantitative u-PLS models for predicting SEC aggregation content from EEM$_{\parallel}$ measurements (Figure 4.8, Section 7.7.3), which suggested that a quantitative correlation can be obtained with reasonable accuracy . It should be noted that the model cross validation improves significantly when the 2-8 minute mechanically stressed samples were excluded from the calibration set which suggests that a different type of aggregate was formed. (Table 7.6 and Table 7.7) Mechanically stressed IgG is thought to form aggregates by displacement from the air-water interface [194] which is a different mechanism compared to stress caused by improper storage. The big problem here is that the spread of aggregate content values (as determined by SEC) in this dataset is small (10-24%) with only four groups of values. Further investigations are required to validate and ideally this should be done using a more representative industrial protein sample set where production and orthogonal analytical data is available. However, these proof of concept models demonstrate that it could be feasible to predict the SEC derived aggregation profile from EEM$_{\parallel}$ measurements.

4.2.2. Reproducibility (Intra lot variability):

Sample preparation is critical in protein analysis, and the complex structure and behaviour of these macromolecules does cause some issues. Here the changes are expected to be more subtle than for the different aggregation profiles, and mostly related to concentration, conformational changes and/or low reversible oligomers content caused by uncontrolled variations in sample preparation. For instance, the composition of Lot 1 (stock solutions 1-5) from SEC was very consistent (81.86 ± 0.91 %) with low variability (RSD < 2%) (Table 4.1) The UV-AI and Fl-AI$_{\parallel}$ parameters however indicated a significant variation among these samples (RSD= 53% and 24% respectively) which suggested that the solution tertiary composition was different to that of the SEC analysis. ROBPCA of EEM$_{\parallel}$ required two PCs for describing the variation as expected and the scores plot (Figure 4.9 and Table 7.8) highlighted two trends. The separation along the first PC correlates with the stock solution concentrations (Table 2.3) whereas PC2 does not correlate with any other measurements and thus may be related to some form of sample handling error that we have not been able to identify.

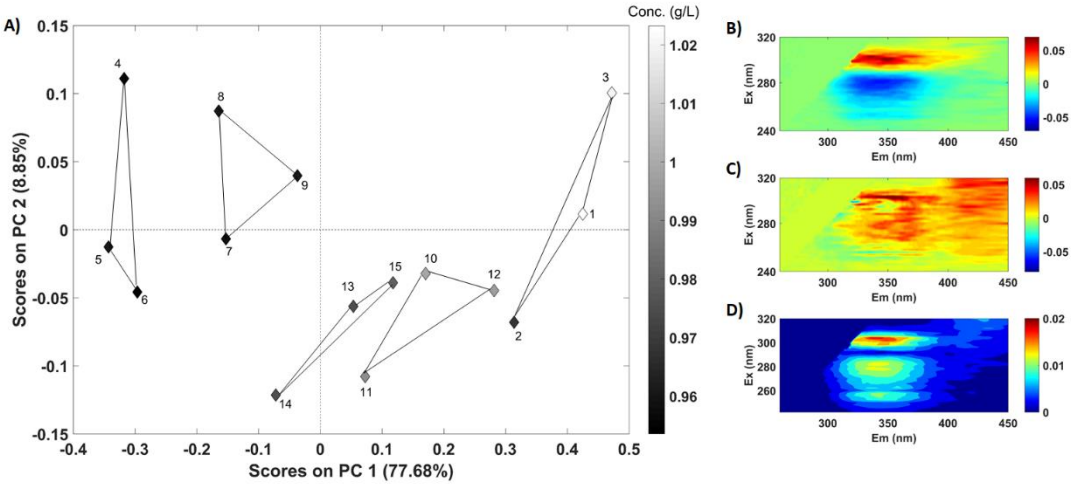


Figure 4.9- (A**)** ROBPC1 versus ROBPC2 scores plot of EEM_I data with samples numbers as per Table 4.1 and the points connected by lines are aliquots from the same stock solution. B and C are the refolded loadings plot (ROBPC1 and 2 respectively) showing the major areas of spectral variability in the rIgG samples from the same bulk lot. Plot (D) is the standard deviation plot calculated for all Lot1 samples.(Compare to the mean and stdev plots of the control Trp solutions, Figure 7.6 and Figure 7.7).

This variation despite being significant in the scores plot, represents a relatively small fraction of the spectral variance (<9%) and the spectral standard deviation calculated for these samples was only twice that of the experimental variability (Figure 7.6). Still, we cannot yet fully exclude the possibility that the PC2 variance might be attributable to the formation of reversible oligomers in solution due to different buffer incubation times and/or handling, however, we do not have measurement methods of sufficient sensitivity to accurately measure these species in-situ. Thus, we can consider this plot to graphically represent the minimum measurement error achievable with the pEEM measurement method for this sample type.

4.3. Protein Concentration

The determination of concentration is another IgG critical quality attribute and the use of absorbance at 280 nm (A_{280}) is the simplest and commonest, non-destructive method.[195] Colorimetric methods such as Bradford assay are alternatives but do require the addition of reagents to the sample and thus are destructive to a certain degree. The use of IPF for *in-situ* protein quantification can be a challenge because of the potentially high working concentrations (>1 g/L) and thus significant IFE effects which will lead to non-linear dependence of fluorescence intensity in relation to the concentration of the protein sample and might limit a simple intensity-based quantification assay.[105, 117] Here, however, one can use normalized pEEM spectra with

u-PLS regression to estimate protein concentration only by considering the changes in EEM shape profile.

For this, samples were split into calibration and test sets using the Kennard-stone algorithm [176] and the concentrations calculated from UV-Vis measurements were used as the nominal values. Four LVs were required for explaining most of the variance in the dataset and resulted in a small $RMSE_{CV}$ (0.02 g/L) which was 2% of the average rIgG concentration and equivalent to the error in the nominal concentration values. The model produced good results for both calibration and prediction sets (R^2_{Cal} =0.94 and R^2_{Pred}=0.96) and small errors (RMSE of calibration and prediction were 1%). The loading plots (Figure 4.10 C/D) highlight the spectral regions with most variation. LV1 represents Tyr excitation but Trp emission (i.e. FRET) and LV2 direct excitation of Trp. The EJCR showed (Figure 4.10B) that the ideal point (1,0) for the slope and intercept was within the ellipse, indicating the absence of bias within the 95 % confidence level.[165, 196] The ellipse was narrow indicating good precision, but it was not centred about the ideal point, which is probably because of the limited concentration range used.

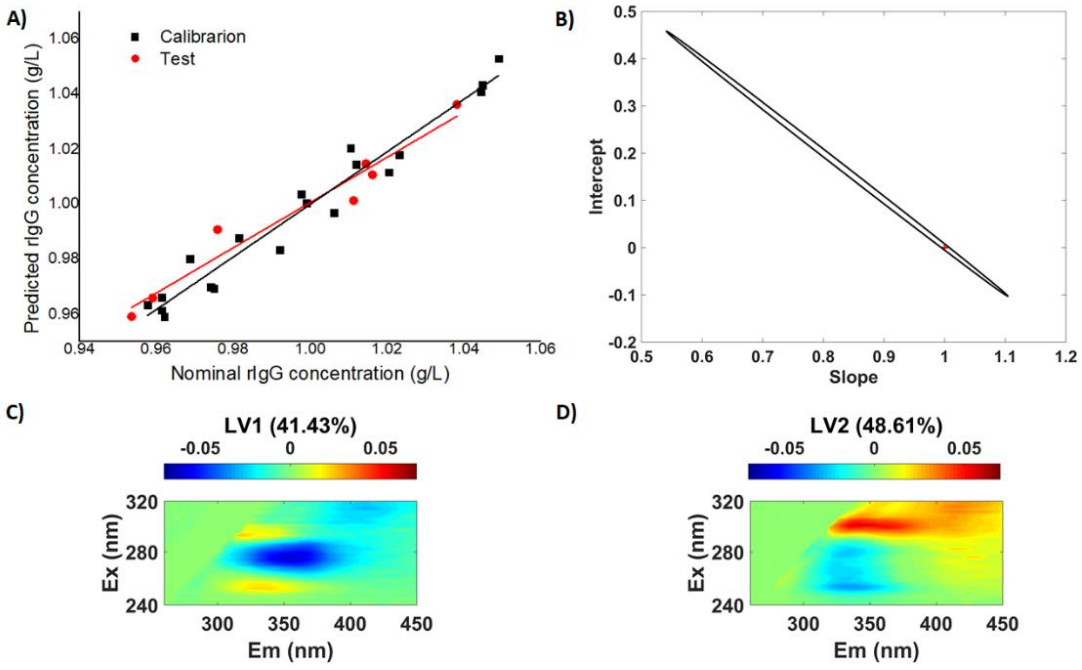


Figure 4.10- Results of the u-PLS regression of EEM used for quantitative analysis: (A) Predicted versus nominal concentration values from UV-Vis for calibration and test set. (B) EJCR plot at 95% confidence level for the regression slope and intercept of predicted versus nominal protein concentrations containing the ideal point (slope=1, intercept=0). Contour plots of reshaped latent variables 1 (C) and 2 (D) loadings.

4.4. Sample Screening

The goal here was to evaluate which of the spectroscopic measurements provided the quickest and most effective screening method for assessing protein variability for routine QC applications. For this, we used the full sample set (n=33) which included 26 unstressed, two storage stressed, and six mechanically stressed samples. The Hotelling T^2 vs. Q residuals plots (also referred to as outliers plot) is an easy method for visualising the performance of the different spectroscopic methods as a screening tool. In these plots the upper right quadrant, should contain the outliers of most significance in terms of compositional differences.[197]

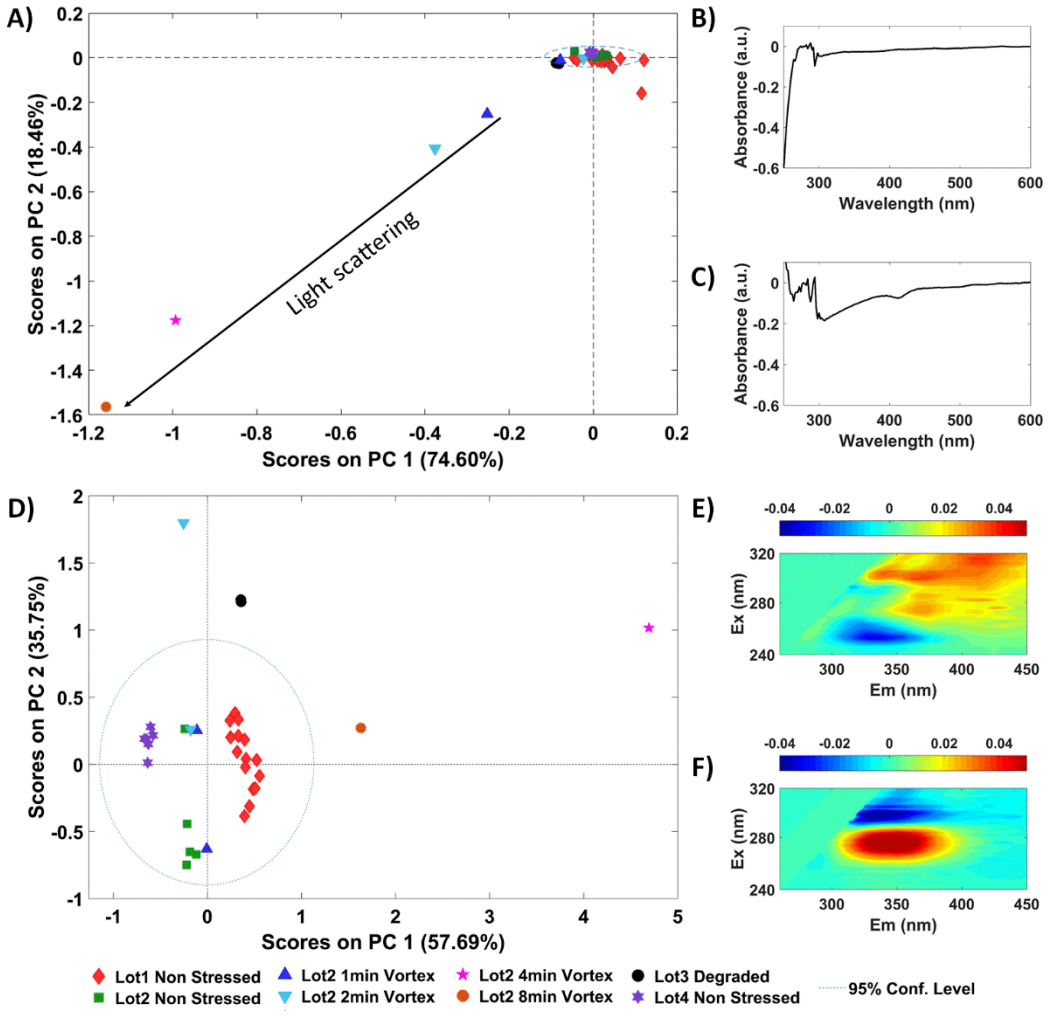


Figure 4.11- ROBPCA of UV-Vis spectra (A) only discriminated samples with increased scattering (larger particles as shown by ROBPC1 (B) and ROBPC2 (C) loadings plot. ROBPCA of EEM‖ (D) identified structural changes due to mechanical stress or degradation which can be seen as changes in emission according to ROBPC1 (E) and 2 (F) loadings plot. Samples are numbered as per Table 4.1

ROBPCA of the absorbance spectra (2 PCs explaining 93.06% of the variance, Table 7.8), easily differentiated some mechanically stressed samples as outliers in Figure 4.12A. As shown by Figure 4.11 B/C, this was a result of increased light scatter at longer wavelengths caused by the increases in particle size/insoluble aggregates and, to a loss of light intensity in the 250-270 nm region due to the increased light scatter. When using the absorbance spectra, the Lot3 storage degraded samples were not identified as outliers because they appear on the confidence interval ellipse boundary in the scores plot (Figure 4.11A). This was possibly because, compared to the other samples, these did not have many insoluble particles (and higher light scattering), but only a higher percentage of soluble oligomers.[k]

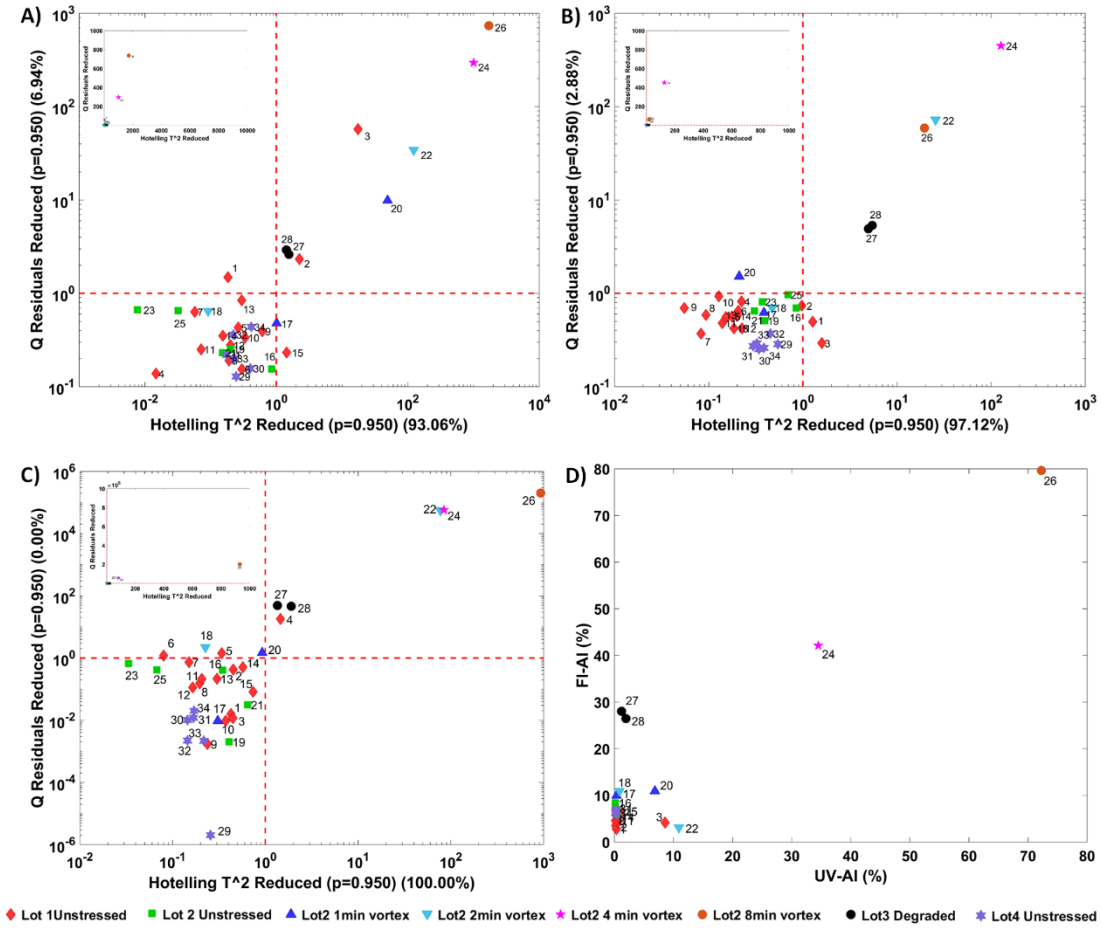


Figure 4.12- ROBPCA outlier diagnostic plot for: (A) UV-Vis data, (B) EEM‖, and (C) RS‖ and sample distribution based on Fl-AI and UV-AI values (D). Plots A-C were plotted using log scale to facilitate outlier visualisation. The sample numbering is provided in Table 4.1.

[k] Unfortunately, the DLS system was not available when this work was done, thus we were unable to validate this hypothesis.

ROBPCA of the normalized pEEM, required four PCs to explain 97% of the total variance and identified the storage degraded and mechanically (4-8 minute) stressed samples as outliers (Figure 4.11D / Figure 4.12A). The spectral regions with significant changes correspond to Tyr, Trp, and the deep blue autofluorescence regions and possibly spectral changes related to FRET.

Based on the SEC data and the stresses applied, one would expect between five (#28, 27, 26, 24, 22) and six (#20) significant outliers. The ROBPCA models built using the absorbance data identified 8 major outliers whereas the models using EEM and RS∥ yielded 5 and 6 respectively. A scatter plot built using the simple univariate indices UV-AI and Flu-AI (Figure 4.12) clearly identifies 4 and 8 minutes mechanically stressed samples (#24 and 26 which), and one can visually separate another five (#22, 3, 20, 28, 27). It is clear that relying on a single measurement to determine outlier protein samples is not perfectly reliable and it seems that measurements involving scatter are the most sensitive, generating a high rate of false positives, for example, the good Lot 1 samples #2, 3, and 4 (Figure 4.12C). The RS∥ spectra used here (Figure 4.5), only covered a relatively small spectral range and thus probably does not accurately capture the physical changes. Furthermore, the data collection methodology used in the fluorescence measurements (single scan per sample) is not ideal and is usually linked with significant noise artefacts. For instance, the RS∥ spectra from the Lot1 samples had an RSD for the area of 6.3 %, which could possibly be reduced by averaging multiple scans, thus leading to more accurate outlier identification.

A closer look at the ROBPCA of the stressed samples shows the spectral regions most influenced by degradation/aggregation processes. For instance, samples from the same stock solution (soln.7) were mechanically stressed for 1, 2, and 4 minutes respectively (samples 3-5 in Figure 4.13) and showed a progressive increase in ROBPC1 scores which was mostly related to the emission at $\lambda_{ex}/\lambda_{em}$> 300/400 nm. Sample 6 from the same soln.7 was stressed for 8 minutes deviated from this trend and showed a smaller contribution of ROBPC1 compared to the 8min stressed sample.

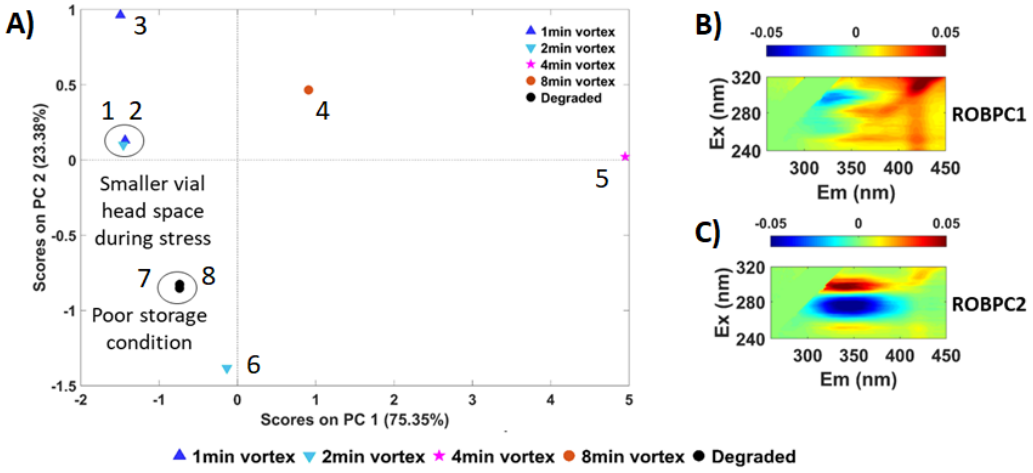


Figure 4.13- (A) ROBPCA1 and 2 scores and(B,C) refolded loadings plots of samples stressed for different lengths of time. Samples 1,2 (from soln.6) were stressed for 1 and 2 min, samples 3, 4, 5, 6 (from soln. 7) were stressed for 1,2,4 and 8 min and samples 1 and 8 were stored under poorly controlled conditions.

Samples 1 and 2 (1 and 2 minutes stress respectively from soln.6) do not seem to follow the same gradual trend and this behaviour could be explained by the use of a vial with smaller headspace during mechanical stressing, which is known to affect aggregation process. The type of stress also plays a role in aggregation path, structure and rate and this could be the reason for the differences of the degraded samples 7 and 8 in the scores plot, as these samples seem to be more influenced by changes in Trp emission (Figure 4.13C).

4.5. Conclusions

Here, we compare the efficacy of several simple spectroscopic methods, and SEC with pEEM spectroscopy for variance analysis of rIgG solutions from multiple batches of a commercially supplied rabbit IgG model protein. In contrast to a previous study using EEM and chemometrics[141] for assessing protein aggregation, here our goal was to evaluate the potential of pEEM to assess more subtle changes and also analyse the fluorescence signal and RS separately, as independent sources of information about sample quality. Also, this is a completely different approach to what has been published in the group.[198, 199] For IgG type proteins, PARAFAC analysis of either pEEM or polarized TSFS data is rather complicated and does not, as yet offer a robust approach to routine protein variance analysis. The key reasons are that the large numbers of fluorophores in IgG result in extensive FRET and the presence of residual light scatter both hinder component resolution.

Despite being the standard method for protein aggregation measurements, SEC analysis has limitations that affect protein analysis, and in particular that of reversible oligomers/aggregates in solution. The SEC sample preparation (e.g. dilution and mobile phase composition) and chromatographic separation process often lead to the disassociation of these reversible, non-covalently bound, aggregates.[200] All of these factors may lead to an aggregation profile which may not be representative of the situation in the protein stock solution. This is a critical advantage for spectroscopic based methods which can probe, non-destructively the protein solution without perturbing the balance between reversible and non-reversible aggregates.[201]

This was the case here where there was no good correlation between the UV-AI and Fl-AI measurements and the SEC results, despite these simple spectroscopic measurements clearly indicating the presence of varying amounts of aggregates in solution. Absorbance measurements and UV-AI, although fast, easy, and non-destructive, has poor detection limits and low selectivity. Fl-AI has better sensitivity and reproducibility making it a useful qualitative measure of aggregation,[67] however, this measurement does not provide much conclusive information about protein structure changes, particularly when these were relatively small.

The pEEM measurement in combination with conventional chemometric data analysis can be considered to provide a more comprehensive source of information about protein quality in solution. Here we have shown that it can provide information about concentration, gross structure variation, more subtle structure changes, and aggregate/particle formation (from the RS) in a single measurement. The use of normalized data makes the method a robust screening method which minimises the effects of lamp intensity variation, allowing its use as a rapid screening method to select samples for more detailed characterisation by a more time-consuming SEC reference method. It should be highlighted that, if one wants to rely on the long wavelength region for assessing changes, a better measurement set-up to improve SNR should be considered (e.g. use of a faster detector like cooled charge-coupled-CCD based spectrometer), as well as a more detailed calibration study to better understand the source of emission in that region of the EEM.

This proof-of concept study also shows that it could be possible to produce quantitative predictive models for the non-reversible aggregate content. In particular,

the sensitivity of the pEEM method to protein structural changes makes it very suitable for measuring differences compared to a reference batch. It should also be noted that the polyclonal antibody case is considerably more complex than therapeutic mAbs. For mAbs sourced from biopharmaceutical manufacturing processes, we would expect that the reduction in protein species diversity should lead to better correlations with SEC data, potentially leading to the development of accurate quantitative models for both the total and non-reversible (i.e. that determined by SEC) aggregate content.

Chapter 5. Monitoring IgG PEGylation reactions and products

The main goal of this chapter was to assess the ability of pEEM for monitoring conjugation reactions with IgG using a two-step PEGylation of a polyclonal rIgG as a model system.[a] This model can be considered as a "worst case scenario" because the PEG molecule is spectroscopically inactive from a practical sense[b] and the use of a pAb starting material makes it more technically challenging to characterize using spectroscopic methods, as discussed in the previous chapter. Also, conjugation of small molecules would likely cause only subtle changes in a big and complex molecule like IgG (in comparison to PEGylation of lysozyme, for instance). Finally, the chemical modification of an immunoglobulin should be a more therapeutically relevant model as there are a few IgG conjugates currently in the market (e.g. ADCs and radioimmunoconjugates),[45, 46] and the method, if successful here, could be potentially useful for assessing these other types of conjugation reactions. PEG is widely available commercially and used for many applications including the use as linkers for the production of ADCs.[56]

The four objectives of the spectral measurements and chemometric data analysis in this chapter were to:

1) Measure variance of the starting material, intermediate, and final protein product.
2) Follow progress of the reaction,
3) Accurately identify reaction endpoints,
4) Determine the benefits of using pEEM measurements instead of conventional absorbance and 2D fluorescence measurements.

Thus, the focus here was to show the potential of pEEM measurements for assessing conjugation reactions by observing changes in the parent protein, and not to extract reaction kinetics, or characterize final products.

[a] Ideally this model system should mimic a typical IgG conjugation and the small molecule attached should not be fluorescent to avoid any interference with IgG fluorescence because the objective was to assess changes in the protein induced by conjugation. Thus, PEG seemed to be an interesting choice here.

[b] It has neither a chromophore which is easily measured (i.e. strong absorption >250 nm) nor a fluorophore which would facilitate its measurement via absorption or emission spectroscopies.

5.1. The conjugation reaction

We can consider this model rIgG conjugation reaction as reasonably complex because of a few factors:

1) Use of a pAb starting materials;

2) Use of polydisperse PEG;

3) Six potential conjugation sites.

4) Low PEGylation level because of the mild reaction conditions used in the study.

For this study we also introduced some additional variation by using different source lots of starting materials [c] and concentrations of the reducing agent for the first step of the reaction (1.5 and 3.0 molar excess). (Table 7.2) The rationale of varying the concentration of the reducing agent was to produce different levels of conjugation as done for the PEG-LZ system and evaluate the effect on the pEEM. All these factors possibly contributed to the formation of a heterogeneous population of intermediates but final products with very subtle changes in the emission profile. IgG contains intra and interchain disulphide bonds with the inter-chain bridges being more susceptible to reduction because of their greater solvent exposure compared to the buried intra-chain bonds.[16] Thus, the most-likely reduction sites in rIgG are the disulphide bonds linking heavy-heavy chains and heavy-light chains, one in each antibody arm.[21, 202] (Figure 1.3) When reacting with low concentrations of mild reductants such as TCEP this potentially results in two-six thiol groups being available for conjugation with MAL-PEG via formation of a stable succinimidyl thioether between IgG and the PEG molecule.[148] (Figure 2.2) The conditions used here should result in partial rIgG reduction[203] and the formation of mostly di-PEGylated products.[204, 205]

5.2. Screening reaction

We first attempted to carry out the reaction at a slower rate (temperature of 4°C) to identify the points at which significant spectral changes occurred. For this, the reduction step took 14 hours and the alkylation 8 hours, with spectra collected every two. This indicated the most significant changes in pEEM at the latest data points

[c] Lot 1 (fourteen well-controlled reactions using unstressed starting materials, R1–R14), Lot 2 (three poorly-controlled reactions using mechanically stressed rIgG, R15–R17) and Lot 3 (two poorly-controlled reactions using a degraded lot of rIgG, R18–R19). More details in Chapter 2- Materials and Methods.

collected for each step (12-14 hours of reduction and 6-8 hours PEG-rIgG 8hours) compared to rIgG and lead to the decision of monitoring the reaction only at the three main stages: rIgG starting material (rIgG) before any reagent was added, reduction product (Red-rIgG) measurements were made 2:15 hours after TCEP addition, and final product mixture (PEG-rIgG), measurements were made 2 hours after MAL-PEG addition. This would also enable the simultaneous data collection from three reactions.

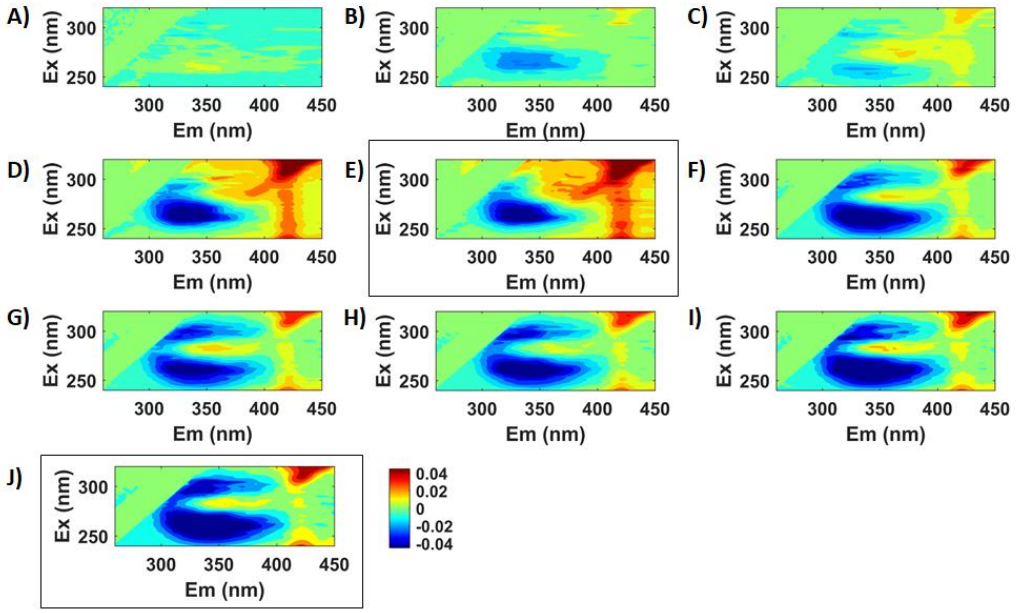


Figure 5.1- Difference spectra calculated using EEM‖ of the starting material minus EEM‖ collected at (A) 0, (B) 2, (C) 4, (D) 12, and (E) 14 hours of the reduction step and at (F) 0, (G) 2, (H) 4, (I) 6, and (J) 8 hours of the alkylation step. EEM‖ was shown here only as an example.

Based on the literature,[85, 148, 206, 207] partial reduction of mAb disulphide bonds is a fast reaction and should be completed within 2 hours at room temperature, which indicates that the data collected should be stable and representative of the reduction product. Similarly, we expect the alkylation to free sulfhydryls, to be completed after 2 hours[85, 148] of MAL-PEG addition, when the PEG-rIgG data was collected. We did not use a quenching solution to stop the reaction because we wanted to avoid any interference in the fluorescence measurements.

5.3. Reaction time course

In order to investigate the spectral changes over these reaction times, we carried out an independent experiment (in duplicate) using only EEM‖ measurements (which

we have determined to be the best for reaction monitoring). For this, EEM‖ spectra was collected every 8 minutes over the full reaction times (135 and 120 min for reduction and alkylation respectively). The spectra from the two reactions were then analysed (after pre-processing and normalization) using ROBPCA.

Figure 5.2 shows the ROBPCA scores and loadings, which indicates that the biggest changes occur during the first few minutes of the reaction after the spectra tend to cluster in PCA space and the spectral variance is reduced, indicating reaction completion. We have to be aware, however, that other processes such as changes in folding state or aggregation could also be impacting the sample and will cause some spectral variation, as discussed in chapter 4, which means that the samples will not perfectly overlap and the scores plots.

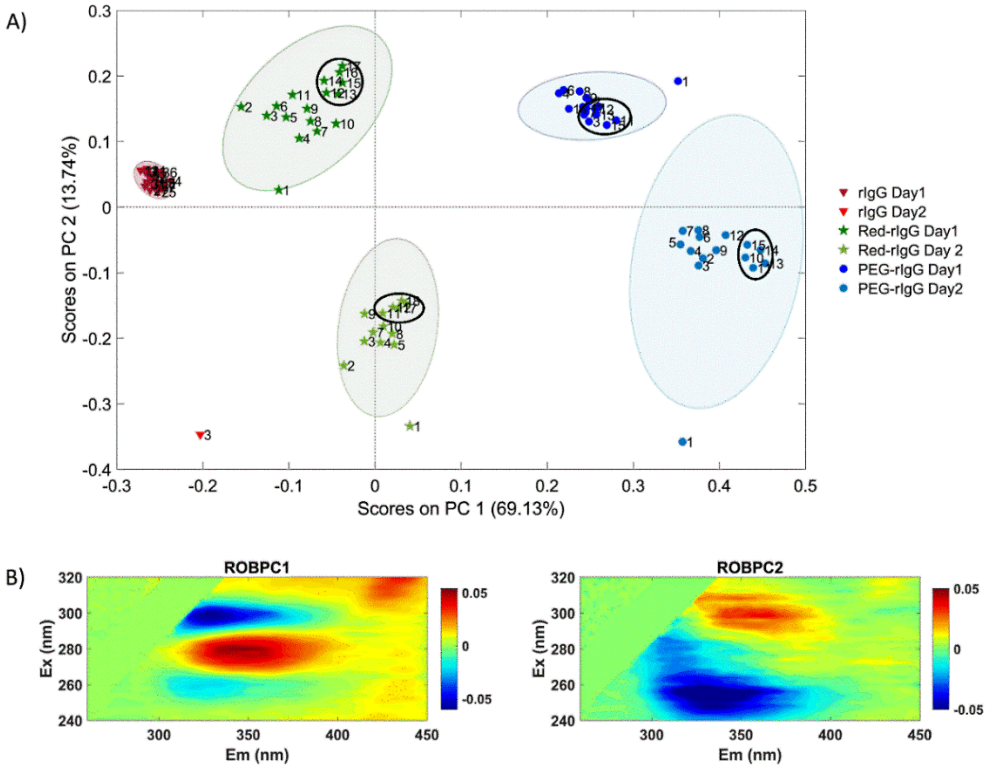


Figure 5.2- Results from EEM‖ time-course measurements (every 8 minutes) of two PEGylation reactions. The numbers in the plot designate the sequence of measurements during each reaction step (1 to 18): (A) ROBPCA scores plot shows that the biggest spectral changes (1→2→3) occurs at the start of the reaction (<~30 min) after which they tend to a cluster, which is indicative of reaction completion. These clusters are similar in size to that of the starting material cluster which indicates the degree of measurement error. The loadings (B) plot shows similar features for this reaction as previously observed for reactions 1-19. Ellipses are the 95% confidence interval for each cluster. Ellipses are the 95% confidence interval for each cluster.

It should also be noted that the separation along PC2 between the two reactions was due to the two rIgG samples used were from the same lot, but from two different

containers. The two reactions were run approximately 2 weeks apart using identical sample handling procedures in a laboratory environment where the temperature did not vary significantly (19 and 21°C). For the lower reaction, only one measurement was made of the rIgG whereas in the second (upper) reaction multiple EEM measurements were made over an extended period to provide some information about the degree of measurement error (i.e. the size of the 95% C.I. ellipse) associated with pEEM.

5.4. SEC and SDS-PAGE

SEC is commonly used for the characterization of protein PEGylation products and was used here to assess changes in size and aggregation/fragmentation profile after conjugation. The raw chromatograms of well-controlled reactions (R1-R14) (Figure 5.3A) presented only small variations in buffer components/reagents concentrations.

There was a reduction of the areas of Peak 4 (monomer) and Peaks 1 and 2 (higher order oligomers and dimers/trimers) after the reaction which was in agreement with UV-Vis results (UV-Vis A_{280} = 0.99±0.02 and 0.93±0.04 for rIgG and PEG-rIgG respectively). These could be ascribed to the dissociation of soluble oligomers under these reaction (or analysis) conditions, and second, precipitation of HMWS aggregates with the reaction. Peak 5 (Rt=13.5 min.) was possibly due to formation of lower molecular weight species via protein fragmentation caused by the reaction conditions.

There was no shift in the main peak (apart from intensity) when comparing the PEGylated species and the rIgG starting material apart from a very small change in peak widths. This was expected as the di-PEGylated species had only ~6% of difference in MW, and this SEC method in unlikely to resolve this small change. Thus, Peak 3 (Rt=10.5min) might be ascribed to multi-PEG species with 4–6 5KDa residues.

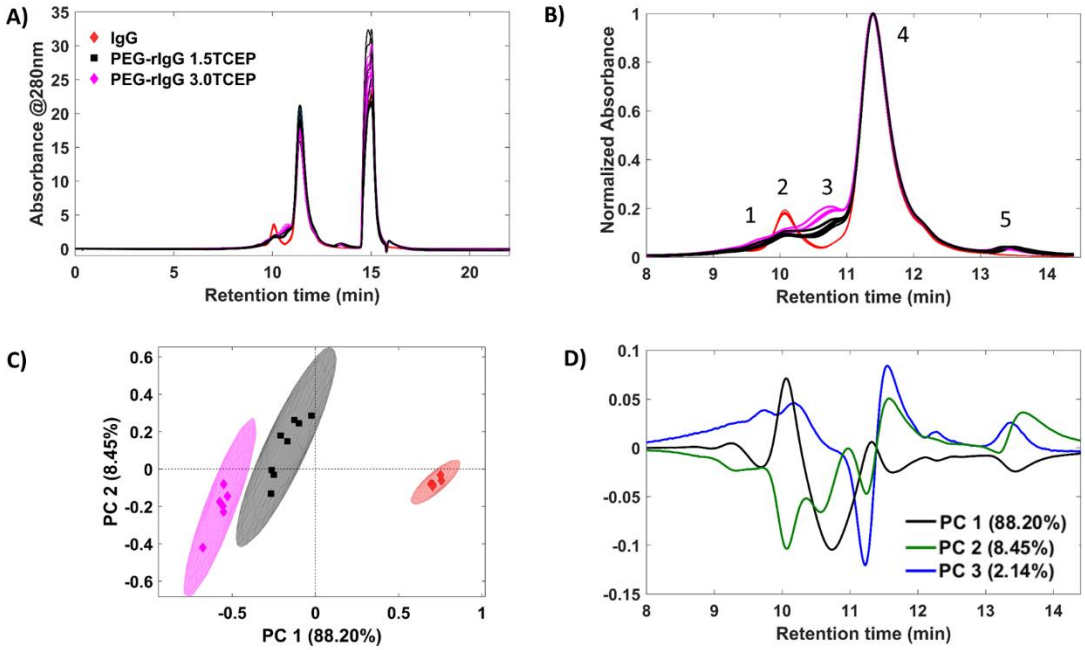


Figure 5.3- Raw chromatograms (A) obtained from SEC analysis of the IgG starting material (red) and the final un-purified reaction products (black and pink); Expanded view of SEC chromatograms for rIgG and PEG-rIgG samples (B); ROBPCA scores plot (C); and loadings (D) of the normalized SEC data. The peak of buffer components/reagents (Peak 5, Rt=15min) was removed prior to analysis.

Overall, this was a complex reaction with possibly multiple species present in both starting materials (monomer and oligomers) and reaction products (populations of PEG-IgG). This was shown with ROBPCA of the normalized chromatograms which required three principal components to explain 99.28% of the data. The scores plot showed, as expected, clear separation between starting materials and PEG samples and some variability for the PEG-rIgG products.

PC1 was the main component describing sample separation (88.20%) and separated the samples mostly due to a change in % of aggregate (9.28 min) and dimer peaks (10.05 min), and the appearance of two additional peaks in the chromatogram (Peaks 3 and 5). PC2 and PC3 (8.45% and 2.14% respectively) mostly described the variability associated with PEG-rIgG reaction products generated by different reaction conditions (C) with evidence of multiple-labelled species, fragments, and changes in dimer contribution. The polyclonal starting material and low yield of the reaction and low MW resolution[d] of this SEC method made it difficult to unambiguously assign

[d] We did not attempt method optimization to improve the separation between the products and SM.

these peaks without undertaking purification of the reaction mixtures which was not feasible here.[e]

We also attempted to analyse the products by HIC (Section 7.8.1), which is one of the key techniques for the characterization of IgG conjugates (e.g. ADCs),[60] but this represented a big challenge because the products had not been purified before we attempted the analysis. This caused a big increase in pressure after injection, especially with consecutive injections, which probably occurred because of the presence of relatively high concentration of free PEG in the unpurified samples. Thus, we decided not to continue with the HIC experiments.

SDS-PAGE with PEG staining of the gel using a 5% barium chloride solution followed by an iodine/iodide solution confirms PEGylation of IgG (Figure 5.4A), because the iodine only reacts with PEG[152] as discussed in Chapter 3. For the IgG case, the denaturing conditions of SDS promotes the dissociation of fragments with the PEG attached generating bands between 100 and 150 kDa and fainter bands with lower MW. These results suggest the formation of mostly 2-PEG species and the presence of intact antibody in solution as shown in the schematic diagram (B), which would agree with the small changes in the SEC chromatograms.

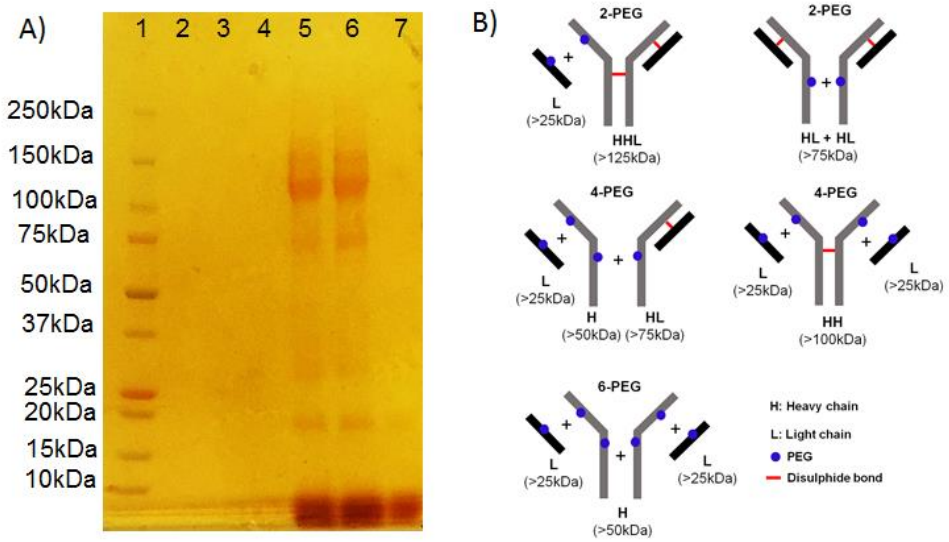


Figure 5.4: (A) SDS-PAGE with iodine staining of reaction samples and controls. Lane (1) Molecular ladder, (2) unmodified rIgG, (3) Red-rIgG (1.5TCEP), (4) Red-rIgG (3.0TCEP), (5) PEG-rIgG (1.5TCEP), (6) PEG-rIgG (3.0TCEP), (7) PEG control. (B) Schematic showing possible PEGylated species present.

[e] We attempted to purify the samples using manually packed columns with Sephadex G-50 resin and gravity flow. Elution was followed using A_{280nm} for protein and colorimetric assay with iodine for PEG. However, the results indicated that PEG and protein were eluting on the same fractions. Unfortunately, we did not have access to another purification method at that time, and because of these difficulties with the purification we decided to focus this first stage of the study on the unpurified reaction mixtures.

Most studies describing the partial reduction of antibodies by TCEP are carried out using IgG1 mAbs (e.g. for the production of ADCs),[148, 204, 208, 209] but here we (incorrectly) applied similar conditions to a rabbit polyclonal antibody expecting that a similar level of conjugation would be achieved. However, SEC and SDS-PAGE results indicate a low yield of the reactions, which suggests different reactivity of rabbit and the towards TCEP reduction. The comparison of Ellman's test results carried out with rIgG and IgG1 mAb[f] after reduction using same conditions suggests a different susceptibility of both IgG types to reduction. For instance, while rIgG reduction with 10 molar excess of TCEP at 20 °C produced an average of 1.34 free thiols per Ab, for the reaction with IgG1 at the same conditions more than double the concentration was obtained (3.9). (Table 7.9) Firstly, the two IgGs are structurally different, have different disulphide bond arrangements (Figure 1.3), and the fact that the rabbit IgG is polyclonal and thus less pure, can contribute to the lower reactivity observed. Also, rIgG has only one inter-heavy disulphide bond in the hinge region, (which is more accessible for reacting), which may also limit the number of free sulfhydrils produced under mild conditions. In addition, most antibody conjugates are produced using smaller surrogates (usually drugs of <1kDa),[211] whereas we were using a 5kDa PEG, which should cause a much bigger steric hindrance for neighbour sites, potentially reducing the PEGylation degree and reaction yield.

Ideally the products should have been better characterized (e.g. using mass spectrometry) before progressing to the next stages of the study, however, this was not possible with the resources/ knowledge we had at the time of the study.[g] Despite the relatively small overall changes and the low product yield suggested by the available methods, this proof of concept study shows the feasibility of using pEEM to monitor conjugation reactions with more complex proteins like IgG even with low degree of conjugation. Thus, the system study in this chapter should represent the minimum possible variance detected by the method.

[f] A humanized mAb is a modified antibody composed of constant and framework regions from human antibody and the CDR from a different species (usually murine), which results in reduced immunogenic response.[210] The humanized IgG1 mAb used in this experiment was donated in 2020.

[g] For instance, the DLS equipment that could have been used for validating particle formation was only acquired after this study was finished.

5.5. Spectroscopic Analysis:

The next step was to use several different spectroscopic methods to identify any significant physicochemical changes in the protein during the course of the two-step reaction. The absorption spectra showed only small changes in concentration (RSD<2.5%) between samples and, as expected, a high similarity between the rIgG starting materials (*SimI*>0.94, all lots[h]). During the reaction, scatter corrected A_{280} values decreased (Table 7.2) and the largest differences were found between rIgG and Red-IgG (4% decrease, *SimI* <0.89). An increase in absorbance at longer wavelengths (λ>320 nm) was also observed, which indicated formation of soluble or insoluble large particle, but there were no significant profile changes that could be used to monitor structural changes, or attachment of PEG residues. Thus, UV-Visible absorbance spectroscopy is a relatively insensitive method for monitoring IgG PEGylation via thiol-maleimide chemistry, which was also observed when we attempted to assess conjugation to a smaller protein in Chapter 3.

Table 5.1- Corrected absorbance at 280nm ($A_{protein,280}$), aggregation index (UV-AI) and SimI calculated from absorbance data (UV-Visible) for the different lots and reaction stages. All values are shown as mean ± stdev of all reaction samples from each Lot at each reaction stage.

Technique	Parameter	Lot	rIgG	Red-rIgG	PEG-rIgG
UV-Visible[i]	**$A_{protein,280}$**	1	1.32 ± 0.03 (RSD 2.31%)	1.30 ± 0.03 (RSD 2.81%)	1.23 ± 0.03 (RSD 2.64%)
		2	1.34 ± 0.01 (RSD 0.77%)	1.30 ± 0.01 (RSD 0.68%)	1.24 ± 0.00 (RSD 0.39%)
		3	1.13 ± 0.00 (RSD 0.01%)	1.06 ± 0.00 (RSD 0.07%)	1.00 ± 0.00 (RSD 0.22%)
	UV-AI (%)	1	0.53 ± 0.27 (RSD 50.78%)	9.74 ± 2.07 (RSD 21.20%)	10.64 ± 2.18 (RSD 20.48%)
		2	0.46 ± 0.37 (RSD 79.93%)	4.10 ± 0.53 (RSD 12.39%)	4.47 ± 0.88 (RSD 19.69%)
		3	1.58 ± 0.53 (RSD 33.25%)	17.20 ± 0.81 (RSD 4.74%)	18.13 ± 0.71 (RSD 3.92%)
	SimI	1	1.0	0.76 ± 0.04 (RSD 5.82%)	0.75 ± 0.05 (RSD 6.37%)
		2	1.0	0.90 ± 0.02 (RSD 2.08%)	0.90 ± 0.03 (RSD 3.33%)
		3	1.0	0.72 ± 0.01 (RSD 0.75%)	0.69 ± 0.00 (RSD 0.29%)

[h] *SimI* of UV-Visible data was calculated taking the rIgG from R1 (Lot1) as the reference sample.
[i] rIgG R04 was an outlier probably due to the UV-Visible measurement and was removed for statistical and AI calculations.

IPF, because it is fundamentally linked to structure and composition, is potentially a more sensitive method for assessing the small changes with each step of the conjugation. As discussed in the previous chapters, emission is determined by excitation wavelength, with λ_{ex} = 280 nm exciting both Tyr and Trp, whereas at λ_{ex}> 290 nm, mostly Trp is excited [105] and when using 280 nm excitation (Figure 5.5B/D/F) we can measure the RS band and the larger fluorescence band, which are used to calculate Fl-AI values.

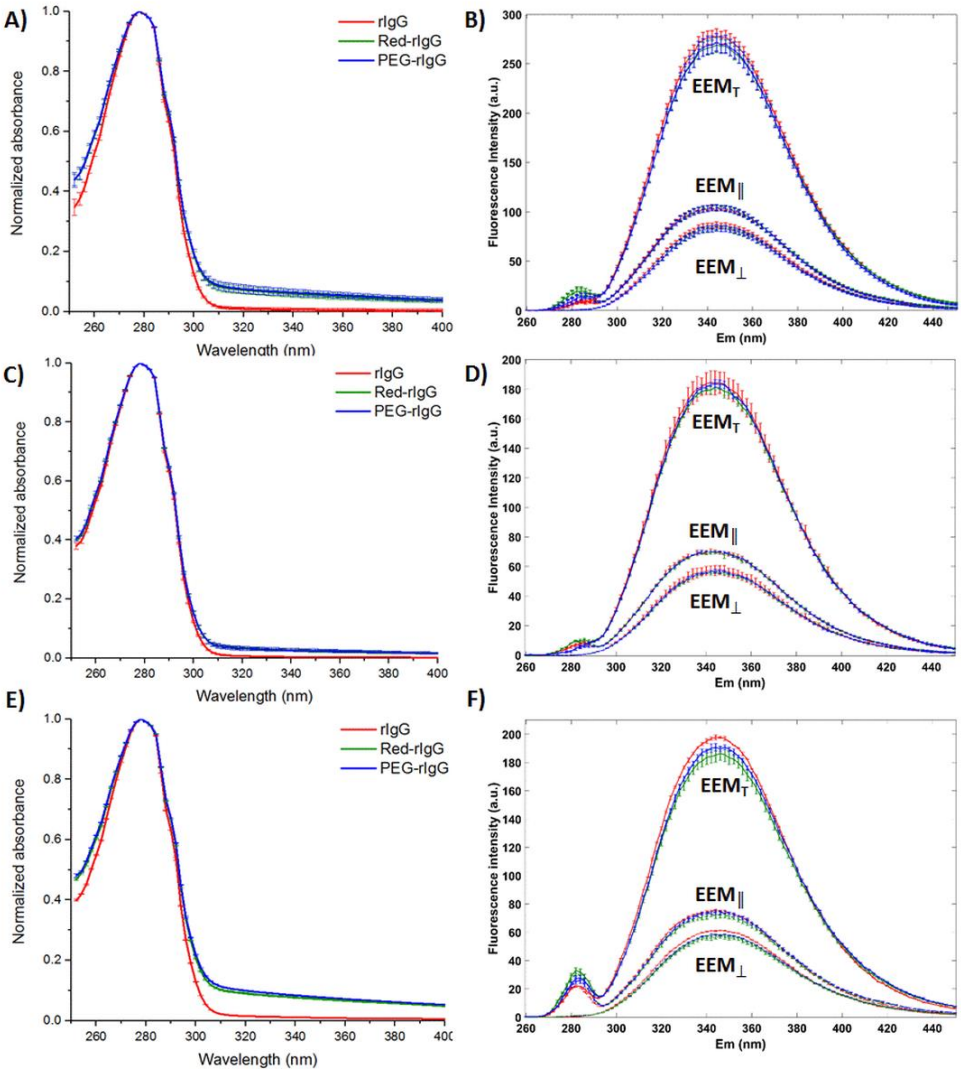


Figure 5.5- Mean (± stdev) of normalized absorbance spectra (A,C,E), and mean (± stdev) blank subtracted 2D-fluorescence spectra (B,D,F) measured at λex=280 nm, for reactions R1-R18. Top row, Lot 1, middle row, Lot 2, and bottom row, Lot 3. rIgG, Red-rIgG and PEG-rIgG are represented by the colours red, green and blue respectively.

The values calculated from the parallel polarization data showed higher signals[j], as expected, and it was observed an increase in Fl-AI of Lot1 samples after reduction (5.89% to 15.96% for SM and Red-rIgG) and then a decrease (8.95%) indicating significant changes in the physical properties of the sample, particularly on the first stage of the reaction. However, the relationship between aggregation/particle size and FI-AI values has not been validated and the results presented in Chapter 4 indicates a poor correlation with % Aggregates obtained from the standard SEC. Thus, this index provides only a qualitative measure of changes in particle size.

The maximum fluorescence intensity decreased (<6%) during the reaction with larger changes for perpendicular polarized and unpolarized EEM compared to the parallel data (Figure 5.5B/D/F and Table 5.2). This change was seen for reactions with rIgG Lot 1, which took place on five different days over six weeks. A smaller change was observed for reactions with rIgG Lot2 and 3 (overall decrease of <2.5%) and a slightly different trend was observed for Lot3, where the intensity decreased after the reduction step (<8%) and increased after the PEGylation (~2%), with, however, an overall change in intensity similar to reactions from Lot1 (~5%).

Despite small intensity differences, significant EEM spectral changes were observed as shown by changes in *SimI* values (calculated using EEM from the respective rIgG starting material as reference). For Lot 1 and Lot 2, it was observed progressive changes in rIgG structure during the reaction (*SimI* $EEM_{\parallel}$=1.0, 0.95 ± 0.01, and 0.93 ± 0.01 for starting material, reduced product, and final product, Table 5.2). Larger changes were observed when degraded samples from Lot 3 rIgG were used (*SimI*=1.0, 0.91 ± 0.00, 0.89± 0.01). In practice, one would use a validated reference batch as the *SimI* reference sample. For instance, when comparing all samples using the same reference spectrum (here, R1 rIgG starting material was arbitrarily selected), *SimI* gives fast information about the gross variability in starting materials (0.95±0.02, 0.90 ± 0.01, and 0.83 ± 0.00 for Lot 1,2,3 respectively) and, consequently, in the final PEG-IgG products (0.90±0.02, 0.88±0.00, and 0.76±0.00). (Section 7.8.4 provides more details on the use of different reference spectra for the calculation of *SimI*).

[j] Fl-AI of rIgG from Lot 1 calculated using parallel polarized data was 5.89 ± 1.22 whereas for perpendicular and unpolarized EEM it was 0.09 ± 0.06 and 2.28 ± 0.50 respectively. This confirms the higher sensitivity of $Fl_{\parallel}$-AI as discussed in Chapter 4.

Table 5.2- Fluorescence intensity maxima, Fl-AI, Total RS area, and *SimI* calculated from EEM data for rIgG lots (1,2,3) at each reaction stage. All values are shown as mean ± stdev of all the samples for each lot. Max. Intensity, Fl-AI, and Total RS area were calculated using data after blank subtraction SimI was calculated taking each reaction starting material as the reference.

Technique	Parameter		rIgG	Red-rIgG	PEG-rIgG
$EEM_{\parallel}$	$Fl_{\parallel}$-AI (%)	1	5.89 ± 1.22 (RSD 20.73%)	15.96 ± 3.36 (RSD 21.03%)	8.95 ± 2.97 (RSD 33.21%)
		2	9.77 ± 1.26 (RSD 12.92%)	12.14 ± 0.72 (RSD 5.96%)	4.56 ± 0.75 (RSD 16.50%)
		3	27.25 ± 1.11(RSD 4.07%)	37.93 ± 2.96 (RSD 7.81%)	29.89 ± 2.98 (RSD 9.97%)
	Total RS area	1	2939 ± 178 (RSD 6%)	5113 ± 498 (RSD 10%)	4281 ± 505 (RSD 12%)
		2	1808 ± 129 (RSD 7%)	2418 ± 53 (RSD 2%)	2524 ± 49 (RSD 2%)
		3	2693 ± 63 (RSD 2%)	4396 ± 261 (RSD 6%)	3957 ± 257 (RSD 6%)
	RS_{296}/Fl_{max}	1	1.45 ± 0.08	2.51 ± 0.29	2.11 ±0.27
		2	1.49 ± 0.10	1.96 ± 0.03	2.02 ± 0.02
		3	2.08 ± 0.23	3.66 ± 0.12	2.66 ± 1.05
	Max. Intensity [a]	1	128.34± 1.72 (RSD 1.34%)	125.44 ± 1.78 (RSD 1.42%)	122.87 ± 2.63 (RSD 2.14%)
		2	153.07 ± 2.27 (RSD 1.49%)	151.22 ± 0.75 (RSD 0.49%)	149.35 ± 0.99 (RSD 0.66%)
		3	154.74 ± 0.53 (RSD 0.34%)	147.26 ± 4.81 (RSD 3.27%)	148.48 ± 2.44 (RSD 1.65%)
$EEM_{\perp}$	Total RS area	1	61 ± 9 (RSD 14%)	153±27 (RSD 18%)	109±20 (RSD 18%)
		2	17 ± 1 (RSD 9%)	30±5 (RSD 17%)	28±2 (RSD 6%)
		3	36±2 (RSD 4%)	114±1 (RSD 0%)	84±6 (RSD 7%)
	Max. Intensity [a]	1	101.07 ± 2.32 (RSD 2.29%)	95.74 ± 2.35 (RSD 2.45%)	95.14 ± 2.66 (RSD 2.80%)
		2	119.89 ± 2.33 (RSD 1.95%)	118.63 ± 1.58 (RSD 1.33%)	119.05 ± 1.20 (RSD 1.01%)
		3	123.57 ± 0.62 (RSD 0.50%)	114.17 ± 4.07 (RSD 3.56%)	116.38 ± 0.98 (RSD 0.84%)
EEM_{T}	Total RS area	1	3060 ± 193 (RSD 6%)	5418 ± 547 (RSD 10%)	4497 ± 542 (RSD 12%)
		2	1820 ± 139 (RSD 8%)	2472 ± 61 (RSD 2%)	2572 ± 53 (RSD 2%)
		3	2755 ± 67 (RSD 2%)	4615 ± 263 (RSD 6%)	4112 ± 270 (RSD 7%)
	Max. Intensity [a]	1	329.01 ± 6.10 (RSD 1.85%)	315.34 ± 5.85 (RSD 1.86%)	311.66 ± 7.98 (RSD 2.56%)
		2	389.36 ± 6.41 (RSD 1.65%)	384.32 ± 2.83 (RSD 0.74%)	381.23 ± 4.43 (RSD 1.16%)
		3	401.72 ± 1.72 (RSD 0.43%)	376.69 ± 13.69 (RSD 3.63%)	382.93 ± 6.42 (RSD 1.68%)

[a] Corrected values calculated using factors 1.80, 1.89, and 1.84 for $EEM_{\parallel}$, $EEM_{\perp}$ and EEM_{T} data respectively. (See intensity correction – Section 7.8)

The RS band extracted from EEM‖ suggested formation of reversible HMWS intermediates with conjugation. For all the lots, the area corresponding to the RS‖ band first increased significantly (e.g. Lot 1, ~74%) on reduction, and then decreased (e.g. Lot 1, 16%) after alkylation.(Table 5.2) However, as previously discussed, absolute RS intensity measurements are less reproducible and it was thus preferable to use a ratio between RS at a specific wavelength and fluorescence intensity maximum (RS_{296}/Fl_{max}).The comparison of the ratios for the various lots indicated big differences between Lot 1 and 2 compared to the degraded rIgG (Lot 3), suggesting different degrees of HMWS content despite a similar trend for all lots over the reaction. An increase in RS_{296}/Fl_{max} was observed after reduction and a small decrease after alkylation, which agreed with the Fl-AI and RS band area trends. As stated in the previous chapters, extracting quantitative information from the RS data was not practical here because of the limitations in terms of both SNR and the wavelength range was not ideal for particle size analysis.

Another way to investigate protein structural changes is via *aniso*-EEM maps which are very sensitive to structural and chemical changes which affect FRET, large changes in molecular weight, or environmental factors which affect rotation correlation time.[120, 126] Here, it is likely that more than one factor is present, which makes it difficult to ascribe the observed anisotropy changes to a specific issue, and only a small proportion of the map area was changing during the reaction. (Figure 5.6) The use of *aniso*-EEM for assessing smaller proteins can be hampered by the very low anisotropy values as discussed in Chapter 3, but for IgG it should provide more information because of the higher anisotropy values. *SimI* analysis with the respective rIgG starting material as the reference was used here to qualitatively assess changes in *aniso*-EEM maps (Figure 5.7, Table 7.12 in Section 7.8.4) during the reaction. The values were calculated using selected regions of the spectra (i.e. Tyr, Tyr+Trp and Trp regions) and the whole emission space. *SimI* for the full emission space decreased from 1 to 0.64±0.04 after reduction and then to 0.52±0.08 after the final step, indicating very significant photophysical changes. While *SimI* of pEEM data (Figure 5.7B) was best for simple reaction stage discrimination[a], *SimI* analysis of full *aniso*-EEM map enabled discrimination *(p-value<0.05)* of unpurified final products (Figure 5.7F) according to

[a] Absorbance spectra did not have any discriminative power for either the reaction stage or reaction condition, as shown by Figure 5.7.

the reaction conditions (1.5 vs. 3 × TCEP). Mean anisotropy values, on the contrary, were essentially identical: 0.110±0.002, 0.113±0.001, and 0.112±0.001 which suggested that protein size changes were small. As shown by Figure 5.7D, anisotropy values varied more at longer excitation wavelengths, λ_{ex}=~290–310 nm.

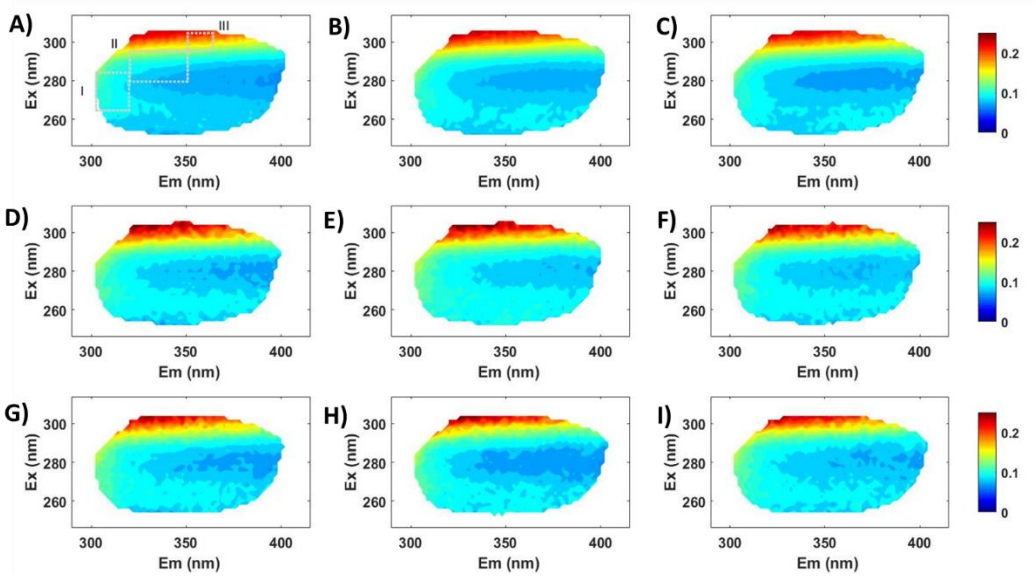


Figure 5.6- Mean *aniso*-EEM maps of rIgG(A,D,G), Red-rIgG, (B,E,H), PEG-rIgG (C,F,I) from reactions using Lot1,2,3 of rIgG respectively.

This was more evident when comparing the variation in the distinct regions of the *aniso* maps: Tyr region ($\lambda_{ex}/\lambda_{em}$= 260-280/300-320 nm, Figure 5.7C), Trp region ($\lambda_{ex}/\lambda_{em}$= 286-304/350-360 nm, Figure 5.7D) and the combined Tyr+Trp region ($\lambda_{ex}/\lambda_{em}$= 280-296/320-350 nm, Figure 5.7E). *SimI* of the Trp emission region showed the largest variation as expected because Trp emission is more sensitive to environmental/structural changes. However, using *SimI* of the full map (Figure 5.7F), provided the most information, and suggested biggest changes in terms of structure when the greater TCEP excess was used. However, since pEEM measurements were made on un-purified reaction mixtures and the orthogonal methods used suggested low yield, more detailed product analysis was not possible.

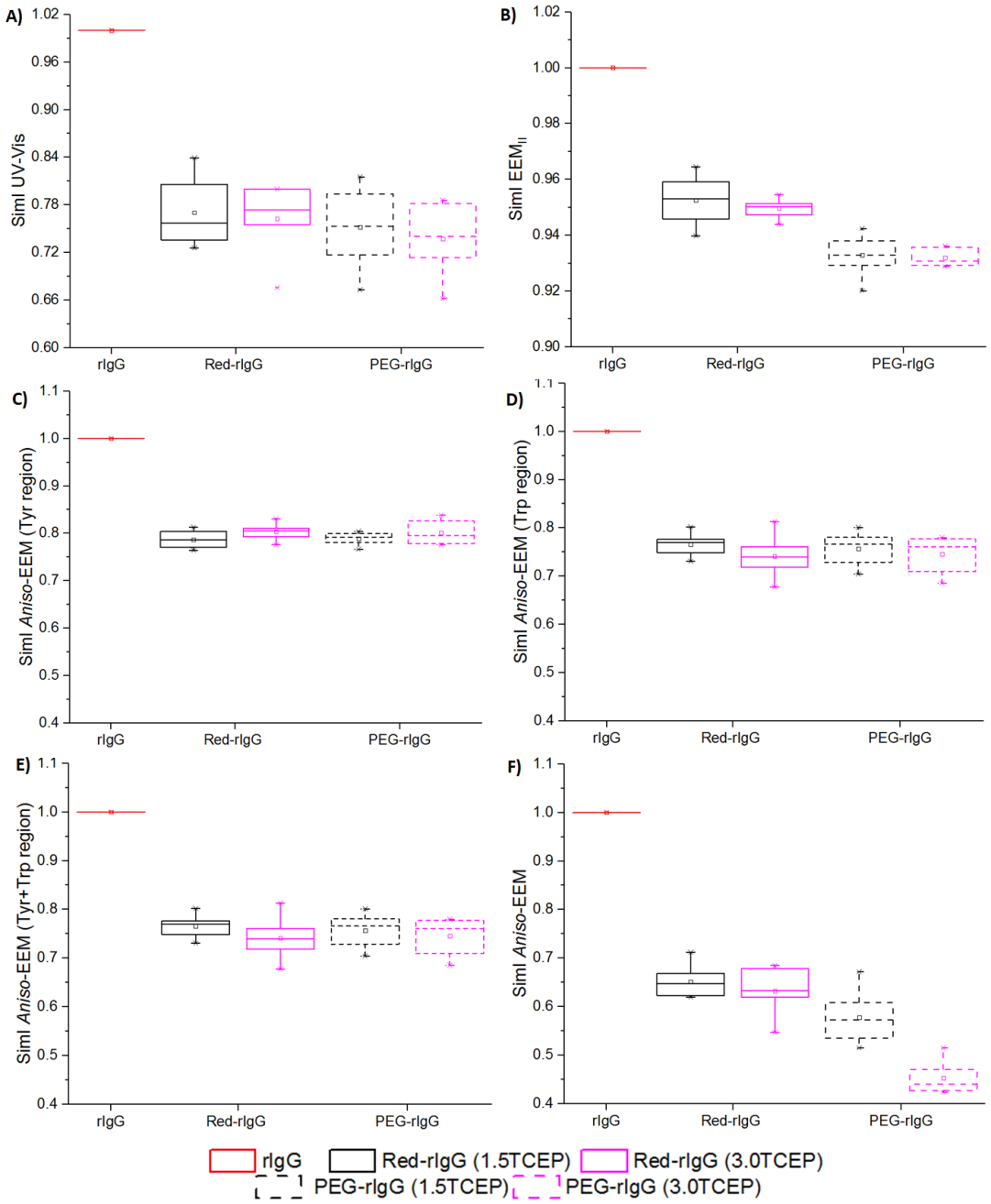


Figure 5.7- *SimI* values calculated for R1-R14 over the reaction from (A) Norm-UV Vis spectra , (B) Norm-EEM‖ and anisotropy maps of the: (A) Tyr region ($\lambda_{ex}/\lambda_{em}$=260−280/300−320 nm); (B) Trp+Tyr region ($\lambda_{ex}/\lambda_{em}$=280−296/320−350 nm); (C) Trp region ($\lambda_{ex}/\lambda_{em}$ = 296−304/350−360 nm); and (D) the full emission space, using the respective IgG starting material (used for each reaction) as the reference.

Both intensity and ratios of RS changed the most after the reduction step, which indicated HMWS formation (or weakly bound species). These were not accompanied by large variations in either fluorescence profile (*SimI* changes of ~7%) or in fluorescence intensity which indicated only small tertiary structure changes (or changes in fluorophores local environment) and a minimal change in fluorophore concentration. The latter might have been caused by protein precipitation during

reduction. Thus, from spectral analysis we suggest that the observed anisotropy changes were due to changes in FRET with minimal variation in size/aggregation. This would agree with PEGylation yielding mostly 2PEG species (and at low concentration), and thus a small overall protein size change which was accompanied by formation of some weakly bound species and, potentially, by local physicochemical effects (i.e. quenching of more externally located Trp residues). It is important to note that anisotropy measurements are very sensitive to SNR and residual scatter and therefore it can be considered the least robust measurement for use as a PAT tool. This coupled with the complexity of the photophysical changes occurring required the use of a more robust measurement like pEEM and multivariate analysis to better understand these changes and develop a method for reaction monitoring for determining the end points of the reduction and alkylation steps.

5.6. Multivariate analysis:

The EEM difference spectra calculated for reactions from each lot (Figure 5.8) highlighted the impact of conjugation in the EEM of rIgG. When compared with the control experiments, which indicate the intrinsic measurement/experiment variability, (Figure 7.6) this suggested small but significant spectral changes during the reaction. (Figure 7.10 in Section 7.8.5 shows the mean±stdev calculated for rIgG, Red-rIgG and PEG-rIgG).

ROBPCA was then applied to better identify the physicochemical changes responsible for pEEM and absorption spectrum variation between the different source lots and at different reaction stages.

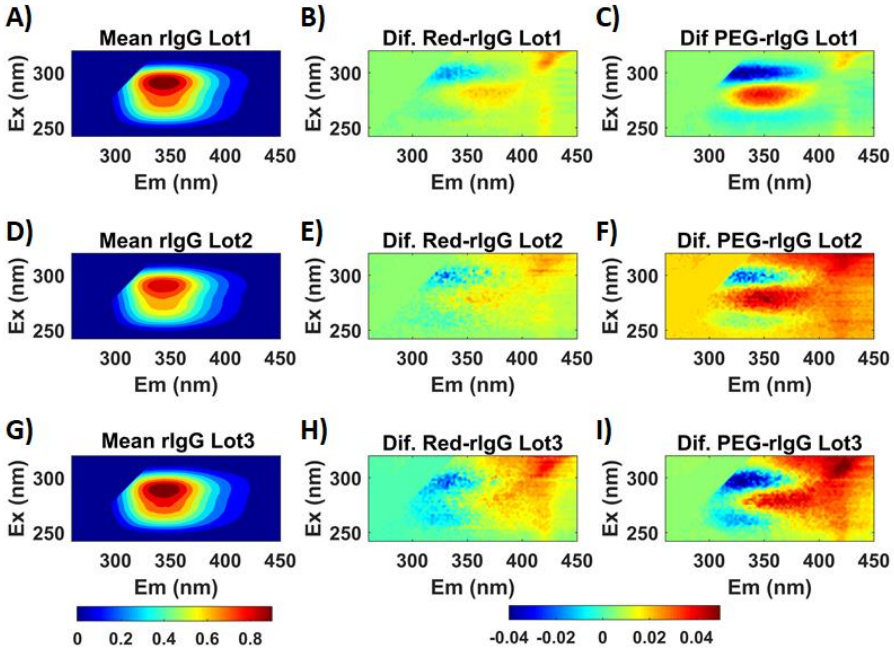


Figure 5.8- Mean EEM‖ spectra of rIgG (A,D,G) and difference spectra calculated using EEM‖ data from the Red-rIgG (B,E,H) and PEG-rIgG (C,F,I) from Lot1,2 and 3 respectively. Mean and difference spectra were calculated using normalized data. ($EEM_{Dif.} = EEM_{Mean\ PROD} - EEM_{Mean\ SM}$).

5.6.1. Source lot screening:

Structural and other variations in the protein starting material is an important factor in determining final product composition, purity, and quality. Therefore, measuring any variation would be an important part of the QbD process, as previously discussed in Chapter 4. The goal here is to use pEEM to assess not only the variation in starting material but the impact on the reaction progression.

Table 5.3- Summary of ROBPCA results using normalized absorbance and pEEM spectra. Model contained samples from all lots and all reaction steps (n=57).

	% variance captured by each ROBPC			
ROBPC	**UV-Vis**	**$EEM_{\parallel}$**	**$EEM_{\perp}$**	**EEM_{T}**
1	99.18	52.01	52.53	64.26
2	0.56	14.84	20.54	12.72
3	-	10.79	5.36	8.03
4	-	4.29	2.70	2.59
Total variance	99.74	81.92	81.13	87.60

ROBPCA of normalized absorbance data yielded two principal components (PC) explaining 99.74% of the variance. PC1 accounted for 99.18% and PC2 0.56%

of the total variance (Table 5.3). The biggest variability in starting material was related to PC2, which according to the loadings plot was related to very small differences in the 250–270 nm region (Figure 5.11B) However, PC2 contribution to total variance was rather small and not reliable as a diagnostic for source lot discrimination. ROBPCA of pEEM on the other hand, seemed to be a better screening tool to identify deviations from a reference sample (e.g. a golden lot. here, Lot1).

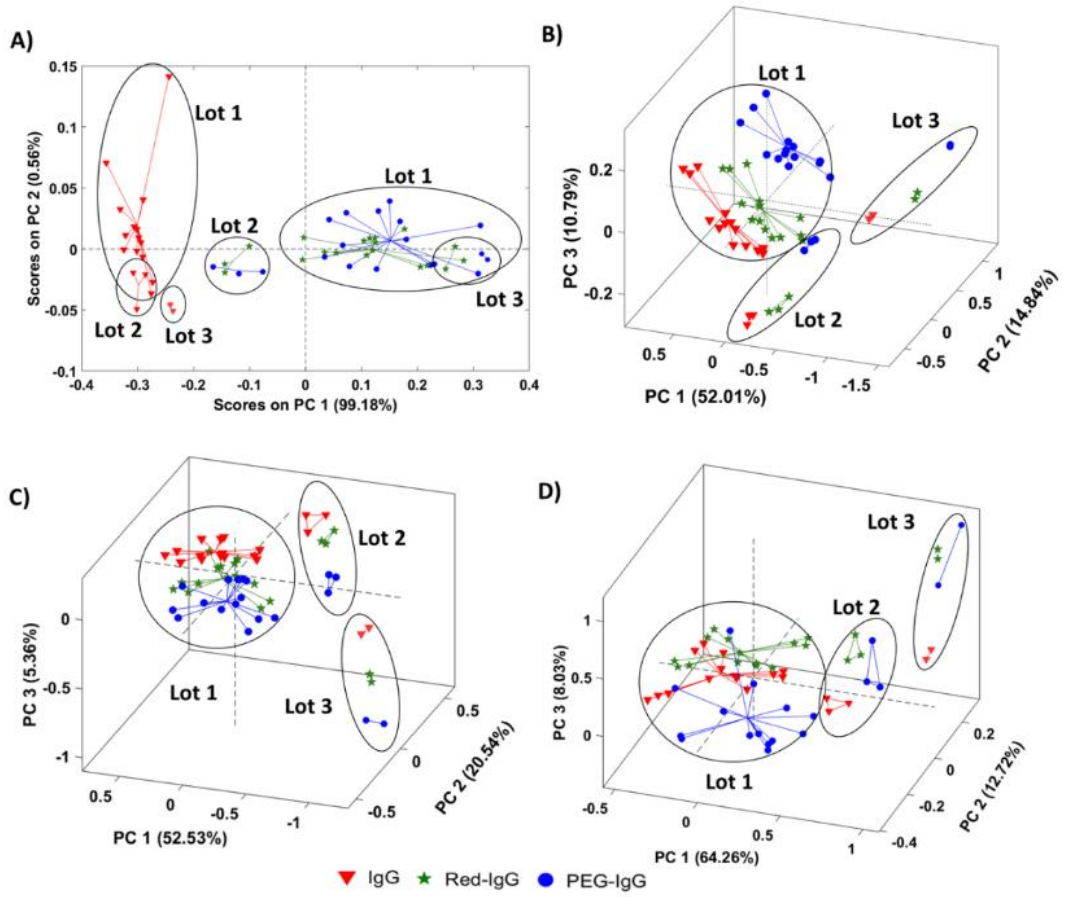


Figure 5.9 - ROBPCA scores plots of PEGylation reactions using normalized spectra from: (A) UV-visible; (B) EEM$_{\parallel}$; (C) EEM$_{\perp}$; and (D) EEM$_{T}$ measurements. The ellipses were manually added to facilitate the visualization of samples from each lot and do not represent the confidence limits for each cluster.

Four PCs accounted for 82% (EEM$_{\parallel}$), 81% (EEM$_{\perp}$), and 87% (EEM$_{T}$) of total variance. This difference in % of explained variance was mostly due to noisier EEM$_{\perp}$ and EEM$_{\parallel}$ data compared to unpolarized EEM. Scores plots (Figure 5.9B/C/D) showed discrimination between reaction stages and source lots. When assessing all reactions (R1-19, n=57) the biggest difference encountered was between the source lots. Starting materials from Lots 2 and 3 were clearly identified as being significantly different (and outliers in Figure 5.10) leading to products with slightly different

structure and spectral profiles, as indicated in the scores plot. The reactions however, showed similar trajectories, loadings plots, and discrimination of reaction products (Figure 5.12 shows ROBPCA results of model with Lot 2 and 3 only).

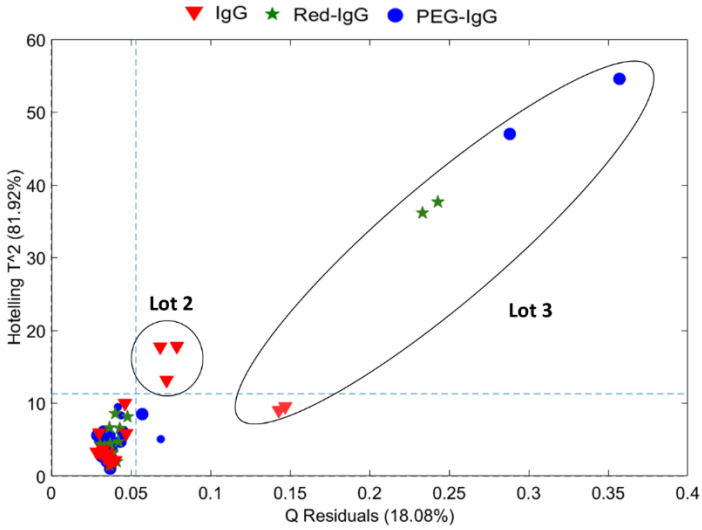


Figure 5.10- Outliers plot of ROBPCA analysis of EEM‖ dataset. The ellipses for the two lots were included as a visual guide to show the two lots and do not represent confidence intervals for a statistical resolution.

In terms of a practical application, pEEM & ROBPCA could be used in a production environment to quickly identify processes that were outside the limits defined by a golden batch (i.e. Lot1 here). Refolded robust PC loadings plots showed the areas of highest spectral variance (Figure 5.11C-H). For all the EEM datasets, a similar ROBPC1 loadings and score trend was observed. The main spectral feature responsible for this separation was the change in contribution between the Tyr/Trp area (λ_{ex} ~280 nm) and Trp only area (λ_{ex} >302 nm) caused by reaction induced changes leading to variations in IFE and FRET rates. For $EEM_{T,}$ an inverted orientation in positive/negative contribution of Tyr/Trp was observed, but the same areas of the map were changing.

ROBPC1 also indicates sample discrimination according to the different reaction conditions and discrimination according to reaction stage. This can be better visualized when each PC scores were plotted separately (Figure 7.11). ROBPC1 scores plot shows a visual separation of samples prepared using different TCEP concentrations and this discrimination was confirmed by the two-sample t-test results using 95% confidence interval. When comparing the scores values of Red-rIgG 1.5TCEP *vs* Red-rIgG 3.0TCEP, PEG-rIgG 1.5TCEP *vs* PEG-rIgG 3.0TCEP and also all Red-IgG *vs* all

PEG-IgG the p-value was lower than 0.05 indicating that the separation by PC1 scores is statistically significant.

The second spectral feature playing an important role in the discrimination of reaction samples seemed to be Trp emission quenching caused by these structural changes. This was represented by ROBPC2 for $EEM_{\perp}$ and EEM_{T}, and PC3 for the $EEM_{\parallel}$. Sample separation was different for each pEEM data, however, all three clearly discriminated according the different rIgG lots.

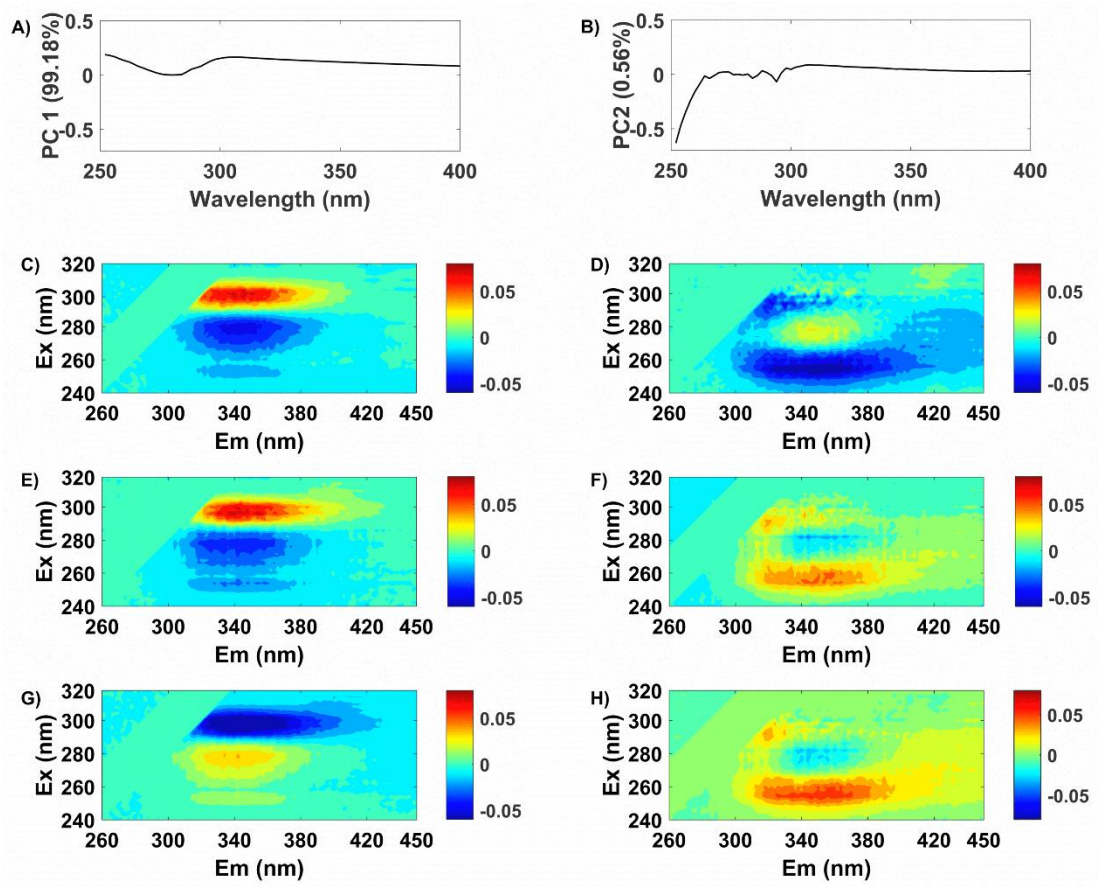


Figure 5.11- ROBPC1 and 2 loadings of normalized UV-visible, refolded ROBPCA 1, 3 loadings of normalized $EEM_{\parallel}$ (C,D), and ROBPCA 1, 2 of normalized $EEM_{\perp}$ (E,F) and EEMT data (G,H). Here samples from all the lots were used in the model.

ROBPC 3 and 4 (2 and 4 for $EEM_{\perp}$) had a smaller contribution to the total variance, but PC3 in particular showed a trend that resembled the emission profile of some stressed samples studied in Chapter 4: changes in Tyr/Trp and in the longer wavelength region. (Figure 7.13) These ROBPCs separate Lot2 and 3 samples, and to some extent, samples at different reactions stages (particularly samples from Lot 2 and 3). (Figure 7.12) This corroborates with our hypothesis that this spectral profile is

characteristic of stressed samples, especially considering that all measurements indicated the formation of HMWS with the reaction.

Poorly controlled PEGylation reactions (Lot2 and Lot3):

When the poorly controlled reactions were analysed separately only three PCs were necessary to explain most of the EEM‖ dataset variance (~90%). Compared to the models where all the samples were taken into account, this showed a lower number of PCs, which can be explained by only one reaction condition (3.0 molar excess TCEP). This was done to reduce the variability in the dataset and try and elucidate variations only arising from the poorly controlled starting material.

ROBPC1 loadings plot Figure 5.12A indicates that the separation of the Lot2 from Lot3 reactions was based on the spectral features of non-native protein samples. (Both a red shift in Trp emission and the deep blue autofluorescence region mentioned in Chapter 4). ROBPC2 (Figure 5.12B) highlights changes in IFE and FRET effects (Trp and Tyr/Trp) used to discriminate samples at the different stages of the reaction, which is a similar trend separating reaction samples in the ROBPCA of all reactions (Figure 5.11D,F,H). ROBPC3 is only responsible for a very small % of variation, but it also seems to be related to changes in emission in longer wavelength region. As we previously highlighted, this is a very noisy region and also not well studied, thus any conclusions based on these changes should be avoided.

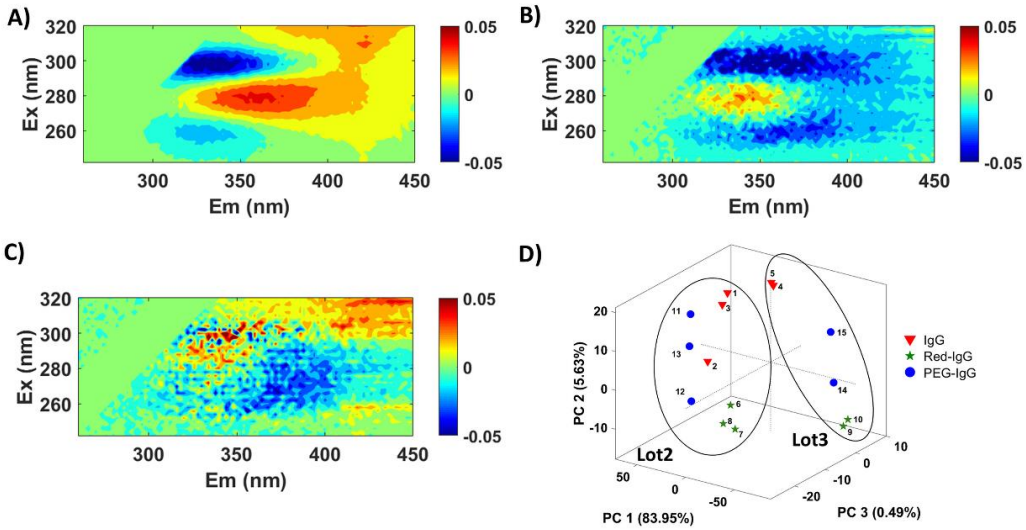


Figure 5.12-ROBPC1,2,3 loadings (A,B,C) and 3D scores plot (D) of normalized EEM‖ from reactions using Lot2 and Lot3 samples. The ellipses for the two lots were included as a visual guide to show the two lots.

5.6.2. Reaction stage discrimination:

Here the main objective was to evaluate the ability of pEEM to discriminate samples according to the reaction stages, thus we consider only the samples from the well-controlled reactions (R1-14). This was done to avoid the big variability coming from the source lot and consequently focus on the variations caused by the reaction.

ROBPCA of normalized absorbance spectra data produced two principal components (PC) explaining 99.37% (PC1) and 0.45% (PC2) of the total variance (Table 5.4). The first PC (Figure 5.13) discriminated rIgG starting materials from the reaction products (Red-rIgG and PEG-rIgG) because of increased light scatter as shown in the PC1 loadings. But apart from these changes in scattering of light (which as discussed in Chapter 4 are related to changes in the physical aspects of the sample, no additional information could be extracted from absorbance spectra, confirming the poor sensitivity of absorption measurements for identifying small protein structural changes. Thus, it was unable to discriminate either source lots or reaction products from R1-14.

ROBPCA of normalized pEEM data required four PCs for capturing the variance of reaction samples from R1–R14, and again, the difference in variance explained between the various measurements was observed: 81% , 75% and 85% of total variance explained for $EEM_{\parallel}$, $EEM_{\perp}$, and EEM_T respectively (Table 5.4). As expected, when the large source lot variation was not present, the model (R1-14) better highlighted the structural changes induced by the reaction (Figure 5.13 in comparison with Figure 5.9).

Table 5.4- Summary of ROBPCA results from normalized UV Visible, $EEM_{\parallel}$, $EEM_{\perp}$,EEM_T data for all reaction steps. Here only reactions from Lot1 were used.

	% variance captured by each ROBPC			
ROBPC	**UV-Vis**	**$EEM_{\parallel}$**	**$EEM_{\perp}$**	**EEM_T**
1	99.37	50.38	43.67	57.30
2	0.45	18.25	21.23	14.19
3	-	10.10	6.42	10.80
4	-	2.17	3.48	3.26
Total variance	99.82	80.91	74.80	85.36

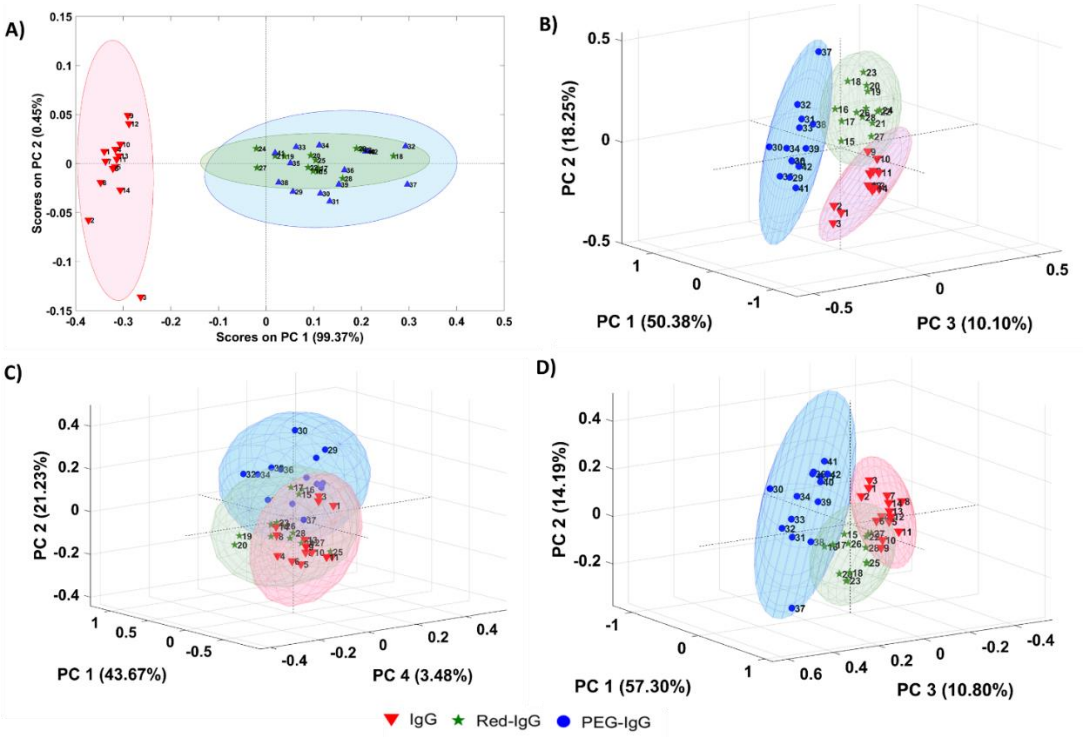


Figure 5.13- ROBPCA analysis of normalized UV-vis spectra indicates good separation of rIgG samples but poor separation of Red/PEG-rIgG or samples generated under different reaction conditions (A) ROBPCA analysis of normalized EEM$_{\parallel}$ (B), EEM$_{\perp}$ (C), and EEM$_{T}$ (D) shows better discrimination between the IgG starting material, reduced form, and PEGylated products. The ellipsoids represent the confidence intervals (95%) calculated for each cluster. The best visual discrimination of reaction stages was obtained using EEM$_{\parallel}$. Here only samples from lot1 were included I the model.

For all EEM models, ROBPCA1 scores seemed to discriminate Lot1 reactions according to the different TCEP concentrations, which was confirmed by statistical hypothesis testing using t-test ($p<0.05$ for samples at both Red-rIgG and PEG-rIgG stages). The later PCs had less discriminating power as the spectral changes are relatively small and thus at least three components were required to correctly evaluate pEEM changes in this process. ROBPCA loadings indicated that the observed pEEM changes, while relatively small, were all significant and related to either reaction induced structural changes or starting material variation. (Figure 5.14 and Figure 7.14)

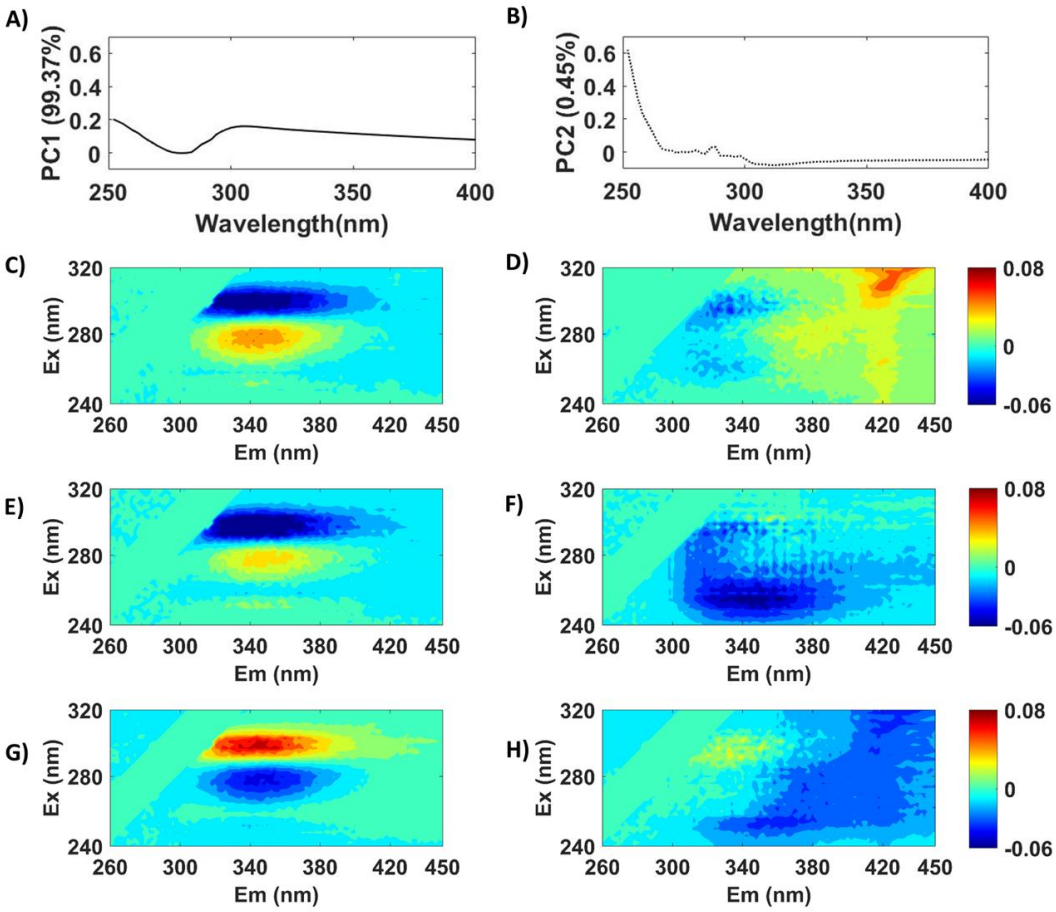


Figure 5.14- ROBPC1/2 loadings of normalized UV-visible (A,B); $EEM_{\parallel}$ (C,D); $EEM_{\perp}$ (E,F); and EEM_{T} data(G,H), respectively showing the areas of largest spectral change during PEG conjugation.

5.6.3. <u>End point determination:</u>

ROBPCA scores plots indicate some degree of overlap between confidence intervals of each cluster. Here we explored the application of linear and non-linear classification algorithms, SIMCA and SVM, to quantitatively assess which pEEM measurement was best at discriminating each reaction stage. The goal being a simple classification method that could be used to validate that reactions were completed before progressing to the next process step. As discussed in Section 5.6.1. of this chapter, Lot 2 and 3 samples were significantly different compared to the Lot1 (our golden batch). We understand that in an industrial context these samples, being known different raw materials would not be used in a process and thus were excluded. Thus, for reaction stage classification, we only used the Lot 1 sample data which were unfolded and randomly separated into two a calibration (n=10) and test set (n=4).

Table 5.5- Confusion matrix obtained by full cross-validation (venetian blind) of UV-Visible, $EEM_{\parallel}$, $EEM_{\perp}$ and EEM_{T} measurements by SIMCA and SVM based classification models. I = IgG starting material; R = reduced product sample; and P = PEGylated product sample.

Models	**Calibration Results**																							
	UV-Visible						**$EEM_{\parallel}$**						**$EEM_{\perp}$**						**EEM_{T}**					
	SVM			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**		
Actual	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**
I	10	0	0	10	0	0	10	0	0	10	0	0	10	0	0	10	5	0	10	0	0	10	2	3
R	0	5	7	0	10	0	0	10	0	1	10	0	0	10	0	10	10	0	0	10	0	0	10	0
P	0	5	3	0	0	10	0	0	10	0	0	10	0	0	10	0	0	10	0	0	10	0	0	10
Models	**Validation Results**																							
	UV-Visible						**$EEM_{\parallel}$**						**$EEM_{\perp}$**						**EEM_{T}**					
	SVM			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**		
Actual	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**	**I**	**R**	**P**
I	4	0	0	3	0	0	4	0	0	3	1	0	4	0	0	4	2	0	4	1	0	3	1	0
R	0	2	2	0	3	1	0	4	0	1	4	0	0	4	1	4	4	0	0	3	0	0	3	0
P	0	2	2	1	3	1	0	0	4	0	0	3	0	0	3	0	0	3	0	0	4	0	0	2

Table 5.6- Results of sensitivity (TPR), Specificity (TNR) and misclassification error (Err) in % for SIMCA and SVM classification models of UV-Visible, $EEM_{\parallel}$, $EEM_{\perp}$, and EEM_T measurements according to the reaction step. I = IgG starting material; R = reduced product sample; and P = PEGylated product sample.

Models	**Calibration Results**																							
	UV-Visible						**$EEM_{\parallel}$**						**$EEM_{\perp}$**						**EEM_T**					
	SVM			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**		
	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P
TPR	100	50	20	100	100	100	100	100	100	100	100	100	100	90	100	100	100	100	100	90	100	100	100	100
TNR	100	60	75	100	100	100	100	100	100	100	95	100	95	100	100	75	50	100	95	100	100	90	100	100
Err.	0	43	43	0	0	0	0	0	0	0	3	0	3	3	0	17	33	0	3	3	0	7	0	0
Models	**Validation Results**																							
	UV-Visible						**$EEM_{\parallel}$**						**$EEM_{\perp}$**						**EEM_T**					
	SVM			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**			**SVM**			**SIMCA**		
	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P	I	R	P
TPR	100	50	50	75	75	25	100	100	100	75	100	75	100	100	75	100	100	75	100	75	100	75	75	50
TNR	100	75	75	100	88	63	100	100	100	88	88	100	100	88	100	75	50	100	88	100	100	88	100	100
Err.	0	33	33	8	16	50	**0**	**0**	**0**	17	8	8	0	8	8	17	33	8	8	8	0	17	8	17

SVM and SIMCA classification performance using absorbance spectra and polarized EEM data is shown in Table 5.5 which details the number of correct and incorrect predictions for calibration and validation models. Table 5.6 summarizes the statistical results of the classification models according to each reaction stage. Using absorbance spectra, a lower percentage of positive cases in the test set were correctly classified. SVM, generated 100% correct identification of starting materials, whereas SIMCA had better performance for the reduction step (16% error). Unfortunately, based on our previous findings, successful discrimination of rIgG starting material was a consequence of increased light scatter rather than significant absorbance changes. Thus, the poor sensitivity of absorbance spectroscopy to small protein structural changes makes it unsuitable for reaction monitoring or end-point determination.

Classification using pEEM data provided significantly better results. Overall $EEM_{\parallel}$ data were better classified than $EEM_{\perp}$ or EEM_{T}, and thus the best choice for end-point determination. SVM performed better than SIMCA which can be attributed to non-linear fluorescence behaviour possibly present in these samples. These effects probably led to SIMCA generating poorer outcomes in terms of classification error for the starting material and reduction stages. In terms of accuracy, the classification errors of prediction obtained for $EEM_{\parallel}$ using SVM (100% correctly classified), demonstrated the superior performance of these measurements to be used for successfully monitoring the reaction stage.

Here, we also evaluated the results if 1) samples from Lot 2 and 3 samples were included in the validation set and 2) Lot 2,3 samples were included with Lot 1 in both calibration and validation sets. For this, only $EEM_{\parallel}$ data was used and SVM carried, as these had been determined to be the best combination for classification in the previous section. Table 7.13 shows a summary of the samples used for calibration and validation in the two situations.

When samples from Lot 2 and 3 were used in the validation dataset, slightly poorer results were obtained prediction errors <15. This was mainly because of Lot3 samples which are not well classified at the starting material and reduction stages. However, even with considerable structural differences, all PEG-rIgG final products were correctly classified (Sensitivity=100%), showing a relatively good predictive power even outside the golden batch. (Table 5.7)

Table 5.7- Summary of performance results obtained by classification of EEM$_I$ using SVM including Lot 2,3 samples in the validation set. I = IgG starting material; R = reduced product sample; and P = PEGylated product sample.

Calibration Results							
Actual	**I**	**R**	**P**		**I**	**R**	**P**
I	**10**	**0**	**0**	**TPR**	100	100	100
R	**0**	**10**	**0**	**TNR**	100	100	100
P	**0**	**0**	**10**	**Err.**	0	0	0
Validation Results							
Actual	**I**	**R**	**P**		**I**	**R**	**P**
I	**7**	**1**	**0**	**TPR**	100	100	100
R	**0**	**6**	**0**	**TNR**	100	100	100
P	**2**	**2**	**9**	**Err.**	0	0	0

On the other hand, when samples from both lot 2 and 3 were included in the original calibration and validation sets and SVM classification was carried out. (Table 5.8) Again, all samples (n=6 from each reaction step) were correctly predicted with no error based on the calibration set (n=13 from each step) showing the good predictive power of the method even outside the well-controlled set of samples.

Table 5.8- Summary of performance results obtained by classification of EEM$_I$ using SVM including Lot 2,3 samples in both validation and calibration sets. I = IgG starting material; R = reduced product sample; and P = PEGylated product sample.

Calibration Results							
Actual	**I**	**R**	**P**		**I**	**R**	**P**
I	**13**	**0**	**0**	**TPR**	100	100	100
R	**0**	**13**	**0**	**TNR**	100	100	100
P	**0**	**0**	**13**	**Err.**	0	0	0
Validation Results							
Actual	**I**	**R**	**P**		**I**	**R**	**P**
I	**6**	**0**	**0**	**TPR**	100	100	100
R	**0**	**6**	**0**	**TNR**	100	100	100
P	**0**	**0**	**6**	**Err.**	0	0	0

Another potential application of the classification methods/EEM$_\parallel$ is the classification of Lot1 reaction samples based on the different reaction conditions. The preliminary results summarized in Table 7.14 show a correct prediction of all samples, which indicates that it might be possible to build a classification model even for very subtle changes as expected here. However, this was an attempt with a very limited number of samples in each cluster, thus further investigations are required to validate these results.

5.7. Conclusions

The initial goal of this PhD project was to evaluate if pEEM and aniso-EEM could be used for the analysis of ADCs. However, because the access to these types of molecules can be difficult particularly because of safety issues, the rational approach was to study a non-toxic model, which could serve as preliminary study to understand how a typical IgG conjugation reaction works, and what the possibilities and challenges for the analysis were. Despite not considering all the factors and their impacts in the initial experimental design, and all the challenges faced, we can consider that these experiments provided interesting insights into the potential use of pEEM for tracking IgG conjugation and assessing the reaction trends via structural/physicochemical variances.

Here the spectral changes were expected to be small not only because of the low yield and low conjugation degree, but also because of the relatively small impact of a 5kDa PEG on the big IgG.[180] This is evident when the two PEGylation systems are compared (LZ-PEG from Chapter 1 and PEG-IgG). Based on our findings, even the lower PEGylation degree in LZ (PPR=0.04) caused a bigger impact in the protein than what was observed here for rIgG, which was evident from the difference spectra. The complexity of the unpurified reaction mixture and the complex emission of IgG because of the many fluorophores in close proximity might also make it harder to assess changes caused by the conjugation.[a]

Nevertheless, we show that the changes, although small, were still significant. ROBPCA of pEEM data indicated that they were mostly due to Trp/Tyr emission fluctuations caused by varying FRET rates induced by structural changes. As for the other systems analysed, the increased RS band during reduction indicated possible formation of soluble and/or insoluble HMWS, some of which dissociated during alkylation.

Aniso-EEM maps are very sensitive to MW changes, to environmental factors affecting diffusion as well as structural and chemical changes, which also affect energy transfer rates.[120, 121] Here, because of the slightly bigger anisotropy values compared

[a] These assumptions are only based on a rough comparison of the two systems. We understand that a direct comparison would only be possible if the two systems had been subjected to the same process and analysed under the same conditions.

to the PEG-LZ system, it was possible to extract more information from aniso-EEM maps, which indicated significant photo physical changes with the reaction (as per *SimI)*. However, *aniso*-EEM measurements are very sensitive to noise and scatter and is the least robust compared to the other three measurements ($EEM_{\parallel}$, $EEM_{\perp}$, and EEM_{T})

Finally, the use of non-linear classification algorithms enabled the correct assignment of starting material, reaction intermediates, and products, which can be used for end-point determination. It was also clear from pEEM measurements were much superior to UV-Vis absorbance spectroscopy and that $EEM_{\parallel}$ measurements were the best option overall, which was also confirmed in the PEG-LZ and IgG quality screening studies. This proof of concept study has demonstrated that there is an advantage for the use of pEEM measurements as a tool for the in-situ analysis of PEGylation, or similar conjugation reactions.

Chapter 6. Conclusions

The goal of this thesis was to determine if pEEM could be used for the meaningful assessment of chemical modification reactions of proteins at different stages of the process. Here we showed that the proposed method could be successfully implemented for assessing protein quality (de Faria e Silva *et al.* [173]), assessing product variance with modification and quantifying the degree of conjugation (de Faria e Silva *et al.* [168]), which are important quality aspects of protein/conjugates. To the best of our knowledge, this is the first time that pEEM has been used for assessing these systems.

The use of EEM measurements has been reported in many studies[212, 213] and the advantages of using chemometrics for process analysis has also been discussed in many reviews.[125, 214] EEM was successfully used in combination with multivariate tools like PLS, PCA and classification algorithms in different areas such as agricultural, food and beverage industry for assessing wine, vinegar or water samples, or in the biopharmaceutical context for monitoring antigen manufacturing process and monoclonal antibody aggregation.(Table 1.6) In our laboratory, it has been applied for quantifying protein in complex media samples, for degradation studies of cell culture components and the ARMES methodology has been studied for protein structural analysis and stability studies.[119-121, 142, 198] Because the polarization adds an extra level of information with which to assess molecular size, local viscosity and fluorophores mobility, it should be more sensitive than conventional MDF for analysing complex protein-based systems.[120, 126]

The analysis of protein in solution can be challenging and demanding, and in the same way characterization of conjugation reactions and products can be difficult because of the relatively small impact on protein structure, and the complexity of the reaction mixtures.[60, 180] Different analytical methods can assess integrity, purity, size, the degree of conjugation, total protein concentration, high molecular weight species (HMWS) content, and lot-to-lot variability of proteins and/or their conjugated forms. However, even the standard techniques have many drawbacks which can preclude their use as in process analysis tools. For instance, DLS despite being non-destructive, sensitive for big changes in size (e.g. large aggregates) and easy to perform, still suffers

from low robustness, low resolution for small size changes and can be very sensitive to contamination.[78] SDS-PAGE, is relatively easy to perform but it is sample destructive and denaturing and requires a staining step. SEC is widely available, there is a relatively low cost associated with the analysis and sensitive for most applications. Despite being the gold standard technique for many applications, SEC also has many disadvantages such as the dilution step which can modify the sample, the possible interactions between matrix and column and relatively long column/equilibration times. (Table 1.4, Section 1.5) Thus, the study and use of orthogonal analytical techniques has been now more encouraged, specially fast, sensitive, high throughput and non-destructive techniques, which could be implemented within the PAT framework.[215]

pEEM measurements offer a convenient alternative to existing spectroscopic techniques for better identification of protein structural changes and variances in particle distribution/aggregate profile which can be difficult to assess using a single conventional measurement technique like SEC. The sensitive intrinsic protein emission combined with RS analysis offers a complete measurement mode for monitoring different aspects of the complete process. While the fluorescence emission component directly probes the smaller structural changes, the RS data can be used to track the formation of larger aggregated species during reactions. This was also demonstrated by Casamayou-Boucau *et al*.[216] who used the fluorescence signal to quantify oligomer content in solution, while RS band volume showed a linear correlation with the average MW of insulin oligomers in solution. The possibility of using the RS as an extra tool for assessing the sample is a compelling advantage of EEM in comparison to TSFS, which however, seemed to be more suitable for factor based analysis of IgG intrinsic fluorescence.[199]

This proof of concept study also shows some advantages in the use of pEEM for assessing PEGylation reactions in comparison to SDS-PAGE, SEC, and DLS which are commonly used for this application.[180] Here we demonstrate that, in the context of PEGylation, the combination of pEEM with simple chemometrics can be used for characterizing PEGylation reaction profiles, accurately identifying reaction endpoints, and assessing raw product variance. Furthermore, one can generate quantitative prediction models for conjugation degree as described by SEC and significantly

decrease the data acquisition time by using variable selection to exclude the non-informative areas of the spectra, which could facilitate in-process implementation.[125]

Here we demonstrated that pEEM offered advantages compared to EEM_T measurements, and $EEM_{\parallel}$ in particular was a better tool to assess changes in the systems studied here. Its fluorescence component was more sensitive to chemical and photophysical changes induced by both aggregation and conjugation and the RS band was, as expected, more sensitive to physical changes, which was also observed by Casamayou-Boucau et al.[216] In addition, because it can be implemented as a single measurement (there is no need for G factor correction), it enables the reduction of measurement time to less than 8 minutes. For use in on or in-line PAT, the longer acquisition time and poorer SNR of the full (four spectra) pEEM measurements compared to unpolarized EEM are the main limitations. We are continuing to develop and enhance this pEEM measurement methodology to generate better quality data with much shorter measurement times to enable its use in High Throughput Screening applications in both academic and industrial applications.

Despite the fact that fluorescence is not a widespread technique for assessing chemical modification processes (it is mostly used to assess the effect of conjugation on higher order structure and stability[43, 60]), we believe that this study could open doors for the application of fluorescence-based reaction monitoring to other systems. pEEM represents an interesting alternative for assessing this type of reactions because it does not rely on the spectroscopic properties of the surrogate, but on changes in the parent proteins with attachment. Because of this, it seems reasonable to think that the same technique could be applied to other protein conjugation systems such as the marketed conjugates mentioned in Table 1.2.

This thesis discussed some preliminary results on the use of pEEM for assessing protein chemical modification reactions, but there is still a lot of possibilities to explore. Lysozyme PEGylation was an interesting model to study conjugation because of it is relatively simple. Thus, one interesting addition to this work would be to use the same system (PEG and LZ) to explore the effect of different conjugation strategies, and conditions. For instance, we could attempt conjugation via disulphide bonds[a], to assess the impact of the reduction step on a smaller protein before studying reduction

[a] Lysozyme has four disulphide bonds[7] (Figure 1.1)

on a bigger protein like IgG. (Section 1.4) Another possible addition to this work, would be to carry out more experiments further varying the level of conjugation (PPR) and/or isolating the individual species by purification[217, 218] to try and better understand the potential applications and limitations of pEEM for assessing the degree of conjugation. In a similar way, it could also be useful to investigate the possibility of monitoring a conjugation reaction of a different active PEG such as aldehyde or NHS derivatives, because, as we showed, quenching by the leaving group from pNPC-PEG limits the use of fluorescence. These two reagents are also more commonly used among the marketed PEGylated proteins.[43, 147]

One interesting addition to the IgG quality screening study would be to develop a similar study using a mAb instead of pAb, and different stress conditions to study the impact of different types of aggregates[219] on the emission profile. As we discussed, an increased emission in the long wavelength region was a recurring change in the EEM of stressed samples, and this is often described as related to the deep blue autofluorescence. However, there is still very little and inconclusive information about this effect on the literature and its source is still controversial.[187, 188] Thus, if we want to better understand and explain this effect, the obvious approach would be to carry out a study designed for this purpose, varying proteins and conditions and using orthogonal techniques to support the findings. Also, an interesting alternative would be to study the application of pEEM for assessing protein quality during the upstream process, by using a system consisting of protein/stressed protein in cell culture media instead of buffer.[220]

The PEGylation of IgG is probably the work that we have more to develop on. The idea of creating a non-toxic model system is still the most rational approach considering the limitations of working with ADCs, but this should be done with a better design. A fundamental step for proving the efficacy of an orthogonal technique is the ability to compare the results with the outcomes of standard tools. Thus, preferably, we should repeat this work with a more adequate protein-surrogate model to enable the characterization and purification using the techniques that we now have available in the laboratory (DLS, SEC, SDS-PAGE, UV-Vis). With this in mind, we have designed another model using a monoclonal IgG1[b] and a smaller surrogate with

[b] A batch of monoclonal IgG1 donated early in 2020 will be used for the follow-on studies.

a chromophore on the structure, which should better mimic the characteristics of an ADC and facilitate the characterization and purification. We selected a non-toxic substitute for the cytotoxic drugs in ADC, but as opposed to previous studies,[85, 221] we opted for a non-fluorescent small molecule to avoid complications in the analysis caused by interference with protein fluorescence. This is a study that we plan to carry out soon.

In case we decided to work again with the PEG-IgG conjugation, ideally, we should use a mAb to decrease the variability in the system and a human or humanized IgG to approximate the work to the conditions being used in the industrial context (for ADCs for example).[222] Also, it would be important to design some screening experiments to determine the best conditions (e.g. Ellman's experiment, Section 7.8.2), develop a purification protocol and use a more informative technique than SDS-PAGE and SEC for the characterization of final products (a good option in this case would be mass spectrometry, Section 1.5).

For all cases, we would certainly benefit from using a CCD based spectrometer like the Aqualog® system from Horiba. One of the main differences between photomultiplier tube (PMT) (used in this work) and CCD detectors is that the first measures a single point in the spectrum at a time, while the latter can operate in multi-channel mode, which means that the complete spectrum can be measured in a single scan. [223-225] With this, we should be able to significantly reduce analysis time to <1 minute, collect more data on a same timeframe and obtain an average spectrum, thus also improving data quality in terms of SNR and resolution. This is particularly important for the noisy RS and aniso-EEM data. Faster measurements would also enable the collection of nearly real-time data and, and thus facilitate the use of EEM measurements as a PAT tool. The Acqualog® system also offers the advantage of simultaneous acquisition of absorbance and EEM spectra and automated multiple sample measurement, which would also simplify experiments of reaction monitoring for example.[c]

[c] Information obtained from Aqualog brochure https://static.horiba.com/fileadmin/Horiba/Products/Scientific/Molecular_and_Microanalysis/Aqualog/Aqualog-Nov13.pdf

Chapter 7. Appendices

7.1. IgG structure

The heavy chain sequence determines the Ab subtype or class. The Greek letters gamma (γ), alpha (α), mu (μ), delta (δ), and epsilon (ε), represents, respectively, the heavy sequences from the five major classes of. The type G is the most abundant found in the serum (70-75%), followed by IgA (15%), IgM (10%) and IgD and IgE (less than 1%).[13] (Figure 7.1) While IgG, IgE and IgD exist in monomeric forms, IgM consists of five subunits of the 2-heavy and 2-light chains and IgA can exist either as a monomer or dimer (rarely a trimer). IgM and polymeric IgA have an additional 15-kD polypeptide called J-chain, which has the function of attaching the different subunits.

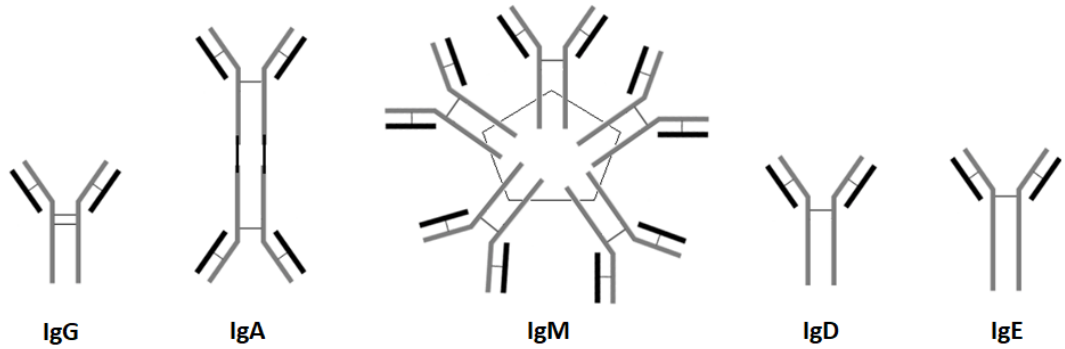


Figure 7.1- Schematic representation of the five classes of immunoglobulins found in human serum underlying structural differences: IgD, IgE and IgG are found in monomeric form, IgA is usually encountered as dimer and IgM in the pentameric form.

While each IgG molecule has around 150kDa, the molecular weight of IgD and IgE is 170kDa and 18k0Da respectively. IgM and IgA will have different MW depending on its polymeric form: the most common form of IgA, a monomer, weights around 160kDa and of IgM, a pentamer, around 900kDa.[226]

The heavy chains in IgA, IgD, and IgG contain, in addition to the constant region domains, the hinge region between C_{H1} and C_{H2}, and the number of amino acids and disulphide bonds in the hinge region varies in the different classes and subclasses of immunoglobulins.[12] Although IgM and IgE structures lack a hinge region, they have an additional domain of 110 amino acids (C_{H2}/ C_{H2}) that has hinge-similar features. Therefore, instead of the sequence C_{H1} / Hinge/ C_{H2} / C_{H3}, IgM and IgE heavy chains have an extra-domain called C_{H4} and are arranged in a C_{H1} / C_{H2} / C_{H3} / C_{H4} form. [227]

The main differences between the four isoforms of IgG ($IgG_{1,2,3}$, and $_4$) are the heavy chain type and sizes as well as the pattern of linkage in the hinge region.[12] The two heavy chains are attached in the hinge region via a number of disulphide bonds which varies according to the isotype: 2 bonds for IgG_1 and IgG_4, 4 bonds for IgG_2 and 11 for IgG_3. Also, while the disulphide bond in IgG_1 connects the last cysteine residue of the light chain and the fifth cysteine residue of the heavy chain, the attachment in IgG of 2, 3 and 4 isotypes occurs between the last cysteine residue of the light chain and the third cysteine residue of the heavy chain. [16] (Figure 7.2)

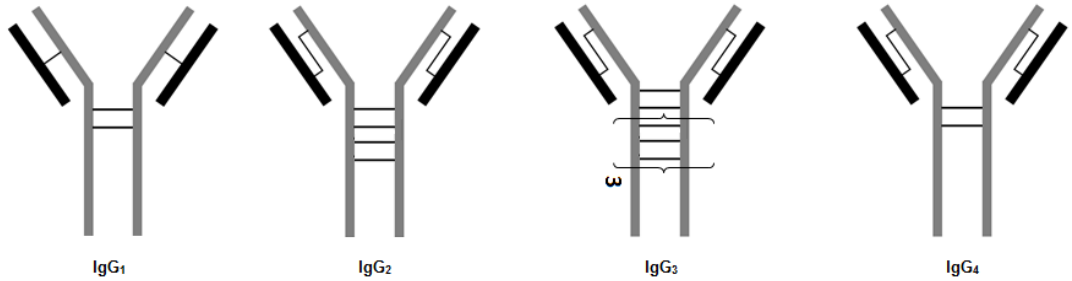


Figure 7.2- Schematic showing the main structural differences between IgG of subtypes 1,2,3 and 4. The main differences are found in hinge region.

Each light chain in all Igs consists of a sequence of around 120 amino acids and a total of 25kDa. The aminoacids sequence of the L constant domain (CL) defines two types of light chain: lambda (λ) and kappa (κ) and variances in the ratio λ: κ are responsible for the differences among Igs from distinct species (e.g. human, rabbit). Whereas, in rabbit, for instance, the average κ to λ ratio is 90:10, in humans it is 67:33. [228] Variance on domains sizes, types and chain ratios are the determinant factors for inter and intra-species variations.

Table 7.1 shows the difference between classes and subclasses of immunoglobulins from different species. Variance on domains sizes, types and chain ratios are the determinant factors for inter and intra-species variations.

Table 7.1- Summary of the main characteristics differentiating IgGs from different species: ratio between lambda (λ) and kappa (κ) in the light chain, heavy chain composition and different classes and subclasses found in each species.

Specie	Light chain ratio (κ/λ) %	Class	Sub-class	H-chain type
Human	67/33	IgG	IgG_1	γ_1
			IgG_2	γ_2
			IgG_3	γ_3
			IgG_4	γ_4
		IgM	-	μ
		IgA	IgA_1	α_1
			IgA_1	α_2
		IgD	-	δ
		IgE	-	ε
Rabbit	90/10	IgG	-	γ
		IgA	-	α
		IgM	-	μ
Sheep	1/99	IgG	-	γ
		IgA	-	α
		IgM	-	μ
Mouse	99/1	IgG	IgG_1	γ_1
			IgG_{2a}	γ_{2a}
			IgG_{2b}	γ_{2b}
			IgG_3	γ_3
		IgM	-	μ
		IgA	IgA_1	α_1
			IgA_1	α_2
		IgD	-	δ
		IgE	-	ε
Bovine	1/99	IgG	IgG_1	γ_1
			IgG_2	γ_2
		IgA	-	α
		IgM	-	μ
		IgD	-	δ
		IgE	-	ε

7.2. Certificate of analysis of lysozyme and rIgG

These certificates of analysis are provided here to allow for the work to be replicated in future by identifying if there is a major change in lot quality.

Certificate of Analysis – lysozyme from chicken egg white

SIGMA-ALDRICH®

sigma-aldrich.com

3050 Spruce Street, Saint Louis, MO 63103, USA
Website: www.sigmaaldrich.com
Email USA: techserv@sial.com
Outside USA: eurtechserv@sial.com

Product Name:

Certificate of Analysis

Lysozyme from chicken egg white – lyophilized powder, protein ≥90 %, ≥40,000 units/mg protein

Product Number:	**L6876**
Batch Number:	**SLCC4285**
Brand:	SIGMA
CAS Number:	12650-88-3
MDL Number:	MFCD00131557
Storage Temperature:	Store at -20 °C
Quality Release Date:	19 APR 2019
Recommended Retest Date:	OCT 2022

Test	Specification	Result
Appearance (Color)	White to Off White	White
Appearance (Form)	Powder	Powder
Solubility (Color)	Colorless	Colorless
Solubility (Turbidity) 10 mg/mL, H2O	Clear to Slightly Hazy	Clear
% Protein (UV)	≥ 90	95
units/mg Protein One unit will produce a change in A450 of 0.001 per min at pH 6.24 at 25 deg C, using a suspension of Micrococcus lysodeikticus as substrate, in a 2.6 mL reaction mixture (1 cm light path).	≥ 40000	40382

Carolyn Baird, Supervisor
Quality Assurance
St. Louis, Missouri US

Sigma-Aldrich warrants, that at the time of the quality release or subsequent retest date this product conformed to the information contained in this publication. The current Specification sheet may be available at Sigma-Aldrich.com. For further inquiries, please contact Technical Service. Purchaser must determine the suitability of the product for its particular use. See reverse side of invoice or packing slip for additional terms and conditions of sale.

Version Number: 1 Page 1 of 1

Certificate of Analysis – IgG from rabbit serum

SIGMA-ALDRICH®

sigma-aldrich.com

3050 Spruce Street, Saint Louis, MO 63103, USA
Website: www.sigmaaldrich.com
Email USA: techserv@sial.com
Outside USA: eurtechserv@sial.com

Certificate of Analysis

Product Name:
IgG from rabbit serum – reagent grade, ≥95% (SDS-PAGE), essentially salt-free, lyophilized powder

Product Number:	**I5006**
Batch Number:	**SLBM2617V**
Brand:	SIGMA
MDL Number:	MFCD00165673
Storage Temperature:	Store at 2 - 8 °C
Quality Release Date:	26 MAR 2015
Recommended Retest Date:	MAR 2018

Test	Specification	Result
Appearance (Color)	White to Off-White	White
Appearance (Form)	Powder	Powder
Solubility (Turbidity) When reconstituted to 10 mg/mL in 0.85% NaCl	Clear to Hazy	Slightly Hazy
Water (by Karl Fischer)	≤ 10 %	3 %
Sodium (Na) By ICP	≤ 1.0 %	< 0.2 %
ID by Immuno-Electrophoresis Single major arc of precipitation in the gamma region versus anti-rabbit whole serum and anti-rabbit IgG. Minor gamma arcs may be present versus anti-rabbit whole serum	Pass	Pass
Purity by SDS-PAGE	≥ 95 %	98 %
Chapter 4(C) - Untreated Product meets European Union requirements for production of technical blood products.	----------	----------

SIGMA-ALDRICH®

sigma-aldrich.com

3050 Spruce Street, Saint Louis, MO 63103, USA
Website: www.sigmaaldrich.com
Email USA: techserv@sial.com
Outside USA: eurtechserv@sial.com

Certificate of Analysis

Product Name:
IgG from rabbit serum – reagent grade, ≥95% (SDS-PAGE), essentially salt-free, lyophilized powder

Product Number:	**I5006**
Batch Number:	**SLBZ5214**
Brand:	SIGMA
MDL Number:	MFCD00165673
Storage Temperature:	Store at 2 - 8 °C
Quality Release Date:	05 DEC 2018
Recommended Retest Date:	DEC 2021

Test	Specification	Result
Appearance (Color)	White to Off White	White
Appearance (Form)	Powder	Powder
Solubility (Turbidity) When reconstituted to 10 mg/mL in 0.85% NaCl	Clear to Hazy	Hazy
Water (by Karl Fischer)	≤ 10 %	3 %
Sodium (Na) By ICP	≤ 1.0 %	0.5 %
ID by Immuno-Electrophoresis Single major arc of precipitation in the gamma region versus anti-rabbit whole serum and anti-rabbit IgG. Minor gamma arcs may be present versus anti-rabbit whole serum	Pass	Pass
Purity by SDS-PAGE	≥ 95 %	97 %
Chapter 4(C) - Untreated Product meets European Union requirements for production of technical blood products.	----------	----------

Sigma-Aldrich warrants, that at the time of the quality release or subsequent retest date this product conformed to the information contained in this publication. The current Specification sheet may be available at Sigma-Aldrich.com. For further inquiries, please contact Technical Service. Purchaser must determine the suitability of the product for its particular use. See reverse side of invoice or packing slip for additional terms and conditions of sale.

Version Number: 1

Page 1 of 2

SIGMA-ALDRICH®

sigma-aldrich.com

3050 Spruce Street, Saint Louis, MO 63103, USA
Website: www.sigmaaldrich.com
Email USA: techserv@sial.com
Outside USA: eurtechserv@sial.com

Certificate of Analysis

Product Name:
IgG from rabbit serum – reagent grade, ≥95% (SDS-PAGE), essentially salt-free, lyophilized powder

Product Number:	**I5006**
Batch Number:	**SLBP7449V**
Brand:	SIGMA
MDL Number:	MFCD00165673
Storage Temperature:	Store at 2 - 8 °C
Quality Release Date:	18 JAN 2016
Recommended Retest Date:	JAN 2019

Test	Specification	Result
Appearance (Color)	White to Off-White	White
Appearance (Form)	Powder	Powder
Solubility (Turbidity) When reconstituted to 10 mg/mL in 0.85% NaCl	Clear to Hazy	Slightly Hazy
Water (by Karl Fischer)	≤ 10 %	4 %
Sodium (Na) By ICP	≤ 1.0 %	0.4 %
ID by Immuno-Electrophoresis Single major arc of precipitation in the gamma region versus anti-rabbit whole serum and anti-rabbit IgG. Minor gamma arcs may be present versus anti-rabbit whole serum	Pass	Pass
Purity by SDS-PAGE	≥ 95 %	96 %
Chapter 4(C) - Untreated Product meets European Union requirements for production of technical blood products.	----------	----------

SIGMA-ALDRICH®

sigma-aldrich.com

3050 Spruce Street, Saint Louis, MO 63103, USA
Website: www.sigmaaldrich.com
Email USA: techserv@sial.com
Outside USA: eurtechserv@sial.com

Certificate of Analysis

Product Name:
IgG from rabbit serum – reagent grade, ≥95% (SDS-PAGE), essentially salt-free, lyophilized powder

Product Number:	**I5006**
Batch Number:	**SLBW8687**
Brand:	SIGMA
MDL Number:	MFCD00165673
Storage Temperature:	Store at 2 - 8 °C
Quality Release Date:	04 APR 2018
Recommended Retest Date:	APR 2021

Test	Specification	Result
Appearance (Color)	White to Off White	White
Appearance (Form)	Powder	Powder
Solubility (Turbidity) When reconstituted to 10 mg/mL in 0.85% NaCl	Clear to Hazy	Hazy
Water (by Karl Fischer)	≤ 10 %	2 %
Sodium (Na) By ICP	≤ 1.0 %	0.2 %
ID by Immuno-Electrophoresis Single major arc of precipitation in the gamma region versus anti-rabbit whole serum and anti-rabbit IgG. Minor gamma arcs may be present versus anti-rabbit whole serum	Pass	Pass
Purity by SDS-PAGE	≥ 95 %	99 %
Chapter 4(C) - Untreated Product meets European Union requirements for production of technical blood products.	----------	----------

Sigma-Aldrich warrants, that at the time of the quality release or subsequent retest date this product conformed to the information contained in this publication. The current Specification sheet may be available at Sigma-Aldrich.com. For further inquiries, please contact Technical Service. Purchaser must determine the suitability of the product for its particular use. See reverse side of invoice or packing slip for additional terms and conditions of sale.

7.3. Protein concentration calculated for the PEG-rIgG study

Sample protein concentration calculated using absorbance at 280 nm corrected for scatter contribution. The calculations for correction are detailed in Section 7.5

Table 7.2- Summary of starting material lots and amount of TCEP used for each PEG-rIgG reaction, and calculated concentrations for each rIgG starting material (rIgG SM), reduction (Red-rIgG) and alkylation product (PEG-rIgG) samples using corrected absorbance values to scatter contribution.

Stock solution	Reaction	SM Lot condition	Molar excess TCEP	Concentration (mg.mL^{-1})		
				rIgG SM	Red-rIgG	PEG-rIgG
IgG stock soln.1	R01	Lot SLBP7449V	1.5	1.02	1.00	0.95
	R02			0.97	1.00	0.94
	R03[a]			1.02	1.00	0.95
IgG Stock soln. 2	R04		3.0	0.96	0.93	0.88
	R05			0.96	0.93	0.88
	R06			0.95	0.93	0.88
IgG Stock soln. 3	R07		3.0	0.96	0.95	0.91
	R08			0.96	0.95	0.90
IgG Stock soln. 4	R09		1.5	1.00	0.97	0.92
	R10			0.99	0.98	0.92
	R11			1.00	0.98	0.92
IgG Stock soln. 5	R12		3.0	0.98	0.96	0.91
	R13		1.5	0.99	0.96	0.91
	R14			0.98	0.96	0.92
Mean ± stdev				0.98 ± 0.02 (RSD 2.31%)	0.96 ± 0.03 (RSD 2.81%)	0.91 ± 0.02 (RSD 2.64%)
IgG stock soln. 6	R15	SLBP7449V	3.0	0.98	0.97	0.92
	R16	SLBP7449V (1min vortex)		1.00	0.97	0.92
	R17	SLBP7449V (2min vortex)		0.99	0.96	0.91
Mean± stdev				0.99 ± 0.01 (RSD 0.77%)	0.96 ± 0.02 (RSD 0.68%)	0.92 ± 0.00 (0.22%)
IgG stock soln. 7 (Lot3)	R18	SLBM2617V	3.0	0.83	0.79	0.74
	R19			0.83	0.79	0.74
Mean± stdev				0.83 ± 0.00 (RSD 0.01%)	0.79 ± 0.00 (RSD 0.07%)	0.74 ± 0.00 (RSD 0.22%)

[a] rIgG R03 was an outlier probably due to the UV-Visible measurement and was taken out from the data for the calculation of mean and stdev of IgG starting material.

7.4. Ellman's reagent calibration curve

The quantification of thiols is based on the reaction of the Ellman's reagent (5,5'-dithio-*bis*-(2-nitrobenzoic acid, DTNB) with a thiol to form a thiol-reagent conjugate with a concomitant release of one 5-thio-2-nitrobenzoic acid (TNB) (per available thiol). (Figure 7.3C) The quantification can be done either via the extinction coefficient of TNB at 412 nm or the correlation to a standard curve of known sulfhydryl concentration.[148, 149] The latter was the choice here, and the absorbance spectra and standard curve are shown in Figure 7.3A/B.

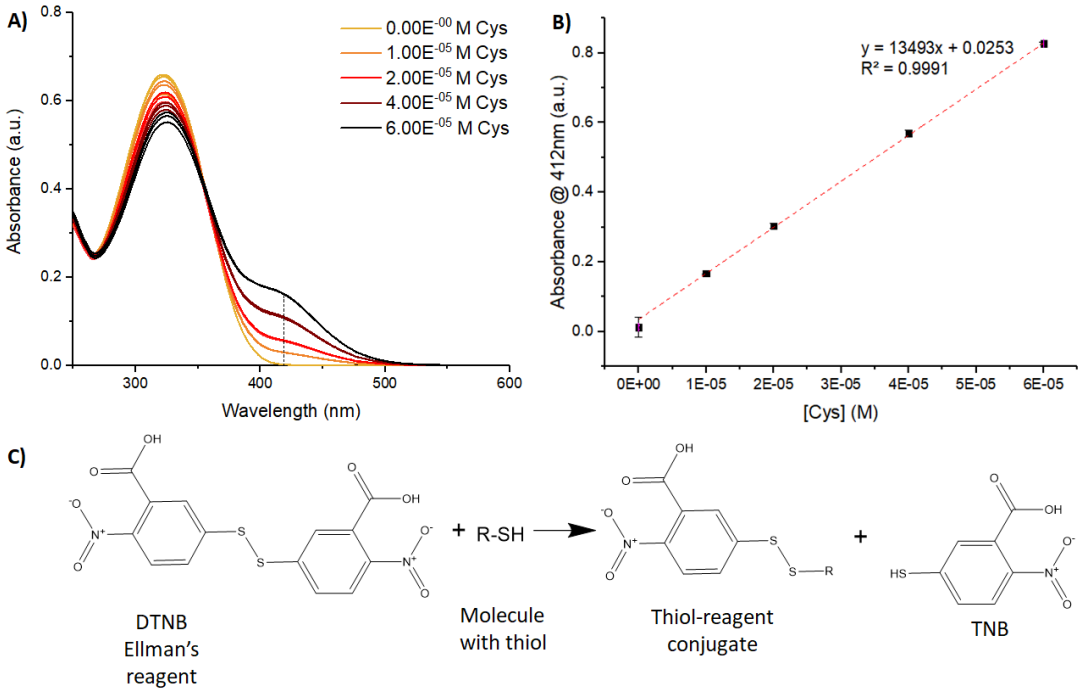


Figure 7.3- (A) Absorbance spectra of cysteine standard at various concentrations, (B) standard curve used for the quantification of free sulfhydryls by the Ellman's method and schematic of chemical reaction between the Ellman's reagent DTNB and the molecule with thiol.[148]

7.5. Protein concentration calculation and correction

Protein concentration was calculated using absorbance at 280 nm and protein molar extinction coefficient ($\boldsymbol{\varepsilon}$) at 280 nm ($1.35\ mg/mL)^{-1}cm^{-1}$) reported in the literature:[229]

$$\mathrm{C\ (mg/mL)} = \frac{\mathrm{A_{280}}}{\varepsilon\ ((\mathrm{mg/mL})^{-1}\mathrm{cm}^{-1})\cdot 1\mathrm{cm}} \qquad \text{Equation 7.1}$$

Prior to the calculation of concentration, absorbance values at 280 nm were corrected for scatter contribution. Absorbance at 320 and 340 nm (where proteins are

not supposed to absorb) were used to extrapolate and find the scatter contribution at 280 nm, as described in equation 7.2 and 7.3. Corrected protein absorbance (A_{280}) was then calculated by the subtraction of $A_{scatter,280}$ from the apparent absorbance at 280 nm following equation 7.4.[230]

$$A_{scatter,280} = 10^{\,m\,.\,\log(280) + b} \qquad \text{Equation 7.2}$$

$$\text{where,}\quad m = \frac{\log(A_{340}) - \log(A_{320})}{\log(340) - \log(320)} \quad \text{and} \quad b = \frac{\log(A_{340}) - m\,(A_{320})}{\log(340) - \log(320)} \qquad \text{Equation 7.3}$$

$$A_{280} = A_{apparent,280} - A_{scatter,280} \qquad \text{Equation 7.4}$$

7.6. Additional information: LZ-PEG study

7.6.1. Control experiments

Control experiments were carried to first quantify residual pNP leaving group in the purified PEG-LZ products and evaluate its effect on the fluorescence spectra (Figure 7.4), and second assess the effect of free PEG (which we had not been able to remove with the purification strategy adopted) in the fluorescence properties of LZ (Figure 7.5).

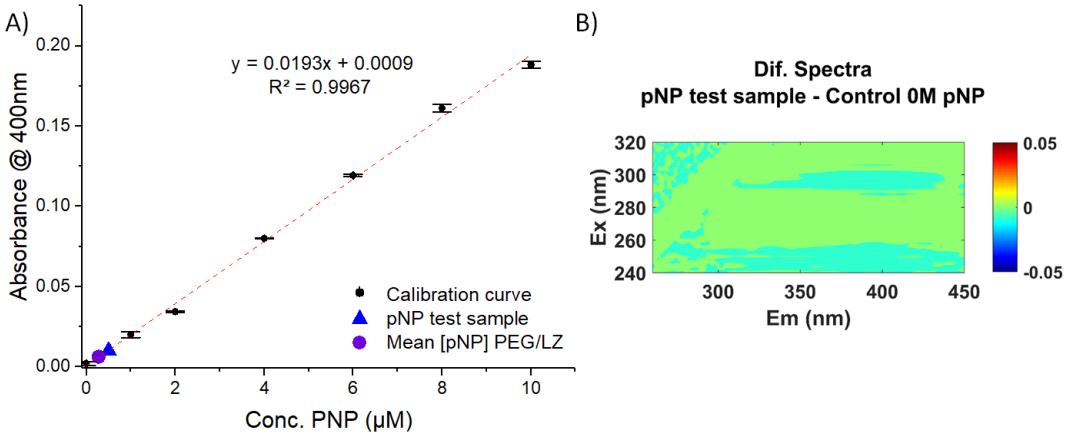


Figure 7.4 - Control experiment carried out to check for the presence of the p-nitrophenol (pNP) leaving group in the purified product samples. (A) Results indicated that there was a minimal concentration of pNP in all purified product samples (<1µM, purple in the calibration curve). (B)This excessive pNP groups did not affect the fluorescence properties of LZ as seen by the difference spectra calculated using a test sample with known concentration of pNP (blue in the calibration curve) as follows: $EEM_{\| Dif.} = EEM_{\| Mean\ Test} - EEM_{\| LZ}$.

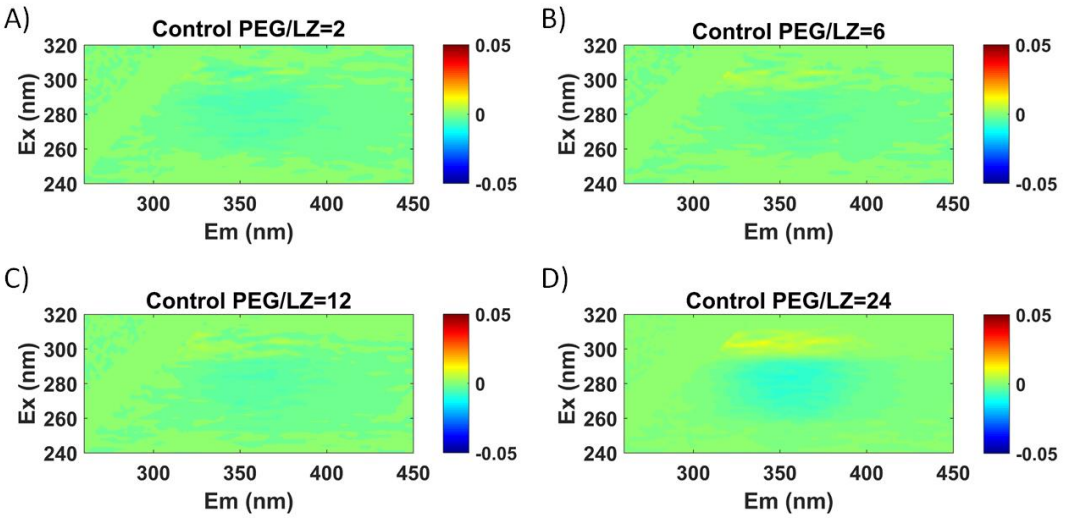


Figure 7.5- Control experiment to check whether the presence of excess free PEG in solution would change the photophysical properties of LZ (Note that the purification strategy adopted here did not allow for the complete removal of the excess PEG reagent in solution). Spectral changes were calculated as the difference between the normalized EEM‖ data from lysozyme:PEG mixtures and the lysozyme in the reaction buffer at different molar excess of non-reactive 5kDa PEG ($EEM_{‖\ Dif.} = EEM_{‖\ Mean\ LZ+PEG} - EEM_{‖\ Mean\ LZ}$): (A) 2 molar excess, (B) 6, (C)12 and (D) 24 , which was the highest amount of PEG added for the reactions.

7.6.2. SEC measurement error

The pooled relative standard deviation (RSD_P, Equation 7.5/Table 7.3*)* gives an indication of the measurement error for each peak of the chromatograms taking into account all samples and respective number of injections (RSD_P <3%). However, a more appropriate approach is to consider the error propagation with the calculation of the parameters PPR and Total PEG-LZ% and calculate the combined uncertainty value for each parameter (Equation 7.6 and Equation 7.7 respectively).[231] Using this approach, the errors were 7.2 and 5.2% for PPR and PEG-LZ% respectively, which would be the minimum error for the prediction of the degree of PEGylation.

$$RSD_P = \sqrt{(n_i - 1).RSD_i \Big/ (n_i - 1)} \qquad \text{Equation 7.5}$$

Where n is the number of injections for each sample (n=3 for all samples) and RSD is the relative standard deviation of the %AUC calculated for each *i* sample.

Table 7.3- Pooled relative standard deviation (RSD_P) calculated for each peak of the SEC chromatograms obtained for the various purified PEG-LZ solutions.

Peak	%LZ	%1PEG-LZ	%2PEG-LZ	%3 PEG-LZ	% Multi PEG-LZ
RSD_P	1.33	1.44	2.59	1.89	2.56

The RSD_P was then used to calculate the combined uncertainty for PPR and Total PEG-LZ% as follows:

$$\text{Uncertainty}_{PPR} = \sqrt{(RSD_{P\,\%\,1PEG-LZ}\,\text{x1})^2 + (RSD_{P\,\%\,2PEG-LZ}\,\text{x2})^2 + (RSD_{P\,\%\,3PEG-LZ}\text{x3})^2 + (RSD_{P\,\%Multi\ PEG-LZ}\text{x4})^2}$$

Equation 7.6

$$\text{Uncertainty}_{Total\ PEG-LZ\,\%} = \sqrt{(RSD_{P\,\%\,1PEG-LZ})^2 + (RSD_{P\,\%\,2PEG-LZ})^2 + (RSD_{P\,\%\,3PEG-LZ})^2 + (RSD_{P\,\%Multi\ PEG-LZ})^2}$$

Equation 7.7

7.6.3. u-PLS model results of individual % of PEGylated species

u-PLS predictive models were built for the quantification of the individual species concentration and results are summarized in Table 7. The % of individual species was defined as a percentage of the area in the SEC data, e.g. %2PEG $=(AUC_{peak3}/AUC_{total})\times100$).

Table 7.4- Summary of u-PLS modelling results obtained for the quantification of individual % of PEGylated species (% of 0,1,2,3, and 4PEG-LZ) using UV-Vis, $EEM_{\perp}$, $EEM_{\parallel}$, and EEM_{T}. The REP parameter calculated for each model, which was the main parameter used for assessing model quality is highlighted in yellow.

		UV-Vis	$EEM_{\perp}$	$EEM_{\parallel}$	EEM_{T}
% 0PEG-LZ (or % LZ SM)	**RMSECal**	7.95	4.71	6.46	6.87
	RMSECV	21.50	10.10	10.58	11.62
	RMSEPred	15.75	16.05	11.07	15.52
	REP (%)	**48**	**49**	**34**	**48**
	R^2 Cal	0.95	0.98	0.97	0.96
	R^2 CV	0.69	0.92	0.91	0.90
	R^2 Pred	0.79	0.77	0.93	0.81
% 1PEG-LZ	**RMSECal**	11.17	7.87	8.53	10.02
	RMSECV	19.18	16.79	14.78	17.33
	RMSEPred	17.83	20.52	15.49	19.70
	REP (%)	**58**	**67**	**51**	**64**
	R^2 Cal	0.54	0.77	0.73	0.63
	R^2 CV	0.02	0.06	0.24	0.04
	R^2 Pred	0.12	0.12	0.25	0.22
% 2PEG-LZ	**RMSECal**	5.53	3.05	3.10	3.31
	RMSECV	9.85	3.93	4.17	3.88
	RMSEPred	7.68	4.00	5.80	4.52
	REP (%)	**39**	**20**	**30**	**23**
	R^2 Cal	0.85	0.95	0.95	0.95
	R^2 CV	0.61	0.92	0.91	0.93
	R^2 Pred	0.72	0.91	0.85	0.90
% 3-PEG-LZ	**RMSECal**	4.27	3.46	3.16	4.24
	RMSECV	5.58	6.33	5.55	6.10
	RMSEPred	5.03	5.60	5.73	6.33
	REP (%)	**42**	**47**	**48**	**53**
	R^2 Cal	0.91	0.94	0.95	0.91
	R^2 CV	0.85	0.80	0.85	0.82
	R^2 Pred	0.88	0.85	0.84	0.88
% Multi-PEG-LZ	**RMSECal**	2.71	2.35	1.97	2.75
	RMSECV	3.38	4.23	3.52	4.03
	RMSEPred	3.44	3.65	3.57	4.37
	REP (%)	**68**	**72**	**70**	**86**
	R^2 Cal	0.86	0.89	0.92	0.85
	R^2 CV	0.78	0.65	0.76	0.68
	R^2 Pred	0.86	0.86	0.84	0.84

7.7. Additional information: IgG quality screening study

7.7.1. Results of control experiments

A solution of L-Tryptophan (prepared in PBS/EDTA buffer pH=7.0) with an equivalent concentration to that of the protein samples (1 g/L) was used to assess intrinsic instrumental and experimental conditions (e.g. different cuvettes, place in cuvette holder, days, etc) variability.[b] For this, the EEM‖ spectra of the Trp solution was measured in triplicate, on different days, with different cuvettes, and positions in the cuvette holder. The figure below shows the standard deviation attributed to the intrinsic measurement error coming from both experiment and instrument.

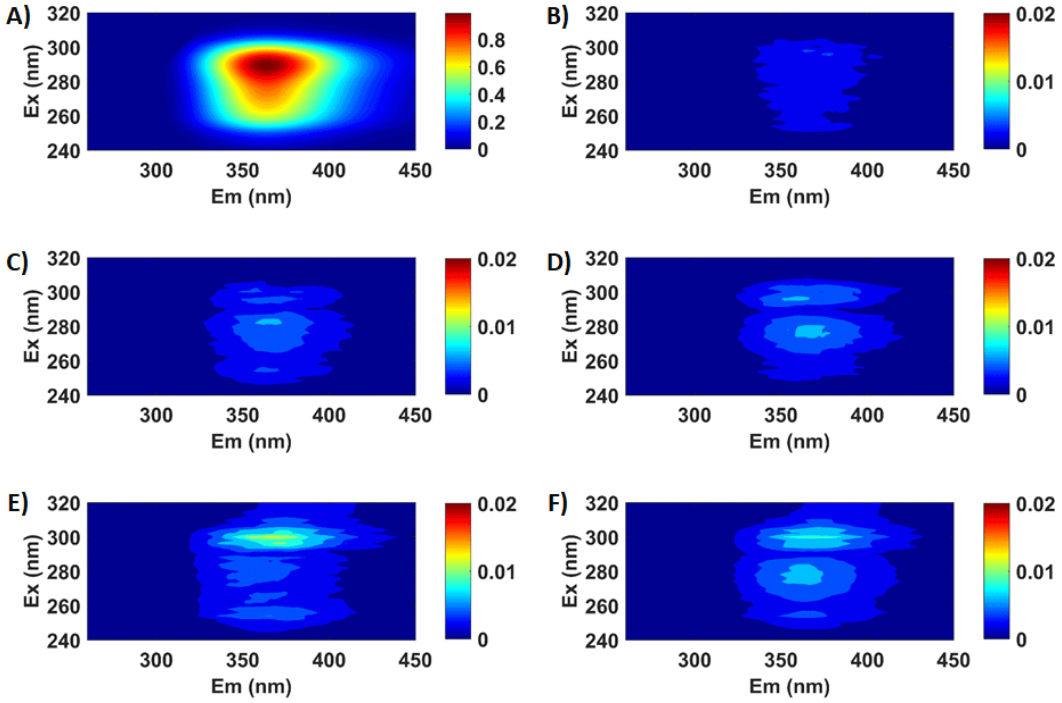


Figure 7.6- (A) Plot of mean signal obtained for Trp solutions collected with the different experimental conditions and standard deviation plots calculated from solutions measured in replicate using both the same cuvette and place in the holder (B, n=12), different cuvettes (C, n=12), different places in the cuvette holder (D, n=12), different days (E, n=12) and all the data in a combined dataset (F, n=39).

We believe that the standard deviation calculated from the dataset containing all sources of variation would be more representative of the error encountered in the rIgG experiment because all these factors were varied during the analysis. In any case, the experimental measurement error was lower than the variability of the IgG samples

[b] The Trp solutions were more intense than the rIgG samples so give a slightly optimistic result. For the unstressed samples which had higher intensity the data is in good agreement, but for the stressed samples where emission intensity was significantly weaker this test solution is not ideal because the maximum intensities are significantly different (Trp signal is >2× more intense than the rIgG).

which is due to real variation in the protein solutions. Furthermore, we measured the instrumental sensitivity to verify that the regions of high spectral variability in the standard deviation plots ($\lambda_{ex}/\lambda_{em}$>300/400 nm) were not due to noise. The average signal calculated over the whole map for the different rIgG datasets was higher than the LOR[c] calculated using 10 blank solutions (LOR=10 × stdev. of 10 blanks).[232]

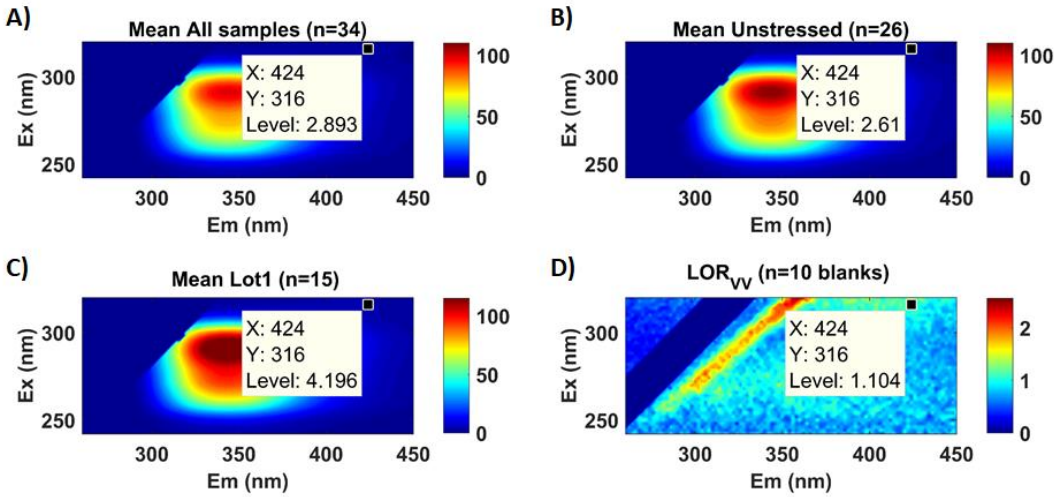


Figure 7.7: Plot of mean signal calculated for each of the rIgG datasets (EEM$_{\parallel}$): all samples (A), Lot1,2 and 4 unstressed samples (B) and lot1 samples only (C). LOR (D) calculated using 10 blank solutions measured using same instrumental settings shows that the observed variation in the datasets is higher than the smallest quantity of analyte that can be quantified (LOR).[232]

7.7.2. Comparison of Fl-AI calculated using different polarized emission measurements

Two Trp solutions (in water) of different concentrations were used here as a negative controls for comparing the different Fl-AI values calculated using the parallel polarized (Fl-AI$_{\parallel}$), perpendicular polarized (Fl-AI$_{\perp}$), and the total, or unpolarized (Fl-AI$_T$) emission spectra extracted from the EEM measurements. Because Fl-AI$_{\parallel}$ values showed the biggest differences in the protein solutions compared to the control Trp solutions, it was determined to be the most sensitive to particle content.

Table 7.5- Fl-AI values calculated using perpendicular, parallel, and unpolarized data of various rIgG starting materials in comparison to two aqueous Trp solutions of two concentrations.

rIgG Stock Solution	Fl-AI$_{\perp}$	Fl-AI$_{\parallel}$	Fl-AI$_T$
Trp (0.01mg/mL)	0.002 ± .000	0.016 ±0.001	0.006 ±0.000
Trp (0.02mg/mL)	0.001 ±0.000	0.010 ± .001	0.003 ±0.000

[c] Limit of reporting (LOR) is the smallest amount of analyte that can be quantified (here, by fluorescence intensity).[232]

soln. 1 (rIgG1,2,3)	0.007 ±0.001	0.186 ± .009	0.075 ±0.003
soln. 2 (rIgG4,5,6)	0.008 ±0.001	0.198 ± .012	0.080 ±0.004
soln. 3 (rIgG7,8, 9)	0.008 ±0.001	0.207 ± .014	0.084 ±0.007
soln. 4 (rIgG10,11,12)	0.008 ±0.001	0.207 ±0.013	0.085 ±0.006
soln. 5 (rIgG13,14,15)	0.008 ±0.001	0.211 ±0.010	0.086 ±0.004
Overall rIgG (n=15)	0.008 ±0.001	0.202 ± 0.014	0.082. ±0.006

7.7.3. u-PLS model for quantification of protein aggregation: results and optimization

Here we summarize the results of u-PLS models for the prediction of % of aggregates as obtained from SEC. The tables show the quality parameters of models using all rIgG samples (Table 7.6) and a reduced sample set (excluding stressed samples.[d] (Table 7.7)

Table 7.6- Summary of u-PLS model quality parameters using a variety of different pre-processing[e] and variable selection procedures. These models used all rIgG samples: Data selected (Kennard stone algorithm) with the validation dataset comprising of 30% of the samples (n=6). The models shown in Figure 4.8 are highlighted here in yellow. These were the best results for quantification of aggregation based on RMSE values and R^2.

Pre-processing	R^2 Cal	R^2 CV	R^2 Pred	RMSE Cal	RMSE CV	RMSE Pred
None	**0.92**	**0.88**	**0.96**	**0.96**	**1.15**	**1.03**
None + VIP	**0.94**	**0.85**	**0.97**	**0.79**	**1.29**	**0.88**
Mean center	0.98	0.70	0.95	0.52	1.83	1.11
Mean center + VIP	0.98	0.77	0.95	0.44	1.59	1.19
Auto scale	0.87	0.79	0.95	1.21	1.53	1.08
Auto scale + VIP	0.84	0.82	0.94	1.29	1.39	1.21

Table 7.7- Summary of u-PLS model quality parameters using a variety of different pre-processing and variable selection procedures. These models excluded the samples from solution 7 (mechanically stressed samples), Validation samples set selected using Kennard stone algorithm and comprised of six samples. The models shown in Figure 4.8 are highlighted here in yellow. These were the best results for quantification of aggregation based on RMSE values and R^2.

Pre-processing	R^2 Cal	R^2 CV	R^2 Pred	RMSE Cal	RMSE CV	RMSE Pred
None	**0.92**	**0.88**	**0.96**	**0.96**	**1.15**	**1.03**
None + VIP	**0.94**	**0.85**	**0.97**	**0.79**	**1.29**	**0.88**
Mean center	0.98	0.70	0.95	0.52	1.83	1.11
Mean center + VIP	0.98	0.77	0.95	0.44	1.59	1.19
Auto scale	0.87	0.79	0.95	1.21	1.53	1.08
Auto scale + VIP	0.84	0.82	0.94	1.29	1.39	1.21

[d] We did not have SEC data for all samples from Lot1 (only from each stock solution), thus, to match the SEC data, we built the models using the average EEM spectra of the 3 samples from each stock solution.

[e] Pre-processing here means the processing of the data for the modelling only (besides blank subtraction + missing data + normalization to the maximum peak + unfolding)

7.7.4. Qualitative multivariate data analysis of rIgG:

Table 7.8- ROBPCA results from the analysis of UV-Vis, EEM‖ and RS‖ datasets of the rIgG starting material samples. The results are divided into models of only unstressed samples, only Lo1 samples, only stressed samples and a model combining stressed and unstressed.

ROBPC	Unstressed samples (n=28)		Lot1 Samples (n=15)		All samples(n=34)			Stressed samples (n=8)
	UV-Vis	EEM‖	UV-Vis	EEM‖	UV-Vis	EEM‖	RS‖	EEM‖
1	79.90	81.21	87.48	77.68	74.60	57.69	94.25	75.35
2	13.09	14.40	9.39	8.85	18.46	35.75	5.23	23.38
3	2.89	1.41	0.86	1.63	-	2.84	0.25	0.18
4	-	0.55	-	-	-	0.84	-	-
Total variance	95.88	97.58	97.72	88.16	93.06	97.12	100.00	98.91

7.8. Additional information: Results of PEG-IgG study

7.8.1. Preliminary results of HIC of unpurified PEG-rIgG

We attempted the analysis of unpurified PEG-rIgG reaction mixtures using HIC, but this represented a big challenge because samples we observed a big increase in pressure after reaction mixture samples were injected. The results obtained from the first HIC runs are discussed below. (Figure 7.8)

Despite the low resolution, the chromatograms suggest the formation of two or three species with different hydrophobicity as seen by the new peaks at retention time between ~11-13min. If we consider these as the conjugates, we could use the %AUC to calculate the yield of the reactions: ~31% and 33% for reactions using 1.5 and 3.0TCEP, respectively. However, it is important to highlight that these results are not very reliable because of the problems with the analysis and the fact that we did not run samples in replicate. We tried to carry out a method optimization varying salt concentration, gradient times and % of organic modifier but unfortunately the pressure was drastically increasing with injections and thus we stopped the analysis to avoid damaging the column.

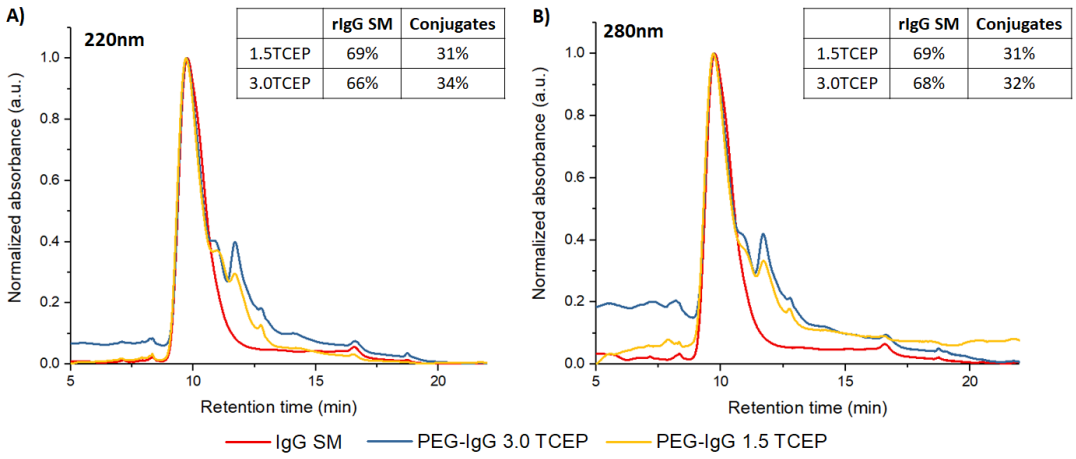


Figure 7.8- HIC chromatograms at (A) 220 nm and (B) 280 nm of rIgG starting material (red) and unpurified PEG-rIgG produced using 1.5TCEP (blue) and 3.0TCEP (yellow). The tables indicate the % AUC of SM and conjugates calculate for each wavelength of detection.

HIC of all samples was performed in a Protein-Pak Hi Res HIC column (4.6 x 100 mm, 2.5 µm) from Waters Technologies in an Agilent 1260 HPLC system with a DAD detector. 50µL of sample were injected onto the column at 30 ºC with a flow rate of 0.70mL/min using a gradient from 0 to 100% of mobile phase B as per table below.

The samples were diluted in mobile phase A (50:50) and kept at 25ºC in the sample cell until injection.

Mobile phase A: 2.5 ammonium sulphate + 125mM potassium phosphate pH 6.7

Mobile phase B: 125mM potassium phosphate pH 6.7 +25% isopropanol

Time (min)	%A	%B
0	100	0
15	0	100
17	100	0
30	100	0

7.8.2. Results of free sulfhydryl quantification: Polyclonal rIgG *vs* Monoclonal IgG1

Table 7.9- Results of free -SH quantification using the Ellman's method (details in Section 2.2.3) carried out in Polyclonal rIgG and monoclonal IgG1 at different temperatures indicating differences in reactivity towards reduction by TCEP. The values represent the number generated thiols per antibody. The two conditions discussed in Section 5.4 are highlighted here in yellow.

	20 °C	
Molar excess of TCEP	**Polyclonal (rIgG)**	**Monoclonal (IgG1)**
1.5	0.47	-
2.5	-	1.7
3	0.83	-
7.5	-	2.5
10	1.34	3.9
25	-	8.8
50	8.23	11.2
150	17.01	-
750	17.69	-

7.8.3. Intensity correction factor using *p-terphenyl* solid standard.

p-Terphenyl (Agilent) in PPMA, emitting in the same spectral region as the protein under study, was used as fluorescence solid standard to account for changes in the instrument sensitivity over the long-time gap between the R1-R14 (before) and R15-R19 (after) measurements. This was due to an instrument repair and replacement of the main board electronics. Figure 7.9 illustrates the change in the instrument sensitivity with time showing an overall decrease in intensity, which is clear when comparing the mean for R1-R14 (before) and R15-R19 (after) measurements. Maximum intensity decreased ~45%,49% and 47% for $EEM_{\parallel}$, $EEM_{\perp}$ and EEM_{T} data,

respectively. However, in terms of reproducibility the relative standard deviation indicates that both sets of samples are still comparable.

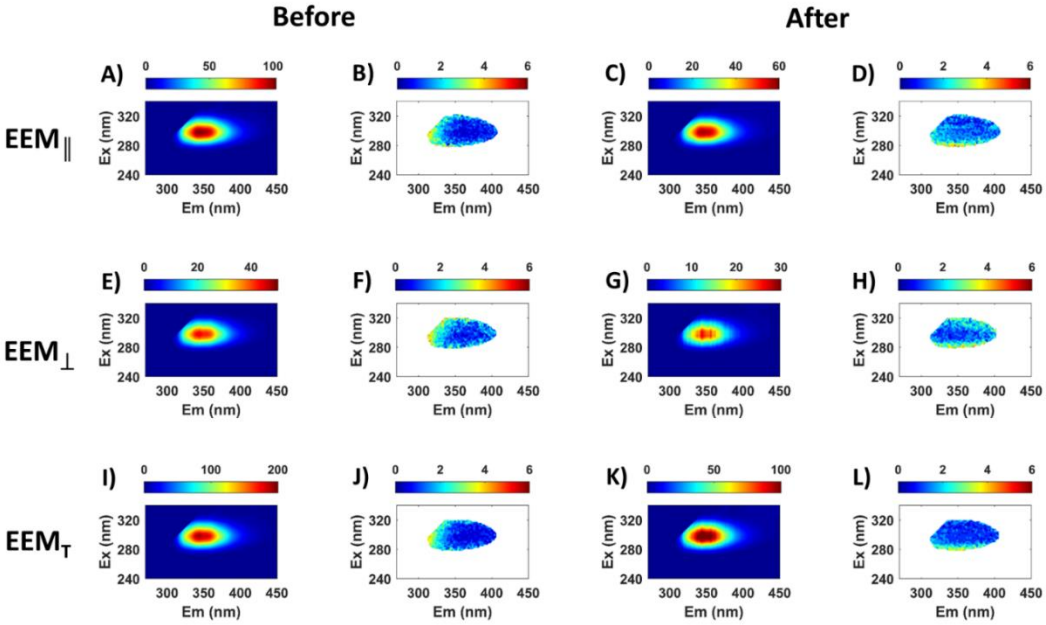


Figure 7.9- Mean (A,C,E,G,I,K) and relative standard deviation (B,D,F,H,J,L) calculated from raw p-terphenyl data collected at different polarization settings for R1-R14 (before) and R15-R19 (after) measurements. The four pEEM spectra (HH,HV,VH,VV) of p-terphenyl were collected over an excitation range of λ_{ex}=240–340 nm and an emission range of 270–450 nm ($\Delta\lambda$ = 2 nm in each case) with 10 nm excitation/emission slit widths, scan rate of 1200 nm min^{-1} and photomultiplier voltage was set at 600V.

The correction factor was obtained for each data set type ($EEM_{\parallel}$, $EEM_{\perp}$ and EEM_T) by calculating the ratio I_{after}/I_{before} (Table 7.10). The intensity value considered for the ratio was calculated as the average of the area around 98% of the maximum intensity peak of p-terphenyl at $\lambda_{ex/em}$=298/344. The data was collected over an excitation range of λ_{ex}=240–340 nm and an emission range of 260–450 nm ($\Delta\lambda$ = 2 nm in each case) with 10 nm excitation/emission slit widths, scan rate of 1200 nm min^{-1} and photomultiplier voltage was set at 600V.

Table 7.10: Correction factors calculated from replicate measurements of the p-terphenyl solid standard.

	R1[f]	R2	R3	R4	R5	Mean ± stdev (n=4)
$EEM_{\parallel}$	**2.00**	1.84	1.77	1.81	1.78	1.80 ± 0.03 (RSD 1.69%)
$EEM_{\perp}$	**2.30**	1.94	1.89	1.91	1.83	1.89 ± 0.04 (RSD 2.32%)
EEM_T	**2.14**	1.89	1.82	1.85	1.80	1.84 ± 0.04 (RSD 2.04%)

[f] R1 was not considered for the calculation because it was significantly different ($p<0.05$) from the other replicate measurements.

Table 7.11- Non-corrected and corrected maximum intensity values obtained from $EEM_{\parallel}$, $EEM_{\perp}$, and EEM_T of rIgG , Red-rIgG and PEG-rIgG used in the IgG quality screening study.

		$EEM_{\parallel}$			**$EEM_{\perp}$**			**EEM_T**		
		rIgG	**Red-rIgG**	**PEG-rIgG**	**rIgG**	**Red-rIgG**	**PEG-rIgG**	**rIgG**	**Red-rIgG**	**PEG-rIgG**
Non corrected	**Lot 2**	85.75	83.55	82.75	64.57	62.85	63.06	214.18	208.46	206.93
		83.58	84.13	82.56	62.12	63.56	62.32	207.64	210.57	204.92
		85.79	84.36	83.59	63.61	61.89	63.59	213.01	207.57	209.72
	Mean	**85.04**	**84.01**	**82.97**	**63.43**	**62.77**	**62.99**	**211.61**	**208.87**	**207.19**
	Stdev	**1.26**	**0.41**	**0.55**	**1.24**	**0.83**	**0.64**	**3.49**	**1.54**	**2.41**
	RSD	**1.49**	**0.49**	**0.66**	**1.95**	**1.33**	**1.01**	**1.65**	**0.74**	**1.16**
	Lot 3	85.76	79.92	81.53	67.40	60.49	62.87	217.66	199.46	205.65
		86.17	83.70	83.45	66.92	63.61	63.62	218.99	209.98	210.58
	Mean	**85.97**	**81.81**	**82.49**	**67.16**	**62.05**	**63.25**	**218.32**	**204.72**	**208.11**
	Stdev	**0.29**	**2.67**	**1.36**	**0.34**	**2.21**	**0.53**	**0.94**	**7.44**	**3.49**
	RSD	**0.34**	**3.27**	**1.65**	**0.50**	**3.56**	**0.84**	**0.43**	**3.63**	**1.68**
Corrected	**Lot 2**	154.35	150.39	148.96	122.03	118.78	119.18	394.09	383.57	380.75
		150.44	151.44	148.62	117.40	120.12	117.78	382.06	387.45	377.05
		154.42	151.84	150.47	120.23	116.98	120.18	391.94	381.93	385.88
	Mean	**153.07**	**151.22**	**149.35**	**119.89**	**118.63**	**119.05**	**389.36**	**384.32**	**381.23**
	Stdev	**2.27**	**0.75**	**0.99**	**2.33**	**1.58**	**1.20**	**6.41**	**2.83**	**4.43**
	RSD	**1.49**	**0.49**	**0.66**	**1.95**	**1.33**	**1.01**	**1.65**	**0.74**	**1.16**
	Lot 3	154.37	143.85	146.76	124.01	111.30	115.69	400.50	367.01	378.39
		155.11	150.66	150.21	123.13	117.05	117.07	402.93	386.37	387.47
	Mean	**154.74**	**147.26**	**148.48**	**123.57**	**114.17**	**116.38**	**401.72**	**376.69**	**382.93**
	Stdev	**0.53**	**4.81**	**2.44**	**0.62**	**4.07**	**0.98**	**1.72**	**13.69**	**6.42**
	RSD	**0.34**	**3.27**	**1.65**	**0.50**	**3.56**	**0.84**	**0.43**	**3.63**	**1.68**

7.8.4. SimI values using different measurements and reference spectrum

Comparison of *SimI* values when considering one selected spectrum (here randomly set as rIgG from R01) as the reference and when setting the respective starting material of each reaction as the reference. When taking a specific sample as the reference, it is possible to analyse how the reactions are varying given a golden batch or reference sample. Taking the respective rIgG as the reference offers a type of "normalization" for each reaction, allowing the study of each reaction individually.

Table 7.12- *SimI* values (mean ± stdev) calculated for reactions using different lots of rIgG. The results highlight the values obtained when different spectra are taken as reference for the calculation.

	Reference:	Lot	rIgG	Red-rIgG	PEG-rIgG
UV-Vis	IgG R01	1	0.97 ± 0.02	0.75 ± 0.05	0.73 ± 0.05
		2	0.98 ± 0.01	0.89 ± 0.01	0.88 ± 0.02
		3	0.94 ± 0.00	0.67 ± 0.01	0.64 ± 0.00
	Respective rIgG starting mat.	1	1.0	0.77± 0.04	0.75±0.05
		2	1.0	0.90 ± 0.02	0.90 ± 0.03
		3	1.0	0.72 ± 0.01	0.69 ± 0.00
EEM∥	IgG R01	1	0.95±0.02	0.92 ± 0.02	0.90±0.02
		2	0.90 ± 0.01	0.90 ± 0.01	0.88 ± 0.00
		3	0.83 ± 0.00	0.79 ± 0.00	0.76 ± 0.00
	Respective rIgG starting mat.	1	1.0	0.95±0.01	0.93±0.01
		2	1.0	0.95 ± 0.00	0.93 ± 0.01
		3	1.0	0.91 ± 0.00	0.89 ± 0.01
EEM⊥	IgG R01	1	0.96 ± 0.02	0.94 ± 0.02	0.92 ± 0.02
		2	0.90 ± 0.01	0.90 ± 0.00	0.88 ± 0.00
		3	0.83 ± 0.00	0.81 ± 0.00	0.78 ± 0.01
	Respective rIgG starting mat.	1	1.0	0.96±0.01	0.94±0.01
		2	1.0	0.96 ± 0.01	0.94 ± 0.01
		3	1.0	0.93 ± 0.00	0.91 ± 0.01
Aniso	IgG R01	1	0.64 ± 0.13	0.53 ± 0.08	0.46 ± 0.11
		2	0.26 ± 0.02	0.25 ± 0.02	0.25 ± 0.01
		3	0.23 ± 0.01	0.26 ± 0.01	0.23 ± 0.01
	Respective rIgG starting mat.	1	1.0	0.64±0.04	0.52±0.08
		2	1.0	0.95 ± 0.01	0.94 ± 0.01
		3	1.0	0.94 ± 0.00	0.91 ± 0.02

7.8.5. **Multivariate data analysis results**

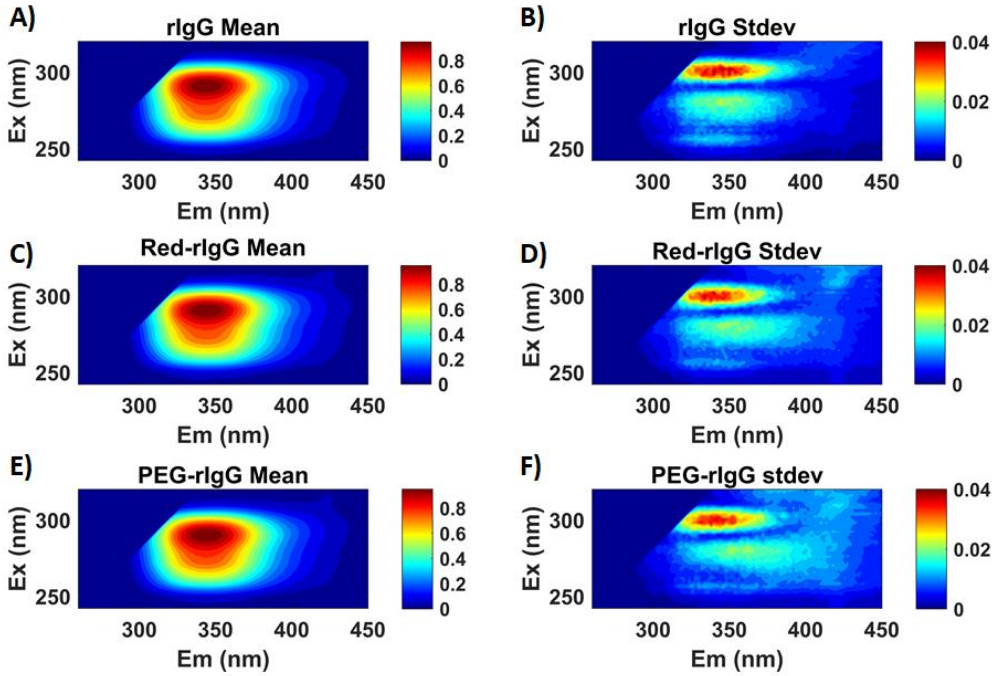


Figure 7.10- Mean and standard deviation plots calculated for all samples (Lot1, 2 and 3) used in the PEG-IgG study. A/B, C/D and E/F are the mean/stdev plots for rIgG, Red-rIgG and PEG-rIgG sample respectively. This figure should be compared with the results for the control experiment using Trp solutions (Figure 7.6) which indicates the intrinsic variability of the experiment.

7.8.6. ROBPCA analysis for source lot screening (Lot1,2,3)

Here only samples from Lot 1 were used to highlight differences between samples according to reaction stage and amount of TCEP used. Figures 7.11 and 7.12 show the scores plotted separately (by samples) for an easier visualization of the contribution of each PC.

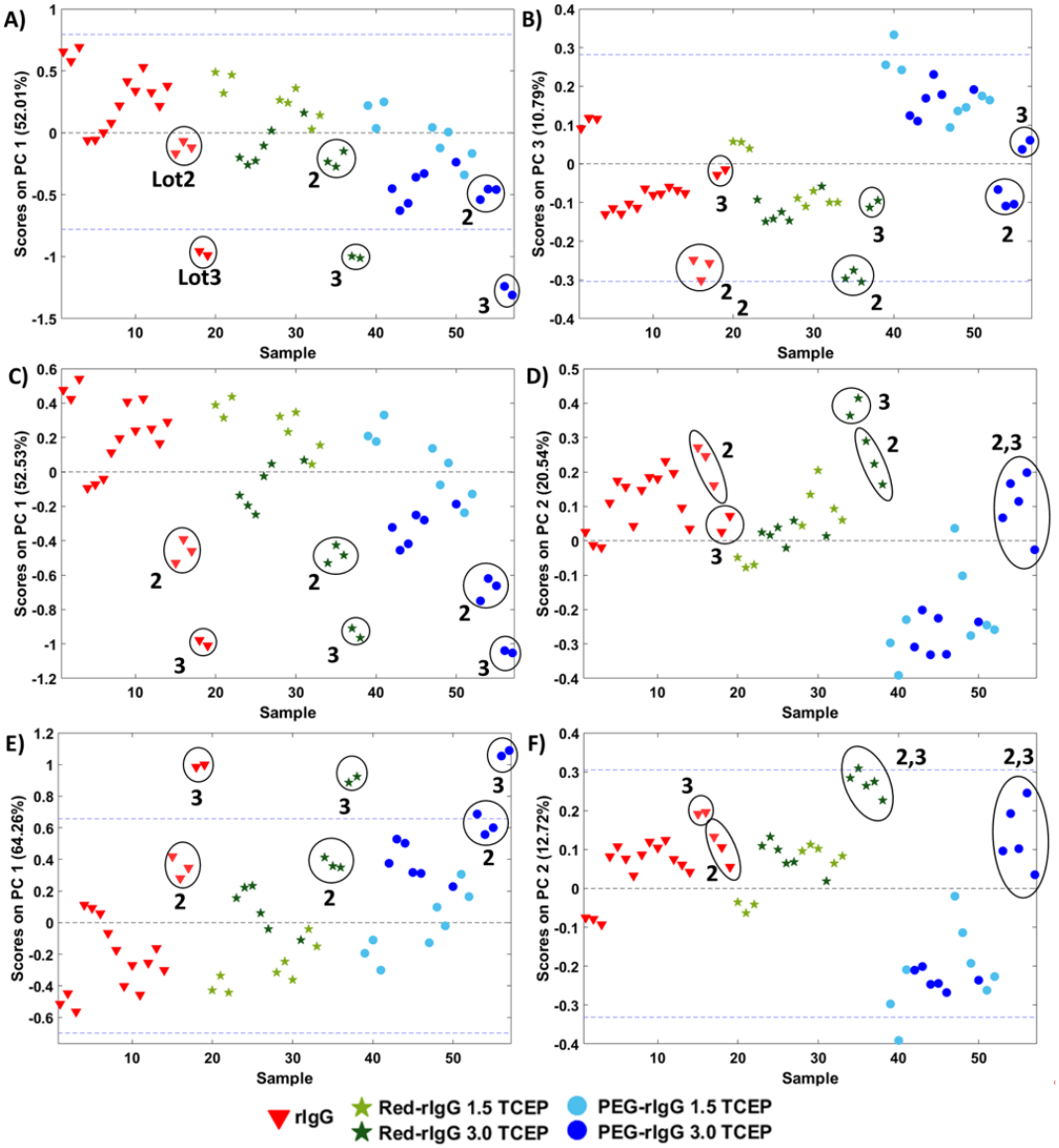


Figure 7.11-ROBPC 1, 3 scores from $EEM_{\parallel}$ (A,B respectively) and 1,2 from $EEM_{\perp}$ (C,D) and EEM_{T} (E,F) shows trends that discriminate samples according to reaction step, amount of TCEP used and rIgG starting material lot. The ellipses were manually added to highlight samples from Lot2 and 3.

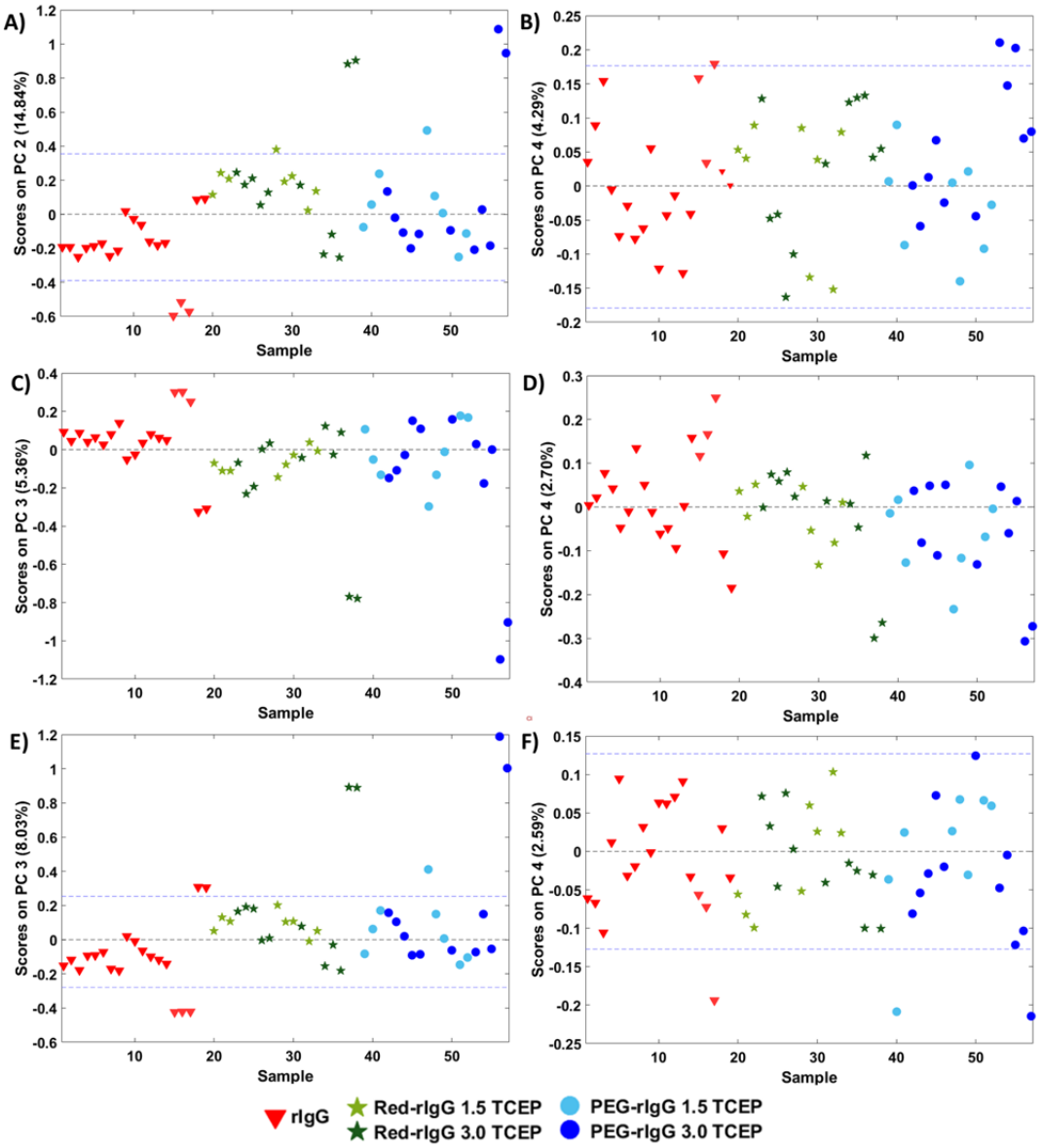


Figure 7.12- ROBPC 2 and 4 scores from EEM$_{\parallel}$ (A,B respectively) and 3 and 4 from EEM$_{\perp}$ and EEM$_{T}$ (C,D,E,F).

.

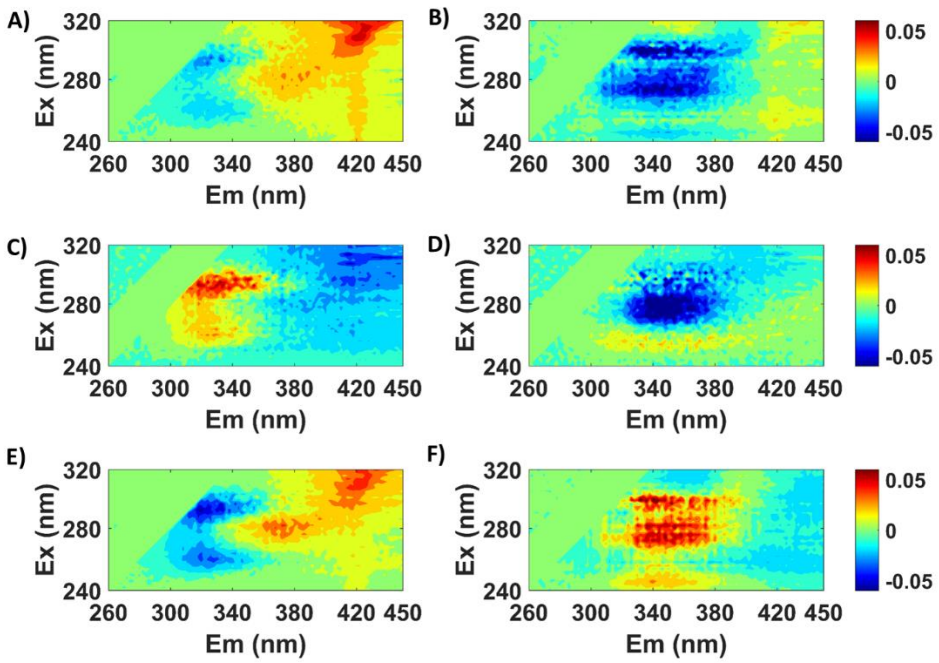


Figure 7.13- ROBPC2/4 loadings of EEM‖ (A,B) data and ROBPC3/4 loadings of EEM⊥ (C,D) and EEM$_T$(D,E). Here, samples from all lots were used.

ROBPCA analysis for reaction stage discrimination (Lot 1 samples only)

Here only samples from Lot 1 were used to highlight differences between samples according to reaction stage and amount of TCEP used.

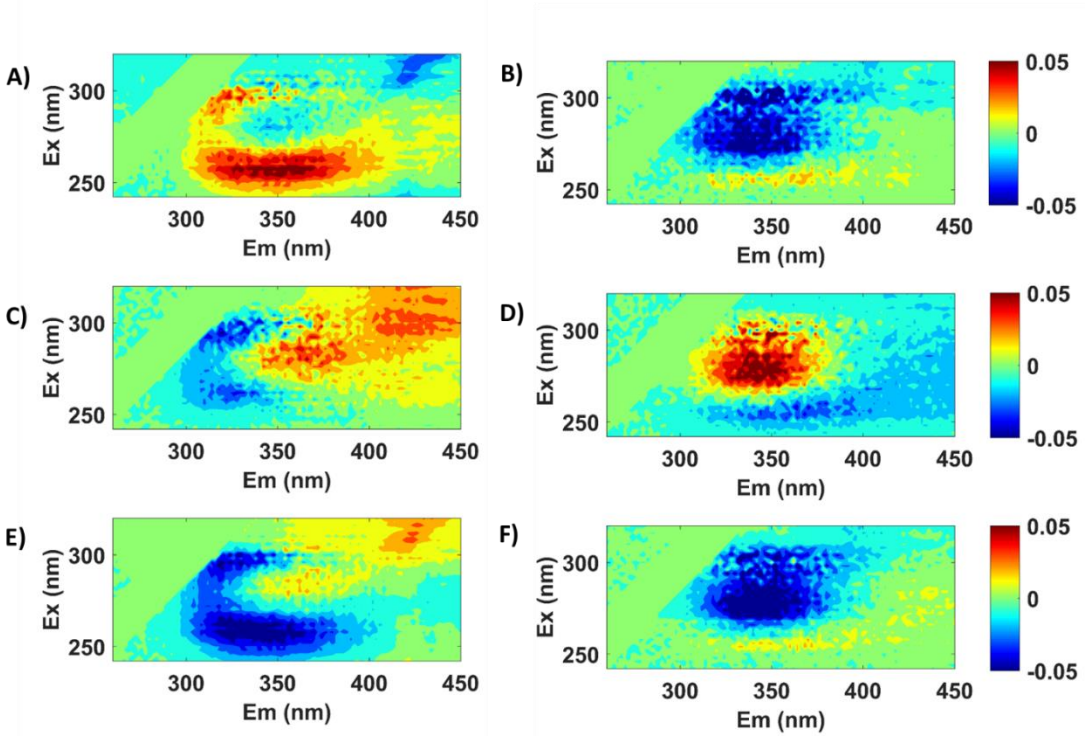


Figure 7.14- ROBPC3/4 loadings plots of EEM‖ (A,B); EEM⊥ (C,D); and EEM$_T$ data(E,F), respectively showing the areas of largest spectral change during PEG to rIgG conjugation.

7.8.7. Samples used in calibration/validation for the classification of reaction stages:

Table 7.13- Summary of number of samples used for calibration and validation sets when Lots 2 and 3 were included in the classification models. In situation 1, samples from Lot2 and 3 were included in the validation set with Lot1, while in situation 2, Lot2 and 3 samples were included in both calibration and validation sets.

Situation 1	Calibration Set (Lot1): rIgG (I): n =10 Red-rIgG (R): n=10 PEG-rIgG (P): n=10	Validation Set (Lot1,2,3): rIgG (I): n =9 Red-rIgG (R): n=9 PEG-rIgG (P): n=9
Situation 2	Calibration Set (Lot1,2,3): rIgG (I): n =13 Red-rIgG (R): n=13 PEG-rIgG (P): n=13	Validation Set (Lot1,2,3): rIgG (I): n =6 Red-rIgG (R): n=6 PEG-rIgG (P): n=6

7.8.8. Results of classification of both reaction stages and conditions using Lot1 samples:

Table 7.14- Confusion matrix summarizing the results of classification of samples according to reaction stage and conditions using Lot1 samples. Modelling was carried out using EEM‖ measurements and SVM. I = IgG starting material; R1.5 and 3.0= reduced product sample using 1.5 and 3.0TCEP; P1.5 and 3.0= pegylated product sample using 1.5 and 3.0TCEP;

Calibration Results											
Actual	**I**	**R1.5**	**R3.0**	**P1.5**	**P3.0**		**I**	**R1.5**	**R3.0**	**P1.5**	**P3.0**
I	**10**	**0**	**0**	**0**	**0**	**Sen.**	100	83	75	83	75
R1.5	**0**	**5**	**0**	**0**	**0**						
R3.0	**0**	**1**	**4**	**0**	**0**	**Spec.**	95	96	96	96	100
P 1.5	**0**	**0**	**0**	**6**	**1**						
P 3.0	**0**	**0**	**0**	**0**	**3**	**Err.**	3	7	7	7	3
Validation Results											
Actual	**I**	**R1.5**	**R3.0**	**P1.5**	**P3.0**		**I**	**R1.5**	**R3.0**	**P1.5**	**P3.0**
I	**4**	**0**	**0**	**0**	**0**	**Sen.**	100	100	100	100	100
R1.5	**0**	**2**	**0**	**0**	**0**						
R3.0	**0**	**0**	**2**	**0**	**0**	**Spec.**	100	100	100	100	100
P 1.5	**0**	**0**	**0**	**2**	**0**						
P 3.0	**0**	**0**	**0**	**0**	**2**	**Err.**	0	0	0	0	0

Chapter 8. Bibliography

1. Leader, B.; Baca, Q. J.; Golan, D. E., Protein therapeutics: A summary and pharmacological classification. *Nat Rev Drug Discov* **2008,** *7* (1), 21-39.
2. Dozier, J. K.; Distefano, M. D., Site-Specific PEGylation of Therapeutic Proteins. *Int J Mol Sci* **2015,** *16* (10), 25831-25864.
3. Cunningham, F. E.; Proctor, V. A.; Goetsch, S. J., Egg-white lysozyme as a food preservative: an overview. *World's Poultry Science Journal* **1991,** *47* (2), 141-163.
4. Proctor, V. A.; Cunningham, F. E., The chemistry of lysozyme and its use as a food preservative and a pharmaceutical. *Crit Rev Food Sci Nutr* **1988,** *26* (4), 359-95.
5. Iacono, V. J.; MacKay, B. J.; DiRienzo, S.; Pollock, J. J., Selective antibacterial properties of lysozyme for oral microorganisms. *Infect Immun* **1980,** *29* (2), 623-632.
6. Lesnierowski, G.; Kijowski, J., Lysozyme. In *Bioactive Egg Compounds*, Huopalahti, R.; López-Fandiño, R.; Anton, M.; Schade, R., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2007; pp 33-42.
7. Canfield, R. E.; Liu, A. K., The Disulfide bonds of egg white lysozyme (Murimidase) *J Biol Chem* **1965,** *240*, 1997-2002.
8. Imoto, T.; Forster, L. S.; Rupley, J. A.; Tanaka, F., Fluorescence of lysozyme: emissions from tryptophan residues 62 and 108 and energy migration. *Proc Natl Acad Sci U S A* **1972,** *69* (5), 1151-5.
9. Liu, J. K. H., The history of monoclonal antibody development - Progress, remaining challenges and future innovations. *Ann Med Surg (Lond)* **2014,** *3* (4), 113-116.
10. Reichert, J. M., Marketed therapeutic antibodies compendium. *Mabs-Austin* **2012,** *4* (3), 413-415.
11. Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S., Antibody Structure, Instability, and Formulation. *J Pharm Sci-Us* **2007,** *96* (1), 1-26.
12. Elgert, K. D., *Immunology : understanding the immune system*. Wiley-Liss: New York, 1996; p x, 468 p.
13. Male, D. K., *Immunology*. 7th ed.; Mosby Elsevier: Philadelphia, 2006; p x, 552.
14. Ponomarenko, N.; Chatziefthimiou, S. D.; Kurkova, I.; Mokrushina, Y.; Stepanova, A.; Smirnov, I.; Avakyan, M.; Bobik, T.; Mamedov, A.; Mitkevich, V.; Belogurov, A.; Fedorova, O. S.; Dubina, M.; Golovin, A.; Lamzin, V.; Friboulet, A.; Makarov, A. A.; Wilmanns, M.; Gabibov, A., Role of kappa ->lambda light-chain constant-domain switch in the structure and functionality of A17 reactibody. *Acta Crystallogr D* **2014,** *70*, 708-719.
15. McAuley, A.; Jacob, J.; Kolvenbach, C. G.; Westland, K.; Lee, H. J.; Brych, S. R.; Rehder, D.; Kleemann, G. R.; Brems, D. N.; Matsumura, M., Contributions of a disulfide bond to the structure, stability, and dimerization of human IgG1 antibody CH3 domain. *Protein Sci* **2008,** *17* (1), 95-106.
16. Liu, H. C.; May, K., Disulfide bond structures of IgG molecules Structural variations, chemical modifications and possible impacts to stability and biological function. *Mabs-Austin* **2012,** *4* (1), 17-23.

17. Lamond, A. I., Molecular biology of the cell, 4th edition. *Nature* **2002,** *417* (6887), 383-383.
18. Lehninger, A. L.; Nelson, D. L.; Cox, M. M., *Lehninger principles of biochemistry*. Macmillan: 2005.
19. Nick Pace, C.; Scholtz, J. M.; Grimsley, G. R., Forces stabilizing proteins. *FEBS Lett* **2014,** *588* (14), 2177-2184.
20. Correia, I. R., Stability of IgG isotypes in serum. *Mabs-Austin* **2010,** *2* (3), 221-232.
21. Rayner, L. E.; Kadkhodayi-Kholghi, N.; Heenan, R. K.; Gor, J.; Dalby, P. A.; Perkins, S. J., The Solution Structure of Rabbit IgG Accounts for Its Interactions with the Fc Receptor and Complement C1q and Its Conformational Stability. *J Mol Biol* **2013,** *425* (3), 506-523.
22. McNaught, A. D.; Wilkinson, A., *IUPAC compendium of chemical terminology : the gold book*. 2nd ed.; Blackwell Scientific Publications, Oxford. Online version (2019-) created by S. J. Chalk., 1997.
23. Branden, C. I.; Tooze, J., *Introduction to Protein Structure*. CRC Press: 2012.
24. Mader, S. S., *Biology*. McGraw-Hill Higher Education: New York, 2010.
25. Gromiha, M. M., Chapter 6 - Protein Stability. In *Protein Bioinformatics*, Gromiha, M. M., Ed. Academic Press: Singapore, 2010; pp 209-245.
26. Doig, A. J.; Sternberg, M. J., Side-chain conformational entropy in protein folding. *Protein Sci* **1995,** *4* (11), 2247-51.
27. Vermeer, A. W.; Norde, W., The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein. *Biophys J* **2000,** *78* (1), 394-404.
28. Kameoka, D.; Masuzaki, E.; Ueda, T.; Imoto, T., Effect of buffer species on the unfolding and the aggregation of humanized IgG. *J Biochem* **2007,** *142* (3), 383-91.
29. Wang, T.; Kumru, O. S.; Yi, L.; Wang, Y. J.; Zhang, J.; Kim, J. H.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B., Effect of ionic strength and pH on the physical and chemical stability of a monoclonal antibody antigen-binding fragment. *J Pharm Sci* **2013,** *102* (8), 2520-37.
30. Dominy, B. N.; Perl, D.; Schmid, F. X.; Brooks, C. L., 3rd, The effects of ionic strength on protein stability: the cold shock protein family. *J Mol Biol* **2002,** *319* (2), 541-54.
31. Vazquez-Rey, M.; Lang, D. A., Aggregates in Monoclonal Antibody Manufacturing Processes. *Biotechnol Bioeng* **2011,** *108* (7), 1494-1508.
32. Wang, W.; Nema, S.; Teagarden, D., Protein aggregation—Pathways and influencing factors. *Int J Pharmaceut* **2010,** *390* (2), 89-99.
33. Roberts, C. J., Therapeutic protein aggregation: mechanisms, design, and control. *Trends Biotechnol* **2014,** *32* (7), 372-380.
34. Wang, W.; Roberts, C. J., Protein aggregation - Mechanisms, detection, and control. *Int J Pharm* **2018,** *550* (1-2), 251-268.
35. Luo, Q. Z.; Joubert, M. K.; Stevenson, R.; Ketchem, R. R.; Narhi, L. O.; Wypych, J., Chemical Modifications in Therapeutic Protein Aggregates Generated under Different Stress Conditions. *J Biol Chem* **2011,** *286* (28), 25134-25144.
36. Vlasak, J.; Ionescu, R., Fragmentation of monoclonal antibodies. *Mabs-Austin* **2011,** *3* (3), 253-263.
37. Ke, Z.; Huang, Q., Effect of protein structure and/or conformation on the dityrosine cross-linking induced by haem-hydrogen peroxide. *Biochim Biophys Acta* **2016,** *1860* (10), 2232-8.

38. Das, T. K.; Narhi, L. O.; Sreedhara, A.; Menzen, T.; Grapentin, C.; Chou, D. K.; Antochshuk, V.; Filipe, V., Stress Factors in mAb Drug Substance Production Processes: Critical Assessment of Impact on Product Quality and Control Strategy. *J Pharm Sci* **2020,** *109* (1), 116-133.
39. Cockrell, G. M.; Wolfe, M. S.; Wolfe, J. L.; Schoneich, C., Photoinduced Aggregation of a Model Antibody-Drug Conjugate. *Mol Pharmaceut* **2015,** *12* (6), 1784-1797.
40. Ratanji, K. D.; Derrick, J. P.; Dearman, R. J.; Kimber, I., Immunogenicity of therapeutic proteins: influence of aggregation. *J Immunotoxicol* **2014,** *11* (2), 99-109.
41. Sharma, B., Immunogenicity of therapeutic proteins. Part 1: impact of product handling. *Biotechnol Adv* **2007,** *25* (3), 310-7.
42. Ross, P. L.; Wolfe, J. L., Physical and Chemical Stability of Antibody Drug Conjugates: Current Status. *J Pharm Sci-Us* **2016,** *105* (2), 391-397.
43. Turecek, P. L.; Bossard, M. J.; Schoetens, F.; Ivens, I. A., PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. *Journal of Pharmaceutical Sciences* **2016,** *105* (2), 460-475.
44. Beck, A.; Goetsch, L.; Dumontet, C.; Corvaia, N., Strategies and challenges for the next generation of antibody drug conjugates. *Nature Reviews Drug Discovery* **2017,** *16* (5), 315-337.
45. Alley, S. C.; Okeley, N. M.; Senter, P. D., Antibody–drug conjugates: targeted drug delivery for cancer. *Curr Opin Chem Biol* **2010,** *14* (4), 529-537.
46. Kraeber-Bodere, F.; Bodet-Milin, C.; Rousseau, C.; Eugene, T.; Pallardy, A.; Frampas, E.; Carlier, T.; Ferrer, L.; Gaschet, J.; Davodeau, F.; Gestin, J. F.; Faivre-Chauvet, A.; Barbet, J.; Cherel, M., Radioimmunoconjugates for the Treatment of Cancer. *Semin Oncol* **2014,** *41* (5), 613-622.
47. Tsuchikama, K.; An, Z., Antibody-drug conjugates: recent advances in conjugation and linker chemistries. *Protein Cell* **2018,** *9* (1), 33-46.
48. Kang, J. S.; Deluca, P. P.; Lee, K. C., Emerging PEGylated drugs. *Expert Opin Emerg Drugs* **2009,** *14* (2), 363-80.
49. LoRusso, P. M.; Weiss, D.; Guardino, E.; Girish, S.; Sliwkowski, M. X., Trastuzumab emtansine: a unique antibody-drug conjugate in development for human epidermal growth factor receptor 2-positive cancer. *Clin Cancer Res* **2011,** *17* (20), 6437-47.
50. Younes, A.; Bartlett, N. L.; Leonard, J. P.; Kennedy, D. A.; Lynch, C. M.; Sievers, E. L.; Forero-Torres, A., Brentuximab vedotin (SGN-35) for relapsed CD30-positive lymphomas. *N Engl J Med* **2010,** *363* (19), 1812-21.
51. Pradhananga, S.; Wilkinson, I.; Ross, R. J., Pegvisomant: structure and function. *J Mol Endocrinol* **2002,** *29* (1), 11-4.
52. Grillo-López, A. J., Zevalin: the first radioimmunotherapy approved for the treatment of lymphoma. *Expert Review of Anticancer Therapy* **2002,** *2* (5), 485-493.
53. Goel, N.; Stephens, S., Certolizumab pegol. *Mabs-Austin* **2010,** *2* (2), 137-47.
54. Mueller, B. M.; Wrasidlo, W. A.; Reisfeld, R. A., Determination of the number of e-amino groups available for conjugation of effector molecules to monoclonal antibodies. *Hybridoma* **1988,** *7* (5), 453-6.
55. Chari, R. V. J., Targeted cancer therapy: Conferring specificity to cytotoxic drugs. *Accounts Chem Res* **2008,** *41* (1), 98-107.
56. Jain, N.; Smith, S. W.; Ghone, S.; Tomczuk, B., Current ADC Linker Chemistry. *Pharm Res* **2015,** *32* (11), 3526-40.
57. Ducry, L. e., *Antibody-drug conjugates*. Methods in Molecular Biology, vol. 1045, Springer Science+Business Media: 2013.

58. Spicer, C. D.; Davis, B. G., Selective chemical protein modification. *Nature Communications* **2014,** *5* (1), 4740.
59. Boutureira, O.; Bernardes, G. J. L., Advances in Chemical Protein Modification. *Chemical Reviews* **2015,** *115* (5), 2174-2195.
60. Wakankar, A.; Chen, Y.; Gokarn, Y.; Jacobson, F. S., Analytical methods for physicochemical characterization of antibody drug conjugates. *Mabs-Austin* **2011,** *3* (2), 161-172.
61. Wagh, A.; Song, H.; Zeng, M.; Tao, L.; Das, T. K., Challenges and new frontiers in analytical characterization of antibody-drug conjugates. *Mabs-Austin* **2018,** *10* (2), 222-243.
62. Wakankar, A. A.; Feeney, M. B.; Rivera, J.; Chen, Y.; Kim, M.; Sharma, V. K.; Wang, Y. J., Physicochemical Stability of the Antibody−Drug Conjugate Trastuzumab-DM1: Changes due to Modification and Conjugation Processes. *Bioconjugate Chem* **2010,** *21* (9), 1588-1595.
63. Cheng, T. L.; Chuang, K. H.; Chen, B. M.; Roffler, S. R., Analytical Measurement of PEGylated Molecules. *Bioconjugate Chem* **2012,** *23* (5), 881-899.
64. Guidance for industry: Q8 (R2) pharmaceutical development, Guideline ICH Harmonised Tripartite. *Current step* **2009,** *4*.
65. In *Guidance for Industry PAT@ A Framework for Innovative Pharmaceutical Development , Manufacturing , and Quality Assurance*, 2004.
66. Alt, N.; Zhang, T. Y.; Motchnik, P.; Taticek, R.; Quarmby, V.; Schlothauer, T.; Beck, H.; Emrich, T.; Harris, R. J., Determination of critical quality attributes for monoclonal antibodies using quality by design principles. *Biologicals* **2016,** *44* (5), 291-305.
67. Raynal, B.; Lenormand, P.; Baron, B.; Hoos, S.; England, P., Quality assessment and optimization of purified protein samples: why and how? *Microb Cell Fact* **2014,** *13*.
68. He, F.; Phan, D. H.; Hogan, S.; Bailey, R.; Becker, G. W.; Narhi, L. O.; Razinkov, V. I., Detection of IgG Aggregation by a High Throughput Method Based on Extrinsic Fluorescence. *J Pharm Sci-Us* **2010,** *99* (6), 2598-2608.
69. Arosio, P.; Rima, S.; Morbidelli, M., Aggregation Mechanism of an IgG2 and two IgG1 Monoclonal Antibodies at low pH: From Oligomers to Larger Aggregates. *Pharm Res-Dordr* **2013,** *30* (3), 641-654.
70. Pai, S. S.; Hammouda, B.; Hong, K. L.; Pozzo, D. C.; Przybycien, T. M.; Tilton, R. D., The Conformation of the Poly(ethylene glycol) Chain in Mono-PEGylated Lysozyme and Mono-PEGylated Human Growth Hormone. *Bioconjugate Chem* **2011,** *22* (11), 2317-2323.
71. Mori, S.; Barth, H. G., *Size exclusion chromatography*. Springer: Berlin ; New York, 1999; p xiv, 234 p.
72. Hughes, H.; Morgan, C.; Brunyak, E.; Barranco, K.; Cohen, E.; Edmunds, T.; Lee, K., A Multi-Tiered Analytical Approach For the Analysis and Quantitation of High-Molecular-Weight Aggregates in a Recombinant Therapeutic Glycoprotein. *Aaps J* **2009,** *11* (2), 335-341.
73. Liu, J.; Andya, J. D.; Shire, S. J., A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. *Aaps J* **2006,** *8* (3), E580-E589.
74. Berkowitz, S. A., Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. *Aaps J* **2006,** *8* (3), E590-E605.

75. Thiagarajan, G.; Semple, A.; James, J. K.; Cheung, J. K.; Shameem, M., A comparison of biophysical characterization techniques in predicting monoclonal antibody stability. *Mabs-Austin* **2016,** *8* (6), 1088-1097.
76. Nobbmann, U.; Connah, M.; Fish, B.; Varley, P.; Gee, C.; Mulot, S.; Chen, J.; Zhou, L.; Lu, Y.; Sheng, F.; Yi, J.; Harding, S. E., Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. *Biotechnology and Genetic Engineering Reviews* **2007,** *24* (1), 117-128.
77. Gokarn, Y. R.; McLean, M.; Laue, T. M., Effect of PEGylation on Protein Hydrodynamics. *Mol Pharmaceut* **2012,** *9* (4), 762-773.
78. Bhattacharjee, S., DLS and zeta potential – What they are and what they are not? *Journal of Controlled Release* **2016,** *235*, 337-351.
79. Zhu, Z. Q. C.; Chen, Y. C.; Ackerman, M. S.; Wang, B.; Wu, W.; Li, B.; Obenauer-Kutner, L.; Zhao, R. L.; Tao, L.; Ihnat, P. M.; Liu, J. P.; Gandhi, R. B.; Qiu, B., Investigation of monoclonal antibody fragmentation artifacts in non-reducing SDS-PAGE. *J Pharmaceut Biomed* **2013,** *83*, 89-95.
80. Lu, C. N.; Liu, D. D.; Liu, H. B.; Motchnik, P., Characterization of monoclonal antibody size variants containing extra light chains. *Mabs-Austin* **2013,** *5* (1), 102-113.
81. Sklepari, M.; Rodger, A.; Reason, A.; Jamshidi, S.; Prokes, I.; Blindauer, C. A., Biophysical characterization of a protein for structure comparison: methods for identifying insulin structural changes. *Analytical Methods* **2016,** *8* (41), 7460-7471.
82. Gokarn, Y.; Agarwal, S.; Arthur, K.; Bepperling, A.; Day, E. S.; Filoti, D.; Greene, D. G.; Hayes, D.; Kroe-Barrett, R.; Laue, T.; Lin, J.; McGarry, B.; Razinkov, V.; Singh, S.; Taing, R.; Venkataramani, S.; Weiss, W.; Yang, D. L.; Zarraga, I. E., Biophysical Techniques for Characterizing the Higher Order Structure and Interactions of Monoclonal Antibodies. *Acs Sym Ser* **2015,** *1201*, 285-327.
83. Noble, J. E., Quantification of Protein Concentration Using UV Absorbance and Coomassie Dyes. In *Methods in Enzymology*, Lorsch, J., Ed. Academic Press: 2014; Vol. 536, pp 17-26.
84. Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T., How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **1995,** *4* (11), 2411-23.
85. Andris, S.; Rudt, M.; Rogalla, J.; Wendeler, M.; Hubbuch, J., Monitoring of antibody-drug conjugation reactions with UV/Vis spectroscopy. *J Biotechnol* **2018,** *288*, 15-22.
86. Hamblett, K. J.; Senter, P. D.; Chace, D. F.; Sun, M. M. C.; Lenox, J.; Cerveny, C. G.; Kissler, K. M.; Bernhardt, S. X.; Kopcha, A. K.; Zabinski, R. F.; Meyer, D. L.; Francisco, J. A., Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clin Cancer Res* **2004,** *10* (20), 7063-7070.
87. D'Atri, V.; Pell, R.; Clarke, A.; Guillarme, D.; Fekete, S., Is hydrophobic interaction chromatography the most suitable technique to characterize site-specific antibody-drug conjugates? *Journal of Chromatography A* **2019,** *1586*, 149-153.
88. Valliere-Douglass, J.; Wallace, A.; Balland, A., Separation of populations of antibody variants by fine tuning of hydrophobic-interaction chromatography operating conditions. *J Chromatogr A* **2008,** *1214* (1-2), 81-9.
89. Queiroz, J. A.; Tomaz, C. T.; Cabral, J. M. S., Hydrophobic interaction chromatography of proteins. *J Biotechnol* **2001,** *87* (2), 143-159.
90. Khawli, L. A.; Goswami, S.; Hutchinson, R.; Kwong, Z. W.; Yang, J.; Wang, X.; Yao, Z.; Sreedhara, A.; Cano, T.; Tesar, D.; Nijem, I.; Allison, D. E.; Wong, P. Y.; Kao, Y.-H.; Quan, C.; Joshi, A.; Harris, R. J.; Motchnik, P., Charge variants

in IgG1: Isolation, characterization, in vitro binding properties and pharmacokinetics in rats. *Mabs-Austin* **2010,** *2* (6), 613-624.
91. Yamamoto, S.; Nakanishi, K.; Matsuno, R., *Ion-Exchange Chromatography of Proteins*. Taylor & Francis: 1988.
92. Cindric, M.; Cepo, T.; Galic, N.; Bukvic-Krajacic, M.; Tomczyk, N.; Vissers, J. P. C.; Bindila, L.; Peter-Katalinic, J., Structural characterization of PEGylated rHuG-CSF and location of PEG attachment sites. *J Pharmaceut Biomed* **2007,** *44* (2), 388-395.
93. Ladwig, P. M.; Barnidge, D. R.; Willrich, M. A. V., Mass Spectrometry Approaches for Identification and Quantitation of Therapeutic Monoclonal Antibodies in the Clinical Laboratory. *Clin Vaccine Immunol* **2017,** *24* (5).
94. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.
95. Feist, P.; Hummon, A. B., Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015,** *16* (2), 3537-63.
96. Chiu, M. H.; Prenner, E. J., Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. *J Pharm Bioallied Sci* **2011,** *3* (1), 39-59.
97. Palm, T.; Esfandiary, R.; Gandhi, R., The effect of PEGylation on the stability of small therapeutic proteins. *Pharm Dev Technol* **2011,** *16* (5), 441-448.
98. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **2006,** *1* (6), 2876-2890.
99. Freitas, D. D.; Abrahao-Neto, J., Biochemical and biophysical characterization of lysozyme modified by PEGylation. *Int J Pharmaceut* **2010,** *392* (1-2), 111-117.
100. Kong, J.; Yu, S., Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochimica et Biophysica Sinica* **2007,** *39* (8), 549-559.
101. Deygen, I. M.; Kudryashova, E. V., New versatile approach for analysis of PEG content in conjugates and complexes with biomacromolecules based on FTIR spectroscopy. *Colloid Surface B* **2016,** *141*, 36-43.
102. Grosshans, S.; Rudt, M.; Sanden, A.; Brestrich, N.; Morgenstern, J.; Heissler, S.; Hubbuch, J., In-line Fourier-transform infrared spectroscopy as a versatile process analytical technology for preparative protein chromatography. *Journal of Chromatography A* **2018,** *1547*, 37-44.
103. den Engelsman, J.; Garidel, P.; Smulders, R.; Koll, H.; Smith, B.; Bassarab, S.; Seidl, A.; Hainzl, O.; Jiskoot, W., Strategies for the Assessment of Protein Aggregates in Pharmaceutical Biotech Product Development. *Pharm Res-Dordr* **2011,** *28* (4), 920-933.
104. Eftink, M. R., Fluorescence Techniques for Studying Protein Structure. In *Methods of Biochemical Analysis*, 1991; pp 127-205.
105. Lakowicz, J. R., *Principles of fluorescence spectroscopy*. 3rd ed.; Springer: New York, 2006.
106. Valeur, B., *Molecular Fluorescence: Principles and Applications*. Wiley: 2002.
107. Ghisaidoobe, A. B. T.; Chung, S. J., Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Forster Resonance Energy Transfer Techniques. *Int J Mol Sci* **2014,** *15* (12), 22518-22538.
108. Burstein, E. A.; Vedenkina, N. S.; Ivkova, M. N., Fluorescence and the location of tryptophan residues in protein molecules. *Photochem Photobiol* **1973,** *18* (4), 263-79.

109. De Lauder, W. B.; Wahl, P., pH dependence of the fluorescence decay of tryptophan. *Biochemistry-Us* **1970,** *9* (13), 2750-2754.
110. White, A., Effect of pH on fluorescence of tryosine, tryptophan and related compounds. *The Biochemical journal* **1959,** *71* (2), 217-220.
111. Wells, T. A.; Nakazawa, M.; Manabe, K.; Song, P. S., A conformational change associated with the phototransformation of Pisum phytochrome A as probed by fluorescence quenching. *Biochemistry-Us* **1994,** *33* (3), 708-12.
112. Lehrer, S., Solute perturbation of protein fluorescence. Quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry-Us* **1971,** *10* (17), 3254-3263.
113. Midoux, P.; Wahl, P.; Auchet, J. C.; Monsigny, M., Fluorescence quenching of tryptophan by trifluoroacetamide. *Biochim Biophys Acta* **1984,** *801* (1), 16-25.
114. Piston, D. W.; Kremers, G.-J., Fluorescent protein FRET: the good, the bad and the ugly. *Trends in Biochemical Sciences* **2007,** *32* (9), 407-414.
115. Wiberg, K.; Sterner-Molin, A.; Jacobsson, S. P., Simultaneous determination of albumin and immunoglobulin G with fluorescence spectroscopy and multivariate calibration. *Talanta* **2004,** *62* (3), 567-574.
116. Kao, S.; Asanov, A. N.; Oldham, P. B., A Comparison of Fluorescence Inner-Filter Effects for Different Cell Configurations. *Instrumentation Science & Technology* **1998,** *26* (4), 375-387.
117. Ryder, A. G.; Stedmon, C. A.; Harrit, N.; Bro, R., Calibration, standardization, and quantitative analysis of multidimensional fluorescence (MDF) measurements on complex mixtures (IUPAC Technical Report). *Pure Appl Chem* **2017,** *89* (12), 1849-1870.
118. Kumar, K.; Mishra, A. K., Application of 'multivariate curve resolution alternating least square (MCR–ALS)' analysis to extract pure component synchronous fluorescence spectra at various wavelength offsets from total synchronous fluorescence spectroscopy (TSFS) data set of dilute aqueous solutions of fluorophores. *Chemometr Intell Lab* **2012,** *116*, 78-86.
119. Casamayou-Boucau, Y.; Ryder, A. G., Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters, validation standards, and Rayleigh scatter removal methods. *Methods Appl Fluores* **2017,** *5* (3).
120. Groza, R. C.; Li, B. Y.; Ryder, A. G., Anisotropy resolved multidimensional emission spectroscopy (ARMES): A new tool for protein analysis. *Anal Chim Acta* **2015,** *886*, 133-142.
121. Groza, R. C. Anisotropy Resolved Multi ‐ dimensional Emission Spectroscopy (ARMES): A new tool for the quantitative and structural analysis of proteins. PhD thesis. National University of Ireland Galway, Galway, Ireland, 2016.
122. Weber, G., Polarization of the fluorescence of macromolecules. I. Theory and experimental method. *The Biochemical journal* **1952,** *51* (2), 145-155.
123. Steiner, R. F., Fluorescence Anisotropy: Theory and Applications. In *Topics in Fluorescence Spectroscopy: Principles*, Lakowicz, J. R., Ed. Springer US: Boston, MA, 2002; pp 1-52.
124. Valeur, B.; Weber, G., Resolution of the fluorescence excitation spectrum of indole into the 1La and iLb excitation bands. . *Photochemistry and Photobiology* **1977,** *25* (5), 441-444.
125. Miller, C. E., Chemometrics in Process Analytical Technology (PAT). In *Process Analytical Technology*, 2010; pp 353-438.

126. Steiner-Browne, M.; Elcoroaristizabal, S.; Casamayou-Boucau, Y.; Ryder, A. G., Investigating Native State Fluorescence Emission of Immunoglobulin G using polarized Excitation Emission Matrix (pEEM) spectroscopy and PARAFAC. . *Chemometrics and Intelligent Laboratory Systems* **2019,** *185*, 1-11.
127. Jolliffe, I., Principal Component Analysis. In *International Encyclopedia of Statistical Science*, Lovric, M., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp 1094-1096.
128. Smilde, A. K.; Bro, R.; Geladi, P., *Multi-way analysis with applications in the chemical sciences*. J. Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2004; p xiv, 381 p.
129. Esbensen, K. H.; Geladi, P., 2.02 - Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice. In *Comprehensive Chemometrics (Second Edition)*, Brown, S.; Tauler, R.; Walczak, B., Eds. Elsevier: Oxford, 2009; pp 3-15.
130. Hubert, M.; Rousseeuw, P. J.; Vanden Branden, K., ROBPCA: A new approach to robust principal component analysis. *Technometrics* **2005,** *47* (1), 64-79.
131. Wold, S.; Sjöström, M.; Eriksson, L., PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab* **2001,** *58* (2), 109-130.
132. Ballabio, D.; Todeschini, R., Multivariate Classification for Qualitative Analysis. *Infrared Spectroscopy for Food Quality Analysis and Control* **2009**, 83-104.
133. Trygg, J.; Holmes, E.; Lundstedt, T., Chemometrics in metabonomics. *J Proteome Res* **2007,** *6* (2), 469-79.
134. Cocchi, M.; Biancolillo, A.; Marini, F., Chapter Ten - Chemometric Methods for Classification and Feature Selection. In *Comprehensive Analytical Chemistry*, Jaumot, J.; Bedia, C.; Tauler, R., Eds. Elsevier: 2018; Vol. 82, pp 265-299.
135. Wold, S.; Sjöström, M., SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. In *Chemometrics: Theory and Application*, American Chemical Society: 1977; Vol. 52, pp 243-282.
136. Cortes, C.; Vapnik, V., Support-Vector Networks. *Mach. Learn.* **1995,** *20* (3), 273–297.
137. Burges, C. J. C., A tutorial on Support Vector Machines for pattern recognition. *Data Min Knowl Disc* **1998,** *2* (2), 121-167.
138. Zavatti, V.; Budman, H.; Legge, R.; Tamer, M., Monitoring of an antigen manufacturing process. *Bioprocess Biosyst Eng* **2016,** *39* (6), 855-69.
139. Claßen, J.; Graf, A.; Aupert, F.; Solle, D.; Höhse, M.; Scheper, T., A novel LED-based 2D-fluorescence spectroscopy system for in-line bioprocess monitoring of Chinese hamster ovary cell cultivations—Part II. *Engineering in Life Sciences* **2019,** *19* (5), 341-351.
140. Pan, H.; Yu, H.; Song, Y.; Zhu, L.; Liu, R.; Du, E., Tracking fluorescent components of dissolved organic matter from soils in large-scale irrigated area. *Environmental Science and Pollution Research* **2017,** *24* (7), 6563-6571.
141. Ohadi, K.; Legge, R. L.; Budman, H. M., Intrinsic fluorescence-based at situ soft sensor for monitoring monoclonal antibody aggregation. *Biotechnol Progr* **2015,** *31* (5), 1423-1432.
142. Li, B. Y.; Shanahan, M.; Calvet, A.; Leister, K. J.; Ryder, A. G., Comprehensive, quantitative bioprocess productivity monitoring using fluorescence EEM spectroscopy and chemometrics. *Analyst* **2014,** *139* (7), 1661-1671.
143. Dramićanin, T.; Lenhardt, L.; Zeković, I.; Dramićanin, M. D., Support Vector Machine on Fluorescence Landscapes for Breast Cancer Diagnostics. *Journal of Fluorescence* **2012,** *22* (5), 1281-1289.

144. Ríos-Reina, R.; Elcoroaristizabal, S.; Ocaña-González, J. A.; García-González, D. L.; Amigo, J. M.; Callejón, R. M., Characterization and authentication of Spanish PDO wine vinegars using multidimensional fluorescence and chemometrics. *Food Chem* **2017,** *230*, 108-116.
145. Azcarate, S. M.; de Araújo Gomes, A.; Alcaraz, M. R.; Ugulino de Araújo, M. C.; Camiña, J. M.; Goicoechea, H. C., Modeling excitation–emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety. *Food Chem* **2015,** *184*, 214-219.
146. Hall, G. J.; Kenny, J. E., Estuarine water classification using EEM spectroscopy and PARAFAC–SIMCA. *Anal Chim Acta* **2007,** *581* (1), 118-124.
147. Veronese, F. M.; Pasut, G., PEGylation, successful approach to drug delivery. *Drug Discov Today* **2005,** *10* (21), 1451-1458.
148. Hermanson, G. T., *Bioconjugate techniques*. 2nd ed.; Elsevier Academic Press: Amsterdam ; Boston, 2008; p xxx, 1202 p.
149. Ellman, G. L., Tissue sulfhydryl groups. *Arch Biochem Biophys* **1959,** *82* (1), 70-7.
150. Han, J. C.; Han, G. Y., A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal Biochem* **1994,** *220* (1), 5-10.
151. Merril, C. R., [36] Gel-staining techniques. In *Methods in Enzymology*, Deutscher, M. P., Ed. Academic Press: 1990; Vol. 182, pp 477-488.
152. Kurfurst, M. M., Detection and Molecular-Weight Determination of Polyethylene Glycol-Modified Hirudin by Staining after Sodium Dodecyl-Sulfate Polyacrylamide-Gel Electrophoresis. *Anal Biochem* **1992,** *200* (2), 244-248.
153. Katayama, D. S.; Nayar, R.; Chou, D. K.; Campos, J.; Cooper, J.; Vander Velde, D. G.; Villarete, L.; Liu, C. P.; Manning, M. C., Solution behavior of a novel type 1 interferon, interferon-tau. *J Pharm Sci-Us* **2005,** *94* (12), 2703-2715.
154. Hawe, A.; Kasper, J. C.; Friess, W.; Jiskoot, W., Structural properties of monoclonal antibody aggregates induced by freeze-thawing and thermal stress. *Eur J Pharm Sci* **2009,** *38* (2), 79-87.
155. Nomine, Y.; Ristriani, T.; Laurent, C.; Lefevre, J. F.; Weiss, E.; Trave, G., A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. *Protein Eng* **2001,** *14* (4), 297-305.
156. Ryder, A. G.; Li, B. Similarity index: a rapid classification method for multivariate data arrays, National University of Ireland, Galway. 7983874, 2009.
157. Resch-Genger, U.; Pfeifer, D.; Monte, C.; Pilz, W.; Hoffmann, A.; Spieles, M.; Rurack, K.; Hollandt, J.; Taubert, D.; Schonenberger, B.; Nording, P., Traceability in fluorometry: Part II. Spectral fluorescence standards. *Journal of Fluorescence* **2005,** *15* (3), 315-336.
158. Ameloot, M.; vandeVen, M.; Acuna, A. U.; Valeur, B., Fluorescence anisotropy measurements in solution: Methods and reference materials (IUPAC Technical Report). *Pure and Applied Chemistry* **2013,** *85* (3), 589-608.
159. Zepp, R. G.; Sheldon, W. M.; Moran, M. A., Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices. *Mar Chem* **2004,** *89* (1-4), 15-36.
160. Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A., Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. *J Chemometr* **2006,** *20* (3-4), 99-105.

161. Rinnan, A.; Booksh, K. S.; Bro, R., First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. *Anal Chim Acta* **2005,** *537* (1-2), 349-358.
162. Zhang, S. Z.; Zhao, F. L.; Li, K. A.; Tong, S. Y., Determination of glycogen by Rayleigh light scattering. *Analytica Chimica Acta* **2001,** *431* (1), 133-139.
163. Savitzky, A.; Golay, M. J. E., Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964,** *36* (8), 1627-1639.
164. Haaland, D. M.; Thomas, E. V., Partial Least-Squares Methods for Spectral Analyses .1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal Chem* **1988,** *60* (11), 1193-1202.
165. Mandel, J.; Linnig, F. J., Study of Accuracy in Chemical Analysis Using Linear Calibration Curves. *Anal Chem* **1957,** *29* (5), 743-749.
166. Wold, S.; Sjöström, M.; Eriksson, L., Partial least squares projections to latent structures (PLS) in chemistry. *Encyclopedia of computational chemistry* **2002,** *3*.
167. Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B., Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy* **2000,** *54* (3), 413-419.
168. de Faria e Silva, A. L.; Elcoroaristizabal, S.; Ryder, A. G., Characterization of lysozyme PEGylation products using polarized excitation-emission matrix spectroscopy. *Biotechnol Bioeng* **2020**, 1–16.
169. Lee, H.; Park, T. G., A novel method for identifying PEGylation sites of protein using biotinylated PEG derivatives. *J Pharm Sci* **2003,** *92* (1), 97-103.
170. Shapiro, A. L.; Viñuela, E.; V. Maizel, J., Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochemical and Biophysical Research Communications* **1967,** *28* (5), 815-820.
171. Bailon, P.; Berthold, W., Polyethylene glycol-conjugated pharmaceutical proteins. *Pharmaceutical Science & Technology Today* **1998,** *1* (8), 352-356.
172. Morgenstern, J.; Baumann, P.; Brunner, C.; Hubbuch, J., Effect of PEG molecular weight and PEGylation degree on the physical stability of PEGylated lysozyme. *International Journal of Pharmaceutics* **2017,** *519* (1-2), 408-417.
173. de Faria e Silva, A. L.; Elcoroaristizabal, S.; Ryder, A. G., Multi-Attribute Quality Screening of Immunoglobulin G using polarized Excitation Emission Matrix Spectroscopy. *Anal Chim Acta* **2019**.
174. Cattani, G.; Vogeley, L.; Crowley, P. B., Structure of a PEGylated protein reveals a highly porous double-helical assembly. *Nature Chemistry* **2015,** *7* (10), 823-828.
175. Masuda, H.; Higashitani, K.; Yoshida, H., *Powder technology : fundamentals of particles, powder beds, and particle generation*. CRC Press/Taylor & Francis: Boca Raton, 2007; p 513 p.
176. Kennard, R. W.; Stone, L. A., Computer Aided Design of Experiments. *Technometrics* **1969,** *11* (1), 137-&.
177. Mehmood, T.; Liland, K. H.; Snipen, L.; Saebo, S., A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems* **2012,** *118*, 62-69.
178. Panigrahi, S. K.; Mishra, A. K., Inner filter effect in fluorescence spectroscopy: As a problem and as a solution. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2019,** *41*.

179. Zhuang, X.; Ha, T.; Kim, H. D.; Centner, T.; Labeit, S.; Chu, S., Fluorescence quenching: A tool for single-molecule protein-folding study. *Proceedings of the National Academy of Sciences* **2000,** *97* (26), 14241.
180. Veronese, F. M., Peptide and protein PEGylation: a review of problems and solutions. *Biomaterials* **2001,** *22* (5), 405-417.
181. Rathore, A. S.; Winkle, H., Quality by design for biopharmaceuticals. *Nat Biotechnol* **2009,** *27* (1), 26-34.
182. Wanning, S.; Suverkrup, R.; Lamprecht, A., Pharmaceutical spray freeze drying. *International Journal of Pharmaceutics* **2015,** *488* (1-2), 136-153.
183. Dotz, V.; Haselberg, R.; Shubhakar, A.; Kozak, R. P.; Falck, D.; Rombouts, Y.; Reusch, D.; Somsen, G. W.; Fernandes, D. L.; Wuhrer, M., Mass spectrometry for glycosylation analysis of biopharmaceuticals. *Trac-Trends in Analytical Chemistry* **2015,** *73*, 1-9.
184. Lipman, N. S.; Jackson, L. R.; Trudel, L. J.; Weis-Garcia, F., Monoclonal versus polyclonal antibodies: Distinguishing characteristics, applications, and information resources. *Ilar J* **2005,** *46* (3), 258-268.
185. Diamant, E.; Torgeman, A.; Ozeri, E.; Zichel, R., Monoclonal Antibody Combinations that Present Synergistic Neutralizing Activity: A Platform for Next-Generation Anti-Toxin Drugs. *Toxins* **2015,** *7* (6), 1854-1881.
186. Warner, I. M.; Davidson, E. R.; Christian, G. D., Quantitative-Analyses of Multicomponent Fluorescence Data by Methods of Least-Squares and Nonnegative Least Sum of Errors. *Anal Chem* **1977,** *49* (14), 2155-2159.
187. Chan, F. T. S.; Schierle, G. S. K.; Kumita, J. R.; Bertoncini, C. W.; Dobson, C. M.; Kaminski, C. F., Protein amyloids develop an intrinsic fluorescence signature during aggregation. *Analyst* **2013,** *138* (7), 2156-2162.
188. Tikhonova, T. N.; Rovnyagina, N. R.; Zherebker, A. Y.; Sluchanko, N. N.; Rubekina, A. A.; Orekhov, A. S.; Nikolaev, E. N.; Fadeev, V. V.; Uversky, V. N.; Shirshin, E. A., Dissection of the deep-blue autofluorescence changes accompanying amyloid fibrillation. *Arch Biochem Biophys* **2018,** *651*, 13-20.
189. Eilers, P. H. C.; Kroonenberg, P. M., Modeling and correction of Raman and Rayleigh scatter in fluorescence landscapes. *Chemometr Intell Lab* **2014,** *130*, 1-5.
190. Bohren, C. F.; Huffman, D. R., *Absorption and Scattering of Light by Small Particles*. Wiley: 1983.
191. Joshi, V.; Shivach, T.; Yadav, N.; Rathore, A. S., Circular Dichroism Spectroscopy as a Tool for Monitoring Aggregation in Monoclonal Antibody Therapeutics. *Anal Chem* **2014,** *86* (23), 11606-11613.
192. Hong, P.; Koza, S.; Bouvier, E. S. P., A Review Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and Their Aggregates. *J Liq Chromatogr R T* **2012,** *35* (20), 2923-2950.
193. Eppler, A.; Weigandt, M.; Hanefeld, A.; Bunjes, H., Relevant shaking stress conditions for antibody preformulation development. *Eur J Pharm Biopharm* **2010,** *74* (2), 139-147.
194. Rudiuk, S.; Cohen-Tannoudji, L.; Huille, S.; Tribet, C., Importance of the dynamics of adsorption and of a transient interfacial stress on the formation of aggregates of IgG antibodies. *Soft Matter* **2012,** *8* (9), 2651-2661.
195. Aitken, A.; Learmonth, M. P., Protein Determination by UV Absorption. In *Protein Protocols Handbook, Third Edition*, Walker, J. M., Ed. Humana Press Inc: Totowa, 2009; pp 3-6.
196. Gonzalez, A. G.; Herrador, M. A.; Asuero, A. G., Intra-laboratory testing of method accuracy from recovery assays. *Talanta* **1999,** *48* (3), 729-736.

197. Li, B.; Ryan, P. W.; Shanahan, M.; Leister, K. J.; Ryder, A. G., Fluorescence excitation-emission matrix (EEM) spectroscopy for rapid identification and quality evaluation of cell culture media components. *Appl Spectrosc* **2011,** *65* (11), 1240-9.
198. Steiner-Browne, M.; Elcoroaristizabal, S.; Casamayou-Boucau, Y.; Ryder, A. G., Investigating native state fluorescence emission of Immunoglobulin G using polarized Excitation Emission Matrix (pEEM) spectroscopy and PARAFAC. *Chemometr Intell Lab* **2019,** *185*, 1-11.
199. Steiner-Browne, M.; Elcoroaristizabal, S.; Ryder, A. G., Using polarized Total Synchronous Fluorescence Spectroscopy (pTSFS) with PARAFAC analysis for characterizing intrinsic protein emission. *Chemometr Intell Lab* **2019,** *194*, 103871.
200. Philo, J. S., Is any measurement method optimal for all aggregate sizes and types? *Aaps J* **2006,** *8* (3), E564-E571.
201. Chaudhuri, R.; Cheng, Y.; Middaugh, C. R.; Volkin, D. B., High-throughput biophysical analysis of protein therapeutics to examine interrelationships between aggregate formation and conformational stability. *The AAPS journal* **2014,** *16* (1), 48-64.
202. O'Donnell, I. J.; Frangione, B.; Porter, R. R., The disulphide bonds of the heavy chain of rabbit immunoglobulin G. *Biochem J* **1970,** *116* (2), 261-8.
203. Makaraviciute, A.; Jackson, C. D.; Millner, P. A.; Ramanaviciene, A., Considerations in producing preferentially reduced half-antibody fragments. *J Immunol Methods* **2016,** *429*, 50-56.
204. Sun, M. M. C.; Beam, K. S.; Cerveny, C. G.; Hamblett, K. J.; Blackmore, R. S.; Torgov, M. Y.; Handley, F. G. M.; Ihle, N. C.; Senter, P. D.; Alley, S. C., Reduction-alkylation strategies for the modification of specific monoclonal antibody disulfides. *Bioconjugate Chem* **2005,** *16* (5), 1282-1290.
205. Behrens, C. R.; Ha, E. H.; Chinn, L. L.; Bowers, S.; Probst, G.; Fitch-Bruhns, M.; Monteon, J.; Valdiosera, A.; Bermudez, A.; Liao-Chan, S.; Wong, T.; Melnick, J.; Theunissen, J. W.; Flory, M. R.; Houser, D.; Venstrom, K.; Levashova, Z.; Sauer, P.; Migone, T. S.; van der Horst, E. H.; Halcomb, R. L.; Jackson, D. Y., Antibody-Drug Conjugates (ADCs) Derived from Interchain Cysteine Cross-Linking Demonstrate Improved Homogeneity and Other Pharmacological Properties over Conventional Heterogeneous ADCs. *Mol Pharm* **2015,** *12* (11), 3986-98.
206. Gottschalk, U., *Process scale purification of antibodies*. John Wiley & Sons: Hoboken, N.J., 2009; p 1 online resource (xxvi, 430 p.).
207. Cheng, X.; Li, J.; Tanaka, K.; Majumder, U.; Milinichik, A. Z.; Verdi, A. C.; Maddage, C. J.; Rybinski, K. A.; Fernando, S.; Fernando, D.; Kuc, M.; Furuuchi, K.; Fang, F.; Uenaka, T.; Grasso, L.; Albone, E. F., MORAb-202, an Antibody–Drug Conjugate Utilizing Humanized Anti-human FRα Farletuzumab and the Microtubule-targeting Agent Eribulin, has Potent Antitumor Activity. *Molecular Cancer Therapeutics* **2018,** *17* (12), 2665-2675.
208. McDonagh, C. F.; Kim, K. M.; Turcott, E.; Brown, L. L.; Westendorf, L.; Feist, T.; Sussman, D.; Stone, I.; Anderson, M.; Miyamoto, J.; Lyon, R.; Alley, S. C.; Gerber, H. P.; Carter, P. J., Engineered anti-CD70 antibody-drug conjugate with increased therapeutic index. *Mol Cancer Ther* **2008,** *7* (9), 2913-23.
209. Stefano, J. E.; Busch, M.; Hou, L. H.; Park, A.; Gianolio, D. A., Micro- and Mid-Scale Maleimide-Based Conjugation of Cytotoxic Drugs to Antibody Hinge Region Thiols for Tumor Targeting. *Methods Mol Biol* **2013,** *1045*, 145-171.
210. Harding, F. A.; Stickler, M. M.; Razo, J.; DuBridge, R. B., The immunogenicity of humanized and fully human antibodies Residual immunogenicity resides in the CDR regions. *Mabs-Austin* **2010,** *2* (3), 256-265.

211. McCombs, J. R.; Owen, S. C., Antibody Drug Conjugates: Design and Selection of Linker, Payload and Conjugation Chemistry. *Aaps J* **2015,** *17* (2), 339-351.
212. Chen, W.; Westerhoff, P.; Leenheer, J. A.; Booksh, K., Fluorescence Excitation−Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter. *Environ Sci Technol* **2003,** *37* (24), 5701-5710.
213. Coble, P. G., Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar Chem* **1996,** *51* (4), 325-346.
214. Rolinger, L.; Rüdt, M.; Hubbuch, J., A critical review of recent trends, and a future perspective of optical spectroscopy as PAT in biopharmaceutical downstream processing. *Analytical and Bioanalytical Chemistry* **2020,** *412* (9), 2047-2064.
215. Rathore, A. S.; Bhambure, R.; Ghare, V., Process analytical technology (PAT) for biopharmaceutical products. *Anal Bioanal Chem* **2010,** *398* (1), 137-54.
216. Casamayou-Boucau, Y.; Ryder, A. G., Quantitative analysis of weakly bound insulin oligomers in solution using polarized multidimensional fluorescence spectroscopy. *Anal Chim Acta* **2020,** *1138*, 18-29.
217. Yu, D.; Ghosh, R., Purification of PEGylated Protein Using Membrane Chromatography. *J Pharm Sci-Us* **2010,** *99* (8), 3326-3333.
218. Morgenstern, J.; Busch, M.; Baumann, P.; Hubbuch, J., Quantification of PEGylated proteases with varying degree of conjugation in mixtures: An analytical protocol combining protein precipitation and capillary gel electrophoresis. *Journal of Chromatography A* **2016,** *1462*, 153-164.
219. Joubert, M. K.; Luo, Q. Z.; Nashed-Samuel, Y.; Wypych, J.; Narhi, L. O., Classification and Characterization of Therapeutic Antibody Aggregates. *J Biol Chem* **2011,** *286* (28), 25118-25133.
220. Paul, A. J.; Schwab, K.; Hesse, F., Direct analysis of mAb aggregates in mammalian cell culture supernatant. *Bmc Biotechnol* **2014,** *14*.
221. Wagner-Rousset, E.; Janin-Bussat, M.-C.; Colas, O.; Excoffier, M.; Ayoub, D.; Haeuw, J.-F.; Rilatt, I.; Perez, M.; Corvaïa, N.; Beck, A., Antibody-drug conjugate model fast characterization by LC-MS following IdeS proteolytic digestion. *Mabs-Austin* **2014,** *6* (1), 273-285.
222. Abe, Y.; Sugihara, K.; Nakada, T.; Shahidi, J.; Gallant, G. J. A.; Jikoh, T.; Agatsuma, T., ADCs on the Market and in Clinical Development. In *Cancer Drug Delivery Systems Based on the Tumor Microenvironment*, Matsumura, Y.; Tarin, D., Eds. Springer Japan: Tokyo, 2019; pp 155-174.
223. JiJi, R. D.; Cooper, G. A.; Booksh, K. S., Excitation-emission matrix fluorescence based determination of carbamate pesticides and polycyclic aromatic hydrocarbons. *Anal Chim Acta* **1999,** *397* (1), 61-72.
224. Zielinski, O.; Rüssmeier, N.; Ferdinand, O. D.; Miranda, M. L.; Wollschläger, J., Assessing Fluorescent Organic Matter in Natural Waters: Towards In Situ Excitation–Emission Matrix Spectroscopy. *Applied Sciences* **2018,** *8* (12), 2685.
225. Deckert, V.; Kiefer, W., Scanning Multichannel Technique for Improved Spectrochemical Measurements with a CCD Camera and its Application to Raman Spectroscopy. *Appl Spectrosc* **1992,** *46* (2), 322-328.
226. Cohen, S., Antibody structure. *J Clin Pathol Suppl (Assoc Clin Pathol)* **1975,** *6*, 1-7.
227. Kindt, T. J.; Goldsby, R. A.; Osborne, B. A.; Kuby, J., *Kuby immunology*. 6th ed.; W.H. Freeman: New York, 2007; p xxii, 574, I-27 p.
228. Janeway CA Jr, T. P., Walport M, et al, Immunobiology: The Immune System in Health and Disease - 5th edition. In *Garland Science*, New York, 2001.

229. Johnstone, A.; Thorpe, R., *Immunochemistry in practice*. 3rd ed.; Blackwell Science: Cambridge, Mass., 1996; p xv, 362 p., 7 p. of plates.
230. Birdsall, B.; King, R. W.; Wheeler, M. R.; Lewis, C. A.; Goode, S. R.; Dunlap, R. B.; Roberts, G. C. K., Correction for Light-Absorption in Fluorescence Studies of Protein-Ligand Interactions. *Anal Biochem* **1983,** *132* (2), 353-361.
231. ISO; IEC; OIML; BIPM, GUIDE 98-1 Guide to the Expression of Uncertainty in Measurement. Geneva, Switzerland v.122, 1995.
232. Kothawala, D. N.; Murphy, K. R.; Stedmon, C. A.; Weyhenmeyer, G. A.; Tranvik, L. J., Inner filter correction of dissolved organic matter fluorescence. *Limnol Oceanogr-Meth* **2013,** *11*, 616-630.

Publications and Research presentations

List of publications

Ana Luiza de Faria e Silva*, Saioa Elcoroaristizabal, Alan G. Ryder / Multi-attribute quality screening of immunoglobulin G using polarized Excitation Emission Matrix spectroscopy, *Analytica Chimica Acta* 1101 (2020) 99-110. https://doi.org/10.1016/j.aca.2019.12.020

Ana Luiza de Faria e Silva*, Saioa Elcoroaristizabal, Alan G. Ryder Characterization of Lysozyme PEGylation products using polarised Excitation Emission Matrix (pEEM) spectroscopy. – *Biotechnology & Bioengineering* https://doi.org/10.1002/bit.27483

List of research presentations

International

Poster, "Characterising PEGylation products using polarized Excitation Emission Matrix (pEEM) spectroscopy." A. L. F Silva* , S. Elcoroaristizabal, and A.G. Ryder. *Bioprocessing Summit Europe*, July, 2020 (Online conference)

Poster, "Multi-attribute quality screening of immunoglobulin G using polarized Excitation Emission Matrix spectroscopy." A. L. F Silva* , S. Elcoroaristizabal, and A.G. Ryder. *Bioprocessing Summit Europe*, July, 2020 (Online conference)

Oral, "Assessing IgG PEGylation reactions and products using polarized Excitation Emission Matrix (pEEM) spectroscopy." A. L. F Silva* , S. Elcoroaristizabal, and A.G. Ryder. *International Forum Process Analytical Chemistry (IFPAC)*, Maryland, USA, March, 2019.

Poster, 2nd prize - "Assessing IgG PEGylation reactions and products using polarized Excitation Emission Matrix (pEEM) spectroscopy." A. L. F Silva* , S. Elcoroaristizabal, and A.G. Ryder. *International Forum Process Analytical Chemistry (IFPAC)*, Maryland, USA, March, 2019.

Poster, - Assessing IgG quality using polarized Excitation Emission Matrix (pEEM) spectroscopy." A. L. F. Silva* , S. Elcoroaristizabal, and A.G. Ryder. *International Forum Process Analytical Chemistry (IFPAC),* Maryland, USA, March, 2019.

Poster, Characterizing IgG conjugation products using Anisotropy Resolved Multi-Dimensional Spectroscopy (ARMES)." A. L. F Silva* , S. Elcoroaristizabal, Y. Casamayou-Boucau, and A.G. Ryder. *15th Conference on Methods and Applications of Fluorescence: Spectroscopy, Imaging and Probes*, Bruges, Belgium, Sept., 2017.

National

Poster, "Assessing IgG quality using polarized Excitation Emission Matrix (pEEM) spectroscopy." A. L. F Silva* , S. Elcoroaristizabal, and A.G. Ryder. *BioITG Research Showcase,* Dublin, Ireland, April, 2019

Poster, "Characterizing PEGylation reactions using multidimensional fluorescence (MDF) spectroscopy. A. L. F Silva*, S. Elcoroaristizabal, Y. Casamayou-Boucau* and A.G. Ryder, Biopharma Ambition, Dublin, Feb. 2018.

Poster, "Monitoring IgG conjugation reactions by Anisotropy Resolved Multi-Dimensional Spectroscopy (ARMES): preliminary results. A. L. F Silva* and A.G. Ryder, *NUI Galway/UL Alliance Research Day,* April, 2017;