Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork

Application of fluorescence spectroscopy for the in-vial investigation of protein solutions

Thesis presented by

Khaled ElKassas

[0000-0002-2905-4358]

for the degree of

Doctor of Philosophy

University College Cork

School of Pharmacy

Head of School: Prof. Brendan Griffin

Supervisor(s): Prof. Abina Crean, Dr. Sonja Vucen

2024

To my dearest parents
For believing in me, despite all

“The complicated futility of ignorance”

- Kurt Vonnegut

Table of Contents

Declaration

This is to certify that the work I am submitting is my own and has not been submitted |for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.

The construction of the bespoke fluorescence instrument was completed by Dr. Krishnakumar Chullipalliyalil, Centre for Advanced Photonics & Process Analysis, Munster Technical University, Ireland. Experimental work in Chapters 2, 3, and 5 were designed and completed in collaboration with Dr. Krishnakumar Chullipalliyalil, and Dr. Michael McAuliffe at the Centre for Advanced Photonics & Process Analysis, Munster Technical University, Ireland.

This thesis has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund [grant numbers 12/RC/2275(P2) and 18/EPSRC-CDT/3587] and the Engineering and Physical Sciences Research Council UK [grant number EP/S023054/1].

Signed: _____

Khaled ElKassas

Acknowledgements

Dear reader, the following is a feeble attempt to recognise the unending graces of all the people who have contributed to the realisation of this work. Despite my own hands writing what you are about to read, it was only made possible by the presence of those who have supported me and shaped my world in major and minor ways all the same. Thus, I give my utmost thanks and appreciation to the following.

First and foremost, to my parents Hassan ElKassas and Azza Abdelghany who believed in me and my wholly unattainable dreams of being a scientist and fought tooth and nail to somehow make them a possibility. For their unconditional love and support when it was easy, and especially when it was not. A special thanks to my sister Salma, who is my light shining in the darkness, and who has never failed to put a smile on my face. To my one and only Maria, for love and support unending, for bearing with me through thick and thin, and for always believing in me. For always giving her most, and for enriching my life with her mere presence. For always being my *porto de abrigo* when I was most lost. To her family who welcomed me into their home with open arms, particularly her parents Sr. Paulo Monteiro, and D. Manuela Ferreira who have cheered me on wholeheartedly. To her brothers Luis and Tiago, who were the brothers I always wished I had.

To my supervisors, Dr. Abina Crean, and Dr. Sonja Vucen, no thanks I could give could possibly be enough. For always caring, for selflessly toiling, and always prioritising my education above all else. For their unending patience, and the impossible feat of making the past 4 years of study a breeze. For setting the unreasonable standard of what an academic should be, as a researcher and a person. Finally, and above all else, for being kind. Words of thanks pale in comparison to what they have given me.

To Michael McAuliffe for his continued support throughout my studies. To Krishnakumar for his invaluable insights, his infectious curiosity and passion for learning, and for teaching me that following the data will inevitably lead to interesting and meaningful knowledge.

To all the UCC staff who offered support and encouragement throughout the years, Dr. Ken Devine, Dr. Michael Cronin, Ms. Aine Healy, and Dr. Joseph O'Shea. A special thanks to Ms. Noreen Moynihan these thanks are what little I can offer for bearing through the paperwork-induced headaches she suffered through on my behalf. A most humble thanks to Dr. Fatma Farag, and Dr. Waleed Faisal for their invaluable support and encouragement. Finally, to the SSPC and CDT for providing funding for this project, and for introducing me to a wonderful community of welcoming and giving academics and operational staff.

To the UCC cohorts present and past, for teaching me what a good and welcoming research community should be like, to Valeria, Anna-Luiza, Niklas, Laura, and Ziad who made me feel welcome my first time far away from home. To my fellow seniors Ayse, Annachiara and Brian, for making those many long cold days feel shorter and warmer. To my friends John and Andrew, for a laugh shared and time well spent. To Michael, Callum, and Lotte for listening to my waffling, always ready with a sharp reply. To everyone who shared their good times with me, Claudia, Maja, Aida, Lianne, Laura, Jack, and all the rest. To Monique for endless wisdom and a great spirit that kept everyone going during the hard times. To the good folks of various CDT cohorts Toby, Fiona, and all the rest, for all the laughs and banter we shared.

To my partner in crime, and my dearest friend, the soon-to-be Dr. Evangelia Tsolaki for always having a shoulder to spare, and a good laugh ready whenever I most needed them. To Henry for a wit without equal, and a good heart to match.

To the second family I was blessed with but do not deserve, Dr. Gehad and Dr. Rasha, I would not be here without them.

To all my friends back home, who always made me feel cared for, Somaia, and Farah. To Khalid, my best friend, no matter how far we move apart, or how long a time we spend away. To my university brothers, who always kept me going Ashraf and Nassar. To Ahmed and Yasmine for keeping me company when I was loneliest. To all my friends lost in the big world, Galal, Amjad, and Abdelrahman El-Adawy for teaching me to be myself, unapologetically. To my friends abroad and far away, but always close to my heart Abdelrahman, Mohammed Salah, Moemen, Yasser, and Bilal whose perseverance always kept me going knowing that I was not alone. To all my friends in the Cork chess community Eimhan, Piotr, Suneet, Jekabs, and all the rest, for welcoming me with open arms, and teaching me so much in so little time. Finally, and last but in no way least, to a friend lost but never forgotten, to Yostina, for encouraging me to follow my dreams and persevere despite all odds.

I would not have been able to finish this work without any of the aforementioned people, and I hope that my acknowledgment of their efforts, and giving spirits suffices as simple thanks.

Publications and Presentations

Publications

- Khaled ElKassas and Krishnakumar Chullipalliyalil, Michael McAuliffe, Sonja Vucen, Abina Crean

  Fluorescence spectroscopy for the determination of reconstitution time of an in-vial lyophilised product

  International Journal of Pharmaceutics. 2021, Vol 597, 120368

  https://doi.org/10.1016/j.ijpharm.2021.120368

- Krishnakumar Chullipalliyalil and Khaled ElKassas, Michael McAuliffe, Sonja Vucen, Abina Crean

  In-Vial Detection of Protein Denaturation Using Intrinsic Fluorescence Anisotropy

  Analytical Chemistry. 2023, 95, 5, 2774–2782

  https://doi.org/10.1021/acs.analchem.2c03912

Oral Presentations

- Khaled ElKassas, Krishnakumar Chullipalliyalil, Michael McAuliffe, Sonja Vucen, Abina Crean

  In-Vial Detection of Protein Denaturation Using Intrinsic Fluorescence Anisotropy

  4th European Meeting on Pharmaceutics, Marseille, France, 2023

- Khaled ElKassas, Krishnakumar Chullipalliyalil, Michael McAuliffe, Sonja Vucen, Abina Crean

  Spectroscopic Insights into the Behaviours of Protein Formulations

  All-Ireland Schools of pharmacy, Queen's University Belfast, 2023

Poster Presentations

- Khaled ElKassas, Krishnakumar Chullipalliyalil, Michael McAuliffe, Sonja Vucen, Abina Crean

  In-vial detection of protein denaturation using intrinsic fluorescence anisotropy

  UKICRS symposium, online, 2020

List of Figures

List of Tables

Abbreviations

Arg	Arginine
Asn	Asparagine
Asp	Aspartate
BSA	Bovine serum albumin
CD	Circular dichroism
CDNN	Circular dichroism neural networking
CRAN	Comprehensive R archive network
CV	Cross-validation
Cys	Cysteine
DLS	Dynamic light scattering
DSF	Differential scanning fluorimetry
EEM	Excitation-emission matrices
EDM	E-divisive with medians
FC-NNLS	Fast combinatorial non-negative least squares solver
FCS	Fluorescence correlation spectroscopy
FFT CCD	Full frame transfer charge coupled device
FTIR	Fourier transform infra-red spectroscopy
Gln	Glutamine
Glu	Glutamate
His	Histidine
HMW	High molecular weight species
HPLC	High-performance liquid chromatography
IOR	Index of refraction
LED	Light-emitting diode
LMW	Low molecular weight species
LOD	Limit of detection
LOOCV	Leave-one-out cross-validation

Lys	Lysine
mAbs	Monoclonal antibodies
MCR/-ALS	Multivariate curve resolution/- Alternating Least squares
MLR	Multiple linear regression
MVA	Multivariate analysis
PCA	Principal component analysis
Phe	Phenylalanine
pI	Isoelectric point
PLSR	Partial least squares regression
PRESS	Predicted residual error sum of squares
R^2	Coefficient of determination
RMSEC	Root-mean-standard error for calibration
RMSECV	Root-mean-standard error for cross-validation
ROS	Reactive oxygen species
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate - Polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SEM	Scanning electron microscopy
smFRET/FRET	Single molecule/Förster resonance energy transfer
SNV	Standard normal variate
Trp	Tryptophan
Tyr	Tyrosine
UPLC	Ultra-performance liquid chromatography
UV	Ultraviolet
WFI	Water for injection

Abstract

Therapeutic protein formulations are subjected to various stressors during manufacturing, transport, and storage, causing destabilisation, in turn leading to deleterious effects such as immunogenic reactions or inefficacy upon administration. An essential part of this production and supply process is establishing protein formulation stability at all stages.

Spectroscopy is a common laboratory tool for the characterisation of therapeutic proteins. Conventional techniques for determining protein denaturation are lab-based and require sample removal from the sealed primary packaging vial and sample destruction. To overcome the limitations of traditional techniques, the development of a fluorescence-based spectroscopic technique for “in-vial” protein analysis was a key thesis aim.

A fluorescence excitation wavelength of 365 nm was found and confirmed to be transmissible through borosilicate glass vial walls. The excitation resulted in an intrinsic emission at 462 nm for a model protein, bovine serum albumin (BSA). Using a bespoke apparatus, the change in fluorescence signal with time was monitored over the reconstitution period of lyophilised protein formulations, without breach of the vial seal. First, we demonstrated how analysis of changes in fluorescence signal with time during reconstitution, using principal component analysis, could be used to determine an instrumentally quantified reconstitution time. At high protein concentrations (10% w/v BSA) the variability of the reconstitution time measurements was reduced from 80.4% relative standard deviation obtained via the pharmacopeial visual method to 8.2% for the instrumental method. Spectroscopic measurements resulted in longer measured reconstitution time compared to the visual method, possibly owing to the detection of subvisible particulates undergoing dissolution.

In the second part of this work, the bespoke spectroscopic apparatus was modified with polarising filters to monitor protein aggregation and denaturation in-vial. Detector measurement of the emission wavelength (365 nm) allowed measurement of protein aggregation via light scattering. Fluorescence anisotropy (a measurement of optical polarizability at different angles) at 462 nm was used to determine protein conformational changes.

Fluorescence scattering and anisotropy measurements for freshly prepared, aggregated, and denatured BSA solutions were compared to reference circular dichroism (CD) and size-exclusion chromatography (SEC) measurements. Anisotropy analysis of thermal unfolding showed strong correlations with secondary structure as quantified through a CD-neural networking model. Scattering correlated with large molecular weight aggregate measurements via SEC.

A study was conducted to find the relationship between the fluorescent amino acids and the overall protein fluorescence for 4 different proteins. Using a multivariate technique, multivariate curve resolution to resolve 3D emission excitation spectra, the fluorescence of individual amino acid solutions and the total fluorescence of the proteins was correlated.

Finally, a partial least squares regression (PLSR) model relating the anisotropy and scattering measurements to reference CD and SEC measurements was established from a sample set of standardised and stored samples. A total of 72 spectra were collected for the batch standardised protein solutions and the stored stability study samples. Samples were stored under real-world-like stress conditions (temperature, shaking, ambient light exposure). The fluorescence anisotropy and scattering results were included as the predictors vs. CD and SEC measurements as the responses. The PLSR showed strong correlations with an average R^2 of 0.87 and a root mean PRESS of 0.487 at 9 latent variables.

In conclusion, the combination of bespoke apparatus and 365 nm excitation was capable of monitoring protein stability in-vial successfully. The spectroscopic measurement of reconstitution time resulted in more precise measurements. The method overcomes the challenges present in the current pharmacopeial standard measurement as it can detect subvisible particulates and is independent of analyst capabilities and subjectivity. The protein solution stability monitoring provided a holistic non-destructive alternative to multi-instrument analysis while maintaining reasonable correlations with established reference stability indicators. This research provides a platform for a cost-effective portable solution to provide a top-level overview of biopharmaceutical product stability in-vial, from manufacture to the point of administration.

Chapter 1
General Introduction

1. General Introduction

1.1. Protein biopharmaceuticals

Biopharmaceuticals, also known as biologic drugs or biologics, are therapeutics produced as a result of certain metabolic processes in living organisms. The protein biopharmaceutical market includes recombinant gene therapeutics, monoclonal antibodies (mAbs), Fc-based fusion proteins, and vaccines, and has grown considerably over the past years making those products more available and less costly. The rise of biopharmaceuticals can be attributed to their effectiveness in treating persistent conditions that do not traditionally respond to small molecule treatment such as chronic illness, and cancers. For example, monoclonal antibodies are a popular type of biopharmaceuticals with around 1,500 drugs in the product pipeline as of 2016 (Informa, 2016). As the understanding of diseases at the molecular and cellular levels continues to grow, the development of mAbs is expected to expand further. This will result in the continued growth of sales for existing mAb products and the increasing success of mAb product candidates currently in development. It is projected that sales of mAb products and all biopharmaceuticals will continue to rise, with an expected increase of 9.22% by the year 2028 (Ecker et al., 2015).

1.1.1. Protein stability considerations

Proteins are polypeptide molecules consisting of amino acid residues in single or multiple chains. They are considered medically relevant as they serve a host of biological purposes functioning as transport carriers, metabolic catalysts, nucleic acid replicators, and structural supports for cells. Protein chains tend to conform to specific 3-dimensional structures due to interactions between residue moieties. Protein arrangement in 3D space

is divided into levels of primary, secondary, and tertiary structure for all proteins, and quaternary structure for only some (Jaenicke, 1993).

Primary protein structure is the linear chain of amino acids and is based on a covalent amide linkage. Secondary structure features are formed through hydrogen bonding with the polypeptide backbone and include α-helices and β-sheets. Bonding interactions between residues on the polypeptide backbone create the 3-dimensional tertiary structure. Finally, in some cases, interactions between the polypeptide side chains create quaternary structures binding polypeptide subunits in a multi-subunit complex (Guo et al., 2005).

The sequence of amino acids and their 3D conformational positions are essential to protein functions which predominantly occur through binding domains. Combinations of secondary conformational features allow the binding of proteins to specific ligands or receptors, allowing them to perform specific functions. Proteins typically consist of several domains geared towards complementary functions. As an example, an immunoglobulin consists of an antigen-binding domain as well as the regulatory Fc region (Dobson, 2003; Gething and Sambrook, 1992).

1.1.1.1. Conformational stability

The loss of the ordered protein 3D structures is defined as unfolding. Protein function is contingent on the specific sequence, folding, and binding capabilities of a protein, and any disruptions may lead to partial or complete loss of activity.

Temperature

Temperatures above the basal temperatures of proteins' native organisms are known denaturants. Cold denaturation has been suggested as another mechanism of temperature denaturation. However, the cold melting point of proteins is rarely above 0 degrees and is

therefore unlikely to occur before the freezing of water and immobilisation of protein structure. Temperature unfolding of proteins is a well-characterised process (Perrot, 1998; Privalov, 1997). Process predictions have generally used the Gibbs free energy difference between the folded and unfolded states to describe the transfer. The application of the Gibbs free energy equation to protein unfolding as described by Privalov is shown in Eq. 1.1 (Privalov, 1997).

$$\Delta G = \Delta H^o - T\Delta S^o \qquad \textit{Eq. 1.1}$$

As the temperature (T) increases the Gibbs free energy (ΔG^o) decreases till it reaches a negative value where the unfolded state has a lower free energy than the folded one.

Solution pH

Denaturants like acid or base cause protein unfolding, leading to a disordered conformation (Goto et al., 1990). Removal of the denaturant allows most proteins to spontaneously refold to their native state. However, denaturation can become irreversible at pH extremes (Kishore et al., 2012). At pH corresponding to the protein's isoelectric point (pI), the protein will carry no net charge. As the pH shifts away from the pI the repulsion between the similarly charged amino acid residues increases, destabilising the protein structure. At pH extremes the native structure becomes unstable due to its high charge density, and the protein shifts to an unfolded state with lower charge density and lower free energy (Dill, 1990). An additional secondary mechanism is the inability of the protein's internal residues to form salt bridges at pH extremes contributing to the unfolding tendency (Wang, 1999).

It is important to note that the above does not imply the absolute stability of the protein at the pI. The solubility of the protein is also tied to the pH of the solution, as at the pI, proteins can become less soluble, and thus unstable (Shaw et al., 2001).

The presence of acids and bases has been shown to disrupt the hydrogen bonds between sidechains thus leading to denaturation (Yang and Honig, 1993). The most likely sites of protonation are the basic amino acid residues such as lysine (Lys), arginine(Arg), and histidine (His). Whereas the acidic residues such as glutamic and aspartic acids are likely to be the sites of deprotonation (Kelly et al., 1992).

The mechanisms of acid unfolding involve the protonation of the aforementioned groups within the protein structure. Studies have shown that the acid-unfolding reaction proceeds through multiple intermediates and is a result of the higher affinity of binding of internal residues to protons in the unfolded protein state (Tanford, 1968). The abundance of protons in essence increases the pK_a of the protein and unfolded state equilibrium (Fink et al., 1990; Tanford, 1968). In extremely basic conditions (>11.0 pH) protein unfolding has been shown to proceed in a similar mechanism to acidic unfolding (Guckeisen et al., 2021; Prajapati et al., 1998).

Protein functions are strongly tied to the pH of the solution. Studies by Dwevedi et al. and Kishore et al. have demonstrated the drop of catalytic efficiency of proteins under non-ideal pH conditions (Dwevedi et al., 2010; Kishore and Kayastha, 2012).

Chemical degradation

Protein unfolding is essential for common characterisation techniques for protein separation like gel electrophoresis. Chemical denaturants like guanidine hydrochloride, urea, and sodium dodecyl sulphate (SDS) are used to unfold the protein prior to analysis (Camilloni et al., 2008; Parui et al., 2016).

Specifically, binding chemical agents such as guanidine hydrochloride and urea are known to cause changes in secondary structure and denature proteins. Through hydrogen bonding with the amino acid backbone, urea molecules disrupt native backbone bonding.

As a result, a straighter backbone conformation becomes more favourable. On the other hand, guanidine interacts with inner residues and disrupts hydrophobic interactions between π-stacked moieties (Lapidus, 2017; Winogradoff et al., 2020).

Chemical denaturants are often used in a lab setting to induce protein unfolding (Calloni et al., 2008; Chen et al., 2009; Lai et al., 1997; Parui et al., 2016). However, due to the kinetic nature of the protein-denaturant interactions, the initial dilution phase (upon the addition of a denaturant concentrated solution) can cause inhomogeneity, which can complicate interaction predictions (Lapidus, 2017).

Surfactants are also known to unfold proteins, SDS is commonly used for the preparation of proteins for gel electrophoresis. The mechanism of protein denaturation by surfactants has been a matter of contention in bio-physical research communities. However, recent advances in in-silico predictions have allowed researchers to identify the specific interactions between protein molecules and surfactants (Nielsen et al., 2007; Winogradoff et al., 2020).

Winogradoff et al. used multiple all-atom molecular dynamics simulations to demonstrate two mechanisms of interaction of SDS with the proteins β-amylase and titin as well as a micellar interaction. The study indicates that SDS binds electrostatically and can insert between strands and β-sheets and wedge them apart, leading to a disruption in secondary structures. Furthermore, micelles of SDS cause the protein to wind itself into the micellar structure in a “necklace-and-beads configuration”. The micelles are dynamically bound to the protein, and the number and configuration of the micelles lead to different unfolding patterns (Winogradoff et al., 2020).

The unfolding of protein by the addition of surfactant is known to be highly variable. The kinetics of the disruptions can involve a large number of regions and domains (dependent

on the size/complexity of the protein) and can therefore be difficult to predict/reproduce (Grigolato and Arosio, 2019).

Pressure

The pressure dependence (as with temperature) is a well-characterised mechanism of unfolding, producing predictable results. The mechanistic understanding of the pressure unfolding process has allowed for accurate in-silico simulations (Dias and Chan, 2014; Roche et al., 2012). Under very high pressures (100~300 MPa) the voids within protein structures become unstable giving way and unfolding the protein. This process is described by the following equation Eq. 1.2.

$$\left(\frac{\delta G}{\delta p}\right)_T = \Delta V \qquad \textit{Eq. 1.2}$$

Where $\frac{\delta G}{\delta p}$ is the partial derivative of the Gibbs free energy (G) over the pressure range (P). The molar change in volume (ΔV) becomes a negative value causing the Gibbs free energy value of the unfolded state lower than that of the native one (Roche et al., 2012). The process of pressure unfolding occurs at much higher pressure than ambient and is orders of magnitude slower than temperature or chemical unfolding (Dias and Chan, 2014). Therefore, pressure is not a typical consideration for therapeutic protein formulations.

1.1.1.2. Colloidal stability

Proteins can form aggregates under certain environmental conditions. The process may involve protein-protein interactions due to conformational properties (intrinsic) or the environment surrounding the protein (extrinsic). Although the presence of some proportion of aggregates in a protein solution is inevitable, their minimisation is essential to biological activity and avoiding any harmful immunological effects. There are two

major aspects to protein aggregation: pathways of formation, and aggregate morphologies. Both the pathways and morphologies share causes and contributing factors and are thoroughly interlinked. As with unfolding, aggregation will generally occur via multiple pathways simultaneously or in succession (Wang et al., 2010).

Aggregation pathways

Protein aggregation mechanisms can be roughly categorized into three groups according to the causal effects of aggregation as follows: 1) aggregates formed due to unfolded moieties or unfolding intermediates, 2) aggregates formed due to protein self-association (hydrophobic bonds or chemical linkage), 3) aggregates formed due to chemical degradation. It is important to state that some proteins exhibit multiple aggregation pathways, with one effect dominating over others. A diagrammatic representation adapted from the original publication by Wang et al. can be seen in Figure 1.1 (Wang et al., 2010).

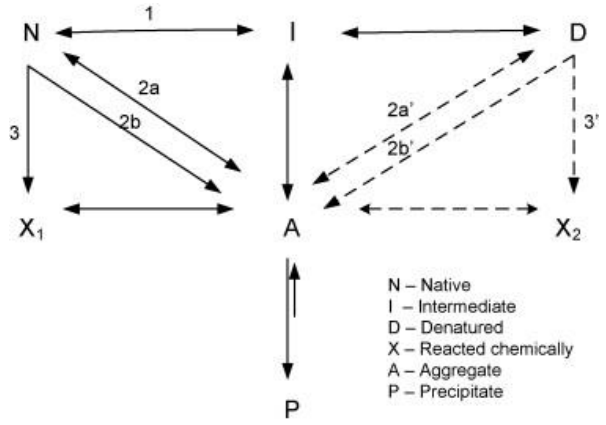


Figure 1.1 Graphic illustration of major protein aggregation pathways by Wang et al.

Protein exists in an equilibrium with relatively small amounts of unfolded and partially unfolded species in solution. As the protein unfolds the hydrophobic core of the protein is exposed and protein-protein hydrophobic binding becomes more favourable (Weiss et al., 2009). Literature suggests that the unfolded intermediates are most likely to form aggregates (Yazdanparast et al., 2007). In contrast, the native protein provides little access

to the hydrophobic chains buried within its core, while the fully unfolded protein's hydrophobic chains are too scattered to provide any bonding capabilities. The conditions for unfolding are therefore considered the rate-limiting step for the unfolded aggregation (Wang et al., 2010).

Hydrophobic aggregates formed in this way are likely to be initially soluble. However, as certain size and solubility limits are reached, solubilisation becomes less favourable, and precipitation becomes inevitable. Precipitates can form ordered (fibrils) or disordered (amorphous) structures (Modler et al., 2003).

Proteins can self-associate through electrostatic and hydrophobic bonding (Weiss et al., 2009). The association does not necessarily proceed from the intermediate unfolded state and may or may not be associated with conformational changes. The association leads to the formation of reversible aggregates, which act as precursors for the irreversible aggregates. Multiple binding sites on the protein surface can be involved in the self-association process (Kanai et al., 2008).

The surface charge of the protein often expressed as the osmotic second virial coefficient (*B*) is a key parameter in the prediction of aggregation behaviours (Alford et al., 2008). If the *B* value is positive the proteins exhibit a repulsion in solution, while a negative value indicates favourable conditions for protein-protein association. Surface charge parameters vary with the pH and ionic strength of the solution, explaining the effect of these parameters on the *B* value and aggregate formation. Often denaturants and even non-electrolytes can affect charge parameters and thus the colloidal stability of the protein (Shaw et al., 2001; Wang, 1999).

It is known that self-association is related to colloidal stability, while unfolded intermediate formation is controlled by conformational stability. Both colloidal stability

and unfolding intermediate kinetics can act as rate-limiting factors for aggregate formation. Because it is experimentally difficult to separate the effects of each pathway, the *B* value can be misleading when determining a protein solution's likelihood of aggregating. However, as mentioned previously, surface charge reduction leads to aggregation and precipitation of proteins (Kanai et al., 2008; Shaw et al., 2001).

The most common pathway for protein chemical associations is through disulfide linkages formed between cysteine (Cys) residues. Although this reaction predominantly occurs in proteins rich in surface cysteine residues, disulfide linkages can also occur in proteins lacking free cysteine residues via β-elimination. The process of β-elimination involves the cleavage of a sigma bond at a β-position leading to the formation of a new pi bond, and is demonstrated in Figure 1.2 (Cabra et al., 2008).

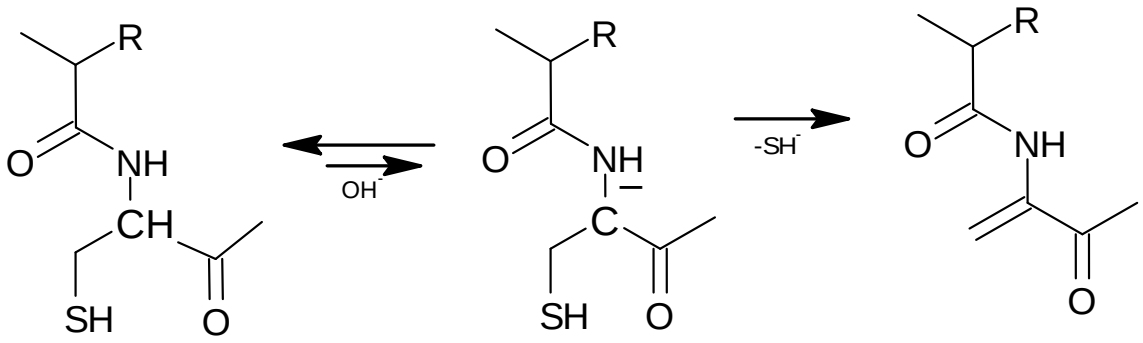


Figure 1.2 Reaction scheme of the chemically driven β-elimination for cysteine residues in proteins as described by cabra et al.

Chemical degradation pathways do not have a unique mechanism of aggregation induction, rather they induce one or more of the pathways discussed previously. Chemical degradation-induced aggregation of the protein molecule can take place via oxidation (Rosenfeld et al., 2009), dimerization (Roostaee et al., 2009), deamidation (Takata et al., 2008), hydrolysis (Van Buren et al., 2008), and glycation (Wei et al., 2009). The study of

aggregate morphology and structure coupled with in-silico simulations can help identify the particular mechanisms associated with a degradation reaction.

Aggregation morphology

Protein aggregation can result in several morphologies including fibrils (ordered with multiple sub-morphologies), particulates (irregular spheroids), and gels (including suspended structures and films). Aggregation morphologies are linked to particular pathways (Giurleo et al., 2008; Natalello et al., 2008). However, solution pH is the most significant factor in determining the type of aggregate formed (Krebs et al., 2009).

According to Krebs et al. the aggregation process is a general one, independent of the protein amino acid sequence (Krebs et al., 2007). Studies performed using bovine serum albumin (BSA), bovine insulin, bovine β-lactoglobulin, and egg white lysozyme all point to the total charge of the solution controlling the aggregate morphology. At high total net charges (which occur at pH much higher than the protein pI) the aggregates will tend towards ordered fibrillation. Conversely, solutions with a low net charge will tend towards disorganised particulates. Finally, for neutrally charged solutions (close to the pI of protein) the weak interactions will initiate gelation (Krebs et al., 2007). Gel properties will be determined by the following multi-phase growth process and are affected by temperature, pH, concentration, and the protein's tendency to form chemical linkages (through the surface availability of linkage-forming residues). Gels have been observed to be clear or turbid. Although gelation can decrease or nullify protein biological activity, they have seen some applications in the literature as drug carriers and delivery systems (Chodankar et al., 2009; Mercade-Prieto and Gunasekaran, 2009).

1.1.1.3. Chemical stability

Chemical degradation refers to the alteration or destruction of covalent bonds within a protein moiety. The chemical stability of proteins has been extensively studied in the solid state as well as in solution (Butreddy et al., 2021). The pathways share common causes and mechanisms under both conditions. The most common forms of chemical denaturation include deamidation, oxidation, β-elimination, peptide bond cleavage, Maillard reactions, and covalent dimerization/trimerization (Ahern and Manning, 1992).

Deamidation

The deamidation of asparagine (Asn) and glutamine (Gln) residues to cyclic residues occurs normally under biological conditions. Some proteins degraded in this manner are targets for destruction by the ubiquitin-proteasome system. The deamidation of proteins over time can therefore act as a sort of biological clock for proteins and facilitate an appropriate turnover. With regards to the efficacy of therapeutic protein formulations, deamidation is less favourable. Deamidation can affect in-vivo half-life, conformation, and can even lead to immunogenic reactions (Kato et al., 2020).

Asn deamidation proceeds through a succinimide intermediate to form aspartate (Asp) or iso-aspartate (Iso-Asp), Figure 1.3.

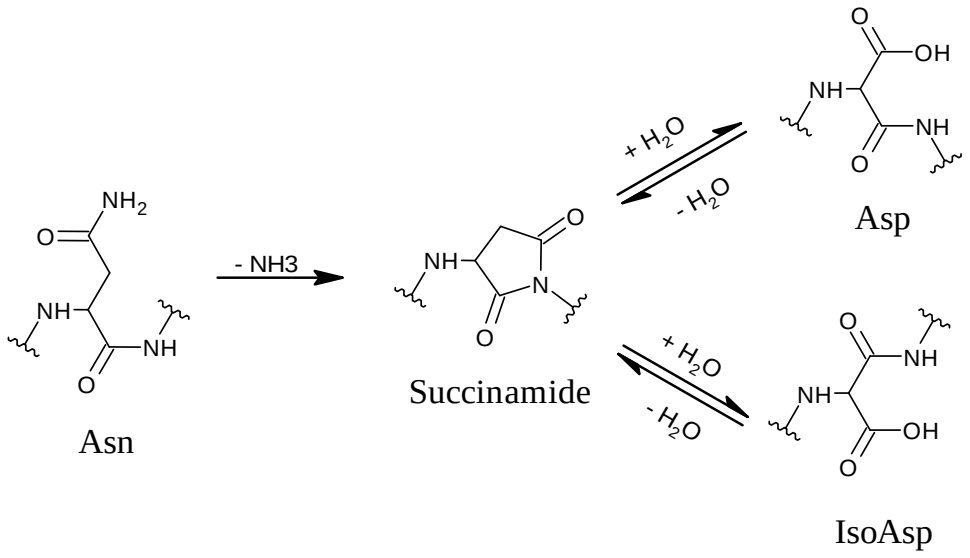


Figure 1.3 Schematic of succinimide intermediated asparagine (Asn) deamidation to aspartate (Asp)and iso-aspartate (IsoAsp).

As for Gln the reaction results in the formation of glutamate (α-Glu) and iso-glutamate (γ-Glu) the reaction is demonstrated in Figure 1.4.

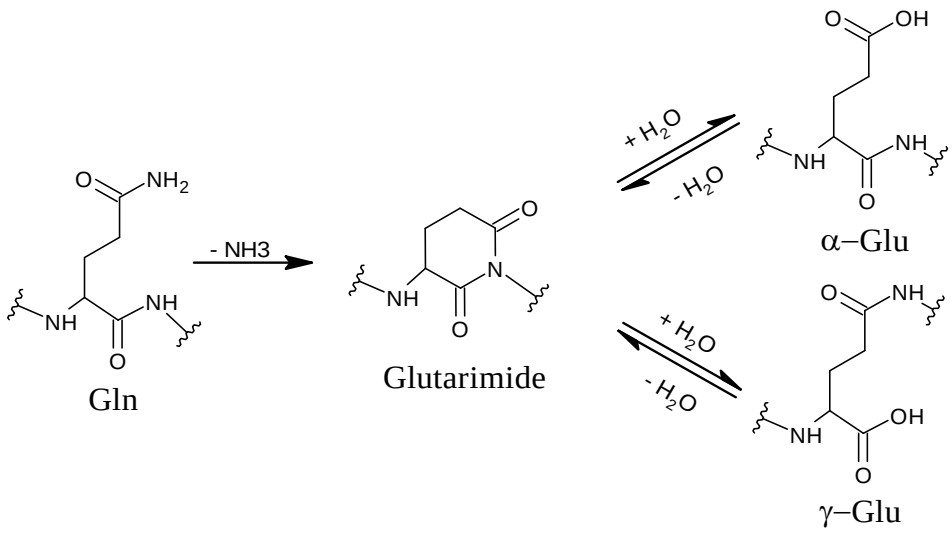


Figure 1.4 Schematic of glutarimide intermediated deamidation of glutamine (Gln) to glutamic acid (α Glu) and iso-glutamic acid (γ Glu)

The reaction is known to occur predominantly in neutral to basic media. The conformational structure plays an important role in mitigating the effects of deamidation as α-helices and β-sheets can shield the reactive moieties.

Oxidation

Oxidation of proteins takes place mainly at amino acid residues containing specific reactive groups including sulphur moieties such as cysteine, and glycoxidation and lipoxidation liable residues as arginine, and lysine. Oxidation can also take place at aromatic rings including tryptophan, tyrosine, phenylalanine and histidine. The reaction requires the presence of an oxidant (reactive oxygen species (ROS)) which can originate from oxygen (O2 •−**)**, hydroxyl (• OH), alkoxyl (RO•) or peroxyl (ROO•) radicals. Non-radical species such as hydrogen peroxide (H_2O_2), and reactive aldehydes and ketones can also act as oxidants (Butterfield and Stadtman, 1997). External stressors such as UV, and higher energy radiations (X-ray, γ-rays, and even visible light in the presence of sensitisers) can result in the initiation or propagation of the formation of ROS (Davies, 2005). Oxidants can react with the protein backbone resulting in conformational changes and fragmentation. Oxidation induces unfolding by increasing surface hydrophobicity, resulting in decreased solubility and increased intra- and inter-protein interactions (Davies et al., 1987). Some oxidation reactions result in the irreversible loss of protein activity through aggregation (Kehm et al., 2021).

Salt effects

Salts can act as protein stabilisers or destabilisers depending on the protein sequence and conformational structure. In 1888, Hoffmeister developed the Hoffmeister series to evaluate the effectiveness of salt ions on protein solubility and classify them accordingly. Since then the literature has reproduced the same series for several effects including denaturation, depolymerisation, dissociation, and activation or suppression of enzymatic activity (Hofmeister, 1888).

The solubility of proteins is influenced by the interactions between the protein and the solvent in the presence of salts. In cases of preferential exclusion of the salts from the monolayer surrounding proteins precipitation becomes more favourable. Whereas salts that exhibit weak binding tendencies with the protein increase solubility.

The mechanism for precipitation involves an increase in water surface tension, to minimise surface-protein interaction the protein takes on a more compact structure. The mechanisms for salting-in are more complex, however, the main effects are understood to be electrostatic in nature. A publication by Neagu et al. explored the salting-in effect, proposing that salts lead to stochastic amplitude fluctuations of the free energy profiles of the protein, proportional to the salt concentration. Neagu et al. describe two types of salt moieties kosmotropes and chaotropes. Kosmotropes tighten inter- and intramolecular interactions and favour aggregation but prevent denaturation, while the opposite is true for chaotropes. Exceptions to this rule are known as some kosmotropes lead to aggregation and destabilisation of conformation due to specific interactions with the protein backbone (Neagu et al., 2001).

Adsorption

The amphiphilic nature of proteins causes them to adsorb to various surfaces, potentially leading to aggregation and denaturation (Andrade, 1985; Burke et al., 1992). Adsorption is of particular concern for low-concentration formulations, as relatively more of the therapeutic protein is eliminated to reach surface saturation. Surface adsorption is also an issue for the transport of protein during manufacture, and transfer from one vessel to another. To control surface adsorption an inert protein like serum albumin may be used to saturate the container surface. Chemical agents such as surfactants, carbohydrates, or

amino acids can also be used to minimise surface interactions (Burke et al., 1992; Fainerman and Miller, 2005; Lu et al., 1999).

Protein instability typically presents in multiple forms simultaneously, often with one factor compounding the others. Different strategies can be used to maintain the efficacy of protein therapeutics, including freezing throughout transport and storage (cold chain), freeze-drying (lyophilisation), spray drying, and the addition of stabilising excipients as buffers and surfactants.

1.1.2. Lyophilisation & reconstitution

1.1.2.1. Lyophilisation

Exposure of therapeutic proteins to a solution environment for long periods can lead to disadvantageous processes like aggregation, or unfolding (Edsall and McKenzie, 1983; Privalov and Gill, 1988). One method of overcoming protein instability is the process of lyophilisation, or freeze-drying of the protein formulation; which, is capable of stabilizing the protein formulation (Allison et al., 1998) . This process involves freezing an aqueous solution and reheating it below the triple point of water, under a vacuum, causing water to sublimate directly into vapour. Thus, removing most of the water content in the formulation while avoiding any heat damage to the dried material (Snowman, 1997). Lyophilisation reduces the water content of the material to below 3% to impart stability to the protein (Chansanroj and Thanawattanawanich, 2015; Deutscher, 1990; May, 2004). However, some protein degradation via oxidation, deamination, or aggregation can still occur in the solid state, post-lyophilisation (Carpenter et al., 1997; Zhang et al., 2010)

In addition, the lyophilisation process carries risks of instability of its own. This is caused by multiple mechanisms. Firstly, pH shifts caused by buffer crystallization during

freezing. Also, freeze (cryo-)concentration can occur, which is the formation of local areas of high concentration due to non-homogenous freezing of the bulk solution, leading to the formation of areas of higher concentration within the cake during the freezing drying process (Akyurt et al., 2002), causing increased intermolecular collisions between protein molecules which in turn increases the probability of protein-protein interactions, as well as, protein-interface interactions (Chi et al., 2003; Jones et al., 1997; Randolph and Jones, 2002). Finally, due to the interactions of the protein at the liquid interface, during the crystallization of water and the subsequent removal of water molecules from the solution, the protein protective monolayer of water molecules can disperse partially or completely leading to destabilization of the protein tertiary structure and therefore unfolding reactions can proceed (Jones et al., 1997). All of the previously mentioned processes can lead to protein unfolding and aggregation (Manning et al., 2010). Therefore, protein formulations need to be designed to eliminate or minimise these instabilities during freeze drying, storage and transport; by utilizing buffers, cryoprotectants, and lyoprotectants respectively (V Gervasi et al., 2018). In summary, lyophilisation of proteins is a complex process influenced by many variables affecting the quality of the final product.

1.1.2.2. Reconstitution

Lyophilised protein formulations are porous solid cakes, with low moisture content (below 3%) (May, 2004), that require solvation with a specific solvent (traditionally, water for injection (WFI)) before it can be administered to the patient. The process is termed “reconstitution”, and is a combination of multiple physical processes occurring simultaneously in the form of wetting, disintegration, and finally dissolution; therefore,

understanding the factors controlling each of these processes is integral in the design of lyophilisation cycles that lead to shorter and more convenient reconstitution times.

For the convenience of the administrator and patient, the reconstitution step should be designed to be as short as possible, especially in a situation where a long reconstitution could lead to complications, e.g., in the case of some anti-venoms (Hill et al., 2001). There has been much progress with regard to techniques used to reduce reconstitution time without any sacrifices to stability, through the optimisation of cake properties including protein density, cake morphology, internal surface properties, and other factors relating to the permeation of the reconstitution fluid through the cake (Cao et al., 2013; Geidobler et al., 2013; Kulkarni et al., 2018; Wang, 2000).

It was shown that formulations containing less than 50 mg/ml of protein are capable of reconstituting reliably and rapidly without the need for reconstitution enhancement techniques or components (Kulkarni et al., 2018). However, at higher protein concentrations the reconstitution time increases significantly (Kulkarni et al., 2020). Wherein reconstitution times could range from ~10 minutes (Cao et al., 2013)up to one or even several hours (Bhambhani and Blue, 2010; Blue and Yoder, 2009; Tchessalov et al., 2010). A look at protein formulations currently in use in the EU reveals that while the traditional IV route is still a popular approach for the administration of protein formulations, there is a trending increase in the use of subcutaneous administration of Mabs (Bhambhani and Blue, 2010; Cui et al., 2017). This is very likely due to the ease of self-administration of subcutaneous products, eliminating the need for administration in a clinical environment thus increasing patient compliance. This necessitates low administration volumes (<1.5ml) and higher concentrations (>100mg/ml) (Bhambhani and Blue, 2010; V Gervasi et al., 2018; Shire et al., 2004). The long reconstitution times

of high-concentration formulations encourage more vigorous shaking or stirring of the vial which is likely to cause aggregation of the protein (Kulkarni et al., 2020).

At present the majority of the literature represents or includes data on lyophilised products and their reconstitution taking a stability-first approach (Butreddy et al., 2021; Preston and Randolph, 2021; Thakral et al., 2021; Wahl et al., 2016). Such analyses would have been sufficient for low-concentration lyophilised proteins with fast reconstitution times. However, with recent trends in the biopharmaceutical markets tending towards higher concentration formulations for subcutaneous administration (V Gervasi et al., 2018), reconstitution expediting strategies have become increasingly relevant.

The effects of the lyophilisation cycle variations on the final dried product have been examined extensively in the literature with regard to their effect on stability and reconstitution (Sane et al., 2020; Snowman, 1997; Zimmermann et al., 2000). By studying the factors affecting both the stability of a lyophilised formulation and its reconstitution time it could be said that the relationship between is that of competing demands with an overt focus on either side leading to compromising the other. This means that an optimal formulation would have an intermediate reconstitution time while maintaining acceptable stability qualities.

Earlier studies on the reconstitution time of lyophilised protein cakes by Shire et al. showed that at a constant mass, lower concentrations (of higher volumes) showed a marked improvement in reconstitution time over higher concentration (lower volume) solutions. At the time this was attributed to the more loosely packed nature (detected by scanning electron microscopy (SEM)) of the lower concentration solution over the denser, higher concentration solutions (Shire et al., 2004). However, further investigations by Cao et al. concluded that morphology may not be the most significant factor affecting

reconstitution time. Furthermore, due to the unquantifiable nature of SEM results, it was concluded that the degree of crystallinity of protein cakes had a more pronounced and measurable effect on the reconstitution times (Cao et al., 2013).

With regards to lyophilised formulations, it was thought that the cake would ideally be completely amorphous, thus minimising the risk of crystallisation upon storage, and allowing the formulation to maintain shelf stability for a longer time (Beech et al., 2015). In some cases, the reconstitution time for fully amorphous cakes is very long due to the low porosity (Beech et al., 2015; Colandene et al., 2007) and high resistance of the cakes (Johnson et al., 2010; Overcashier et al., 1999) in turn, due to the high protein concentration of the formulations.

Studies have shown that certain lyo-cycle conditions and formulation components can promote more controlled crystallisations during the lyo-cycle itself, producing formulations that are still relatively shelf-stable (Overcashier et al., 1999).The improvement in reconstitution time would come at the cost of an increased likelihood of drug crystallisation during storage.

Thus, by controlling all the aforementioned factors we can control the reconstitution properties of the Lyophilised cake. However, to obtain the best possible product for clinical use, conservation of product stability and cake appearance acceptability must also be taken into consideration (Al-hussein and Gieseler, 2013; Patel et al., 2017). Therefore, a balance between the reconstitution time-reducing methods and traditional lyophilisation methods for obtaining acceptable cake appearance and physical stability must be maintained.

1.2. Protein spectroscopy

Spectroscopic analytical techniques are based on the probing of a material using photons, followed by the analysis of some aspect of the resultant photons to obtain chemical and physical information. A photon is a specific quantum (unit) of electromagnetic radiation and is a carrier of energy. The interaction of photons with the probed material can take many forms and can be divided according to the proportion of light received by a detector as a function of the original, incident light (interaction vs. no interaction) (Saleh and Teich, 1991).

Transmittance refers to the proportion of light that passes through a material without any interaction. Reflectance describes how light beams follow the law of reflection, where incident photons reflect at equal angles without losing energy (Hanrahan and Krueger, 1993). Absorbance refers to a photon's absorption by the material, which requires photons to carry an energy that is similar to a certain molecular transition. The energy of a photon is inversely proportional to its wavelength, and as such different molecular transitions will occur depending on a photon’s position along the electromagnetic spectrum(Lakowicz, 1999). Examples of two such possible excitations include electron excitation (Lakowicz, 1999) and vibrational transitions (Nafie, 2008, 2001) is shown in the Jablonski diagram in Figure 1.5.

After an excited state is induced by absorption, the return of an electron to the ground state leads to fluorescence. The emitted photon that causes the fluorescence carries energy that is equal to the difference in energy between the ground and excited states (Eftink, 1991). However, some energy is lost through non-radiative transitions, such as molecular relaxation, resulting in the emitted photon carrying less energy than the incident photon (Eftink, 1991; Lakowicz, 1999).

Scattering refers to an event where photons interact with matter and are emitted in different directions because the incident (absorbed) photon's energy is not resonant with a molecular transition and is instead excited to a virtual state (Clarke and Oprysa, 2004). The relaxation of the excited state leads to the emission of a photon. When the emitted photon carries energy identical to the incident the process is termed elastic scattering, while a change in energy leads to inelastic scattering (Clarke and Oprysa, 2004; Lewis and Edwards, 2001). The Jablonski diagram in Figure 1.5 illustrates both effects, with elastic scattering shown by I, and inelastic scattering shown by II and III (Klijn and Hubbuch, 2021).

Inelastic scattering results in the emission of a photon with a different energy than the incident photon. At lower energies, the scattering is referred to as Raman Stokes (II, Figure 1.5) scattering. Higher energy scattering is referred to as anti-Stokes (III, Figure 1.5). Inelastic Raman scattering accounts for only about 1 in 10^{10} incident photons (Lewis and Edwards, 2001; McCreery, 2005; Wokaun, 1996).

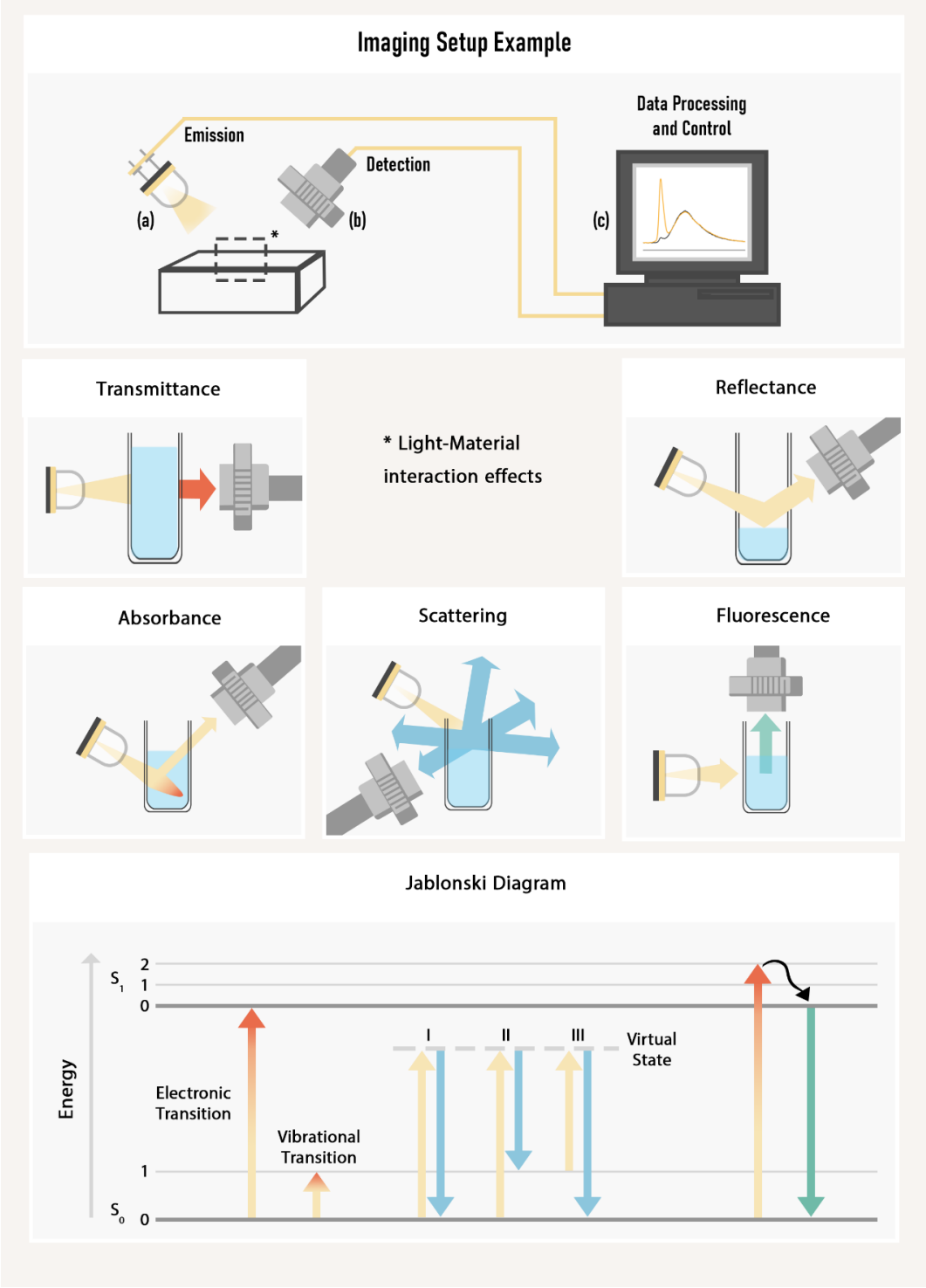


Figure 1.5 Diagrammatic representation of different light interactions during spectroscopic measurements. The imaging setup includes a light source (a), a detector (b), control units and connectors (c). The interactions shown include transmittance (red), reflectance (yellow), scattering (blue), and fluorescence (green) of an incoming light beam. The Jablonski diagram illustrates the transitions pertaining to absorbance, scattering (Rayleigh scattering (I), Raman Stokes scattering (II), and Raman anti-Stokes scattering (III)), and fluorescence.

Analytical spectrophotometric techniques relevant to biomolecular analysis and utilising the previously stated material interactions are discussed in further detail in the following sections.

1.2.1. Protein quantification

Protein quantification is an essential aspect of therapeutic formulation development and manufacture, for example determination of yield during purification, as well as an indirect measure of formulation specific activity. The most common spectrophotometric techniques used for protein quantitation include UV absorbance, and dye-binding assays based on colourimetry and fluorescence (Spalding et al., 2018). Spectrophotometric techniques are the standard in biochemical laboratories. To measure concentration a reference standard or amino acid sequence is required. The choice of assay depends on the buffer matrix in which the protein is dissolved, as well as a host of other factors including sample volume, sample availability and cost, required throughput rate, robustness, existent chemical modification (glycosylation/PEGylation), and aggregation tendencies.

1.2.1.1. Intrinsic UV absorbance

UV-absorbance at 280 nm can be used to quantify proteins carrying UV-absorbing residues including tryptophan, tyrosine, and phenylalanine (Noble and Bailey, 2009). This method is highly selective and is compatible with many excipients commonly found in protein formulations including acids/bases, inorganic buffers, amino-acid buffers, sodium chloride. Relative UV-absorbance can be used to give an estimation of protein concentration, however, if the molar extinction coefficient is known the concentration can be accurately determined using the Beer-Lambert law, Eq. 1.3 (Swinehart, 1962).

$$A = a_m bc \qquad \textit{Eq. 1.3}$$

Where a_m is the molar extinction coefficient of the protein moiety in $Mol^{-1}.cm^{-1}$ and is a summation of all the photon-absorbing species at a defined wavelength. The molar extinction coefficient can be determined experimentally or calculated from first principles from the protein's amino acid sequence (Edelhoch, 1967; Gill and Von Hippel, 1989). Amino acid absorbance is sensitive to the chemical microenvironment influenced by buffers, excipients, and denaturation. *b* is the path length in cm, and *c* is the molar concentration. If a_m is unknown, the analyte concentration can be interpolated from a standard series of dilutions. UV quantitation of proteins at 280 nm is sensitive over a wide range of concentrations, typically from 20 µg/ml to 3000 µg/ml (Noble and Bailey, 2009). Samples can be "blanked" to ensure the absence of interference, however, issues may arise when including some detergents like Triton-X as they are liable to absorb strongly (Thermo Scientific, 2012). The use of quartz crystal allows the full transmission of the light spectrum as well as protecting the solution from any sources of leaching (observed for some plastics). Any turbidity in the sample or aggregation of the protein close to the size of the incident wavelength (250~300 nm) can lead to increased scattering, a decrease in the light reaching the detector and thus an increase in apparent absorption. In such cases, sample filtration, and light scattering correction by baselining to absorbance at lower energy wavelengths can help reduce scattering effects (Leach and Scheraga, 1960). Absorption at 205 nm is useful for low-concentration protein quantitation. Peptide bonds mainly absorb below 210 nm. However, due to the broad absorption peak, longer wavelengths as 214 nm at 220 nm can be used in the presence of interfering buffers and excipients. As the absorbance at 205 nm is more intense than 280 nm, much lower concentrations can be quantified (1-100 µg/ml) (Scopes, 1974).

1.2.1.2. Intrinsic fluorescence

Similar to UV absorbance, amino acids with aromatic side groups are responsible for the intrinsic fluorescence of proteins. The quantitation is conducted similarly to UV using standard curves of known concentration of the queried protein or a protein of similar primary and secondary structure. The parameters for aromatic amino acid fluorescence can be seen in Table 1.1. Quantum yield signifies the level of fluorescence efficiency, which is the ratio of emitted photons to absorbed photons. Phenylalanine's fluorescence is unreliable due to its low quantum yield. Tyrosine's fluorescence could appear to be comparable to tryptophan's, but the signal may be affected by quenching and interference from the protonation on aliphatic side chains. Therefore, tryptophan emissions are the most commonly studied for protein quantification via intrinsic fluorescence (Hawkins and Honigs, 1987).

Table 1.1 Fluorescent properties of aromatic amino acids (Hawkins and Honigs, 1987)

Amino acid	*Excitation wavelength*	*Emission wavelength*	*Quantum yield*
Phenylalanine	260 nm	283 nm	0.04
Tryptophan	285 nm	360 nm	0.20
Tyrosine	275 nm	310 nm	0.21

Although intrinsic fluorescence is most commonly used for qualitative measurements, using a protein standard of similar enough primary structure to the queried protein allows quantitative measurements. As proteins with similar number of fluorophores and of similar conformational structures will produce comparable fluorescence, in some cases a more available protein can be used as a standard for the quantification of a rarer variety. However, the tertiary structure of the biomolecule must be taken into consideration as

quenching can occur due to alterations in conformation by a host of factors e.g., adjacent protonated acidic groups quenching tryptophan fluorescence (Freifelder, 1982).

1.2.1.3. Factors affecting both UV & fluorescence.

The polarity of the solvent is known to affect the spectrophotometric properties of proteins due to the interactions with the absorbing amino acid residues. For example, a decrease in solvent polarity results in a decrease in fluorescence lambda max and an increase in intensity for tryptophan (Varadhan, 1990). Buffers and other excipients can also interfere with absorbance at 280 nm and 205 nm. Common excipients such as ammonium sulphate, ethylenediaminetetraacetic acid, glycerol, 2-mercaptoethanol, sucrose, phosphate and tris buffers, urea, and SDS can absorb at 205 nm due to the presence of carbon–hydrogen bonds or carbon–carbon bonds. Some excipients may absorb in the region of 280 nm and thus can add to the overall sample absorption in this region and can interfere with protein quantification.

Nucleic acids have significant absorbance at 280 nm and can interfere with protein quantitation. To resolve the protein concentration absorbances at 260 nm and 280 nm are measured and used to calculate the pure protein concentration via the formula shown in Eq. 1.4 (Warburg et al., 1941).

$$\text{protein concentration (mg/ml)} = 1.55 \times A_{280} - 0.76 \times A_{260} \quad \textit{Eq. 1.4}$$

1.2.1.4. Colorimetric assays

Although not highly relevant to the results presented in this thesis, it is important to mention in brief the colourimetry tools available for protein quantitation as they offer easy-to-use, efficient, and highly reproducible methods to analyse proteins. The most

frequently employed techniques include the Bradford and the Lowry assays (Bradford, 1976; Lowry et al., 1951). The assays involve protein binding to or reacting with a dye molecule leading to a detectable and quantifiable colour change in the solution. Reactions can occur at low concentrations of protein (1-20 µg/ml) and can be performed rapidly using 96-well plate readers (Simonian and Smith, 2006).

However, the Lowry reactions depend on the similarity of the assayed protein to the standard protein (BSA) and the presence of a certain number of tyrosines/µg. Lower or higher concentrations of tyrosine can lead to over or under-estimation by the assay as it relies on the reaction of these residues with the reagent. The assays also suffer from the interference of excipients within the solution and their reactions with the assay reagents can be difficult to account for or resolve (Noble and Bailey, 2009; Simonian and Smith, 2006).

1.2.2. Conformational monitoring

To measure conformational differences of protein molecules using a spectroscopic method, there must be a measurable signal difference for the different states that occur during unfolding. As previously discussed in the section 1.1.1.1, protein conformation is highly sensitive to changes in the environment surrounding the protein. Some spectroscopic techniques, such as absorbance and fluorescence, rely on the interaction of a fluorophore with a photon with a specific quantum of energy to reach an excited state. Changes in the fluorophore's electromagnetic nano-environment through unfolding or solvent interactions will lead to a measurable difference in the energy required to reach an excited state (Michalet et al., 2006), and thus the wavelength of light absorbed and emitted (Anand et al., 2011; Eftink, 1994). Other spectroscopic techniques may rely on

molecular volume (dynamic light scattering (DLS)), optical transitions (circular dichroism (CD)) (Eftink, 1994), kinetic motion in the solution (anisotropy), or bond vibrational energies (Fourier transform infra-red (FTIR))(Spalding et al., 2018).

1.2.2.1. Fluorescence

UV fluorescence analysis is a well-established method for the identification and characterisation of many chemical and pharmaceutical substances. The fluorescence signals that can be monitored include fluorescence intensity ($F_{ex\lambda em\lambda}$) for wavelengths of excitation (ex_{λ}) and emission ($em_{\lambda)}$, quantum yield (Φ), emission maximum (λ_{max}), fluorescence lifetime (τ_i), pre-exponential factor (α_i), fluorescence anisotropy (r), and rotational correlation time (φ). Researchers have used these signals to monitor protein unfolding and the technique has recently become part of routine protein lab techniques (Eftink, 1994, 1991; Gryczynski et al., 1988; Mei et al., 1992).

Fluorescence lambda max shift

Protein conformation can be monitored by measuring the shift in the maximum emission wavelength (λ_{max}). This shift is called a red shift or a blue shift and occurs when a protein containing a fluorophore undergoes a conformational change (Lakowicz, 1999; Meech, 2009). Protein fluorophores are involved in intermolecular interactions such as hydrogen bonding and hydrophobic and electrostatic interactions. Even a minor change in the protein structure can influence the fluorophore's nano-environment, leading to a shift in the λ_{max}. If the shift is towards shorter wavelengths (blue shift), it suggests that the fluorophore has experienced a more hydrophobic nano-environment and may be involved in stronger affinity interactions compared to the protein's previous structure. The inverse

occurs, where a more polar nano-environment will lead to shifting towards a longer wavelength (red shift) (Hellmann and Schneider, 2019).

Fluorescence quenching

Fluorescence intensity at the λ_{max} can be monitored to detect conformational changes as well as fluorophore interactions with quenching/denaturing molecules. When a quencher moiety is added to a protein solution it is common to see both a red/blue shift and a reduction in fluorescence intensity (quenching). As denaturant concentration increases, the protein structure becomes more compact due to the electrostatic interactions. The interaction leads to a change in the environment surrounding tryptophan moieties within the protein's hydrophobic core, thus leading to an increase in fluorescence intensity (Anand et al., 2011). At higher denaturant concentrations, the protein begins to completely unfold, exposing the fluorescent tryptophan residues to the polar environment of the solution, leading to a decrease in fluorescence intensity (Eftink, 1994).

Fluorescence anisotropy

Fluorescence anisotropy *(r)* capitalises on the change in the kinetic rotation (tumbling) time of a fluorescent protein when a ligand or protein partner binds to it (Gijsbers et al., 2016; Rossi and Taylor, 2011; Sijbesma et al., 2020). The method involves using polarisers that are oriented perpendicularly to each other. By exciting the fluorescent protein with isotropic vertically polarised light, the fluorescent protein emits light (also polarized). The emitted light is then passed through a horizontal polariser before detection. The process is repeated using vertical polarisers for both excitation and emission. The difference between the two signals represents the globularity or symmetry of the protein structure.

The fluorescent protein is free to rotate in solution, and the angle of the emitted polarised light changes as it rotates. When a ligand or a partner bind to the protein, its rotation speed slows down (Rossi and Taylor, 2011). Unfolding of the protein can also change its volume or structure, leading to changes in its rotation speed (Chullipalliyalil et al., 2023).

Since the exciter and detector are set up perpendicularly, rapid rotation of the protein leads to less light reaching the detector at any one angle. Less rapid rotation will lead to a single angle being favoured and thus an increase in its intensity. A more detailed explanation and example is provided in Chapter 2.

Fluorescence lifetime

Time-resolved fluorescence spectroscopy is a technique that can be used to monitor interactions and motions that occur over brief time intervals, ranging from picoseconds to nanoseconds (Millar, 1996). There are two main types of time-resolved fluorescence spectroscopy that are commonly used to investigate proteins. The first type involves measuring the decay of the total fluorescence intensity to determine the fluorescence lifetime (τ_i) of a specific fluorophore and its surrounding nano-environment (Beechem and Brand, 1985).

In protein-protein interaction assays, time-resolved fluorescence energy transfer can be used, which combines time-resolved fluorimetry with Förster resonance energy transfer (FRET), to characterize the modes of action, efficacy, and binding affinities of ligands (Harkes et al., 2021; Margineanu et al., 2016). The time-resolved FRET method can quantify concentrations of protein species by characterizing their fluorescence decay signature. The method involves a series of measurements with fluorescence intensity in the range of 250-280 nm, taken with time differences of at least ten nanoseconds, allowing the construction of a fluorescence decay curve (Hales et al., 2021). The overall decay

profile of a mixed sample is the sum of the decay of the components. Thus, knowledge of the characteristic decay curve of a protein can be used to deconvolute the total fluorescence decay curve (Hales et al., 2021).

The second type of lifetime spectroscopy is based on the measurement of the decay of polarization anisotropy, which can be used to characterize molecular movements such as rotational diffusion of proteins or nucleic acids. The picosecond to nanosecond time scale enables the study of reaction mechanisms in real time (Millar, 1996).

Protein fluorescence decay lifetime is a non-destructive method that can be used to study protein folding, dynamics, assembly, and interactions over time. As the fluorescence lasts for nanoseconds, it is faster than most protein conformational transitions, which allows researchers to investigate fast protein conformational changes with great precision.

Non-Intrinsic fluorescence binding assays

Differential scanning fluorimetry (DSF) uses temperature transitions to probe the unfolding interactions of a protein. As discussed previously temperature is a highly effective and reproducible method for unfolding proteins. Using a PCR thermal cycler instrument the proteins are incubated in a well plate with a hydrophobic dye, typically SYPRO Orange (Huynh and Partch, 2015). The temperature is ramped in a highly controlled manner. As the protein unfolds, the hydrophobic core becomes exposed and can bind and activate the fluorescent dye.

One advantage of the DSF technique is the ability to perform high throughput analysis and enables the rapid and concurrent study of multiple experimental conditions, for example, a study by Rodrigues et al. used DSF to probe the effects of an array of ionic liquids on a protein (Rodrigues et al., 2011).

1.2.2.2. Absorbance

Circular dichroism

Dichroism is a property found in certain materials, which results in the absorption of light to different degrees based on the polarisation of the incident beam. When circularly polarised light is absorbed, an additional absorption component known as circular dichroism arises. This component is measured by switching between left and right circularly polarised light and measuring the resulting difference in absorbance (Hammes and Hammes, 2005; Snatzke, 2000).

Circular dichroism can be used to assess conformational changes in the structure of macromolecules. The characteristic absorption spectra of native proteins have subtle changes that are unique to each protein. The shape of the spectrum curve, as well as the positive and negative maxima, provide valuable information about the protein folding behaviour (Whitmore and Wallace, 2008). For instance, peaks in the 200-250 nm wavelength range typically have a "w"-shaped spectrum with troughs around 222 and 208 nm, indicating the presence of α-helical structures. Similarly, peaks in the 217-220 nm wavelength range usually have a "v"-shaped spectrum indicating the presence of β-sheet structures (Kelly et al., 2005).

The peptide bond is the primary cause of absorption in the region of 240 nm and below (Snatzke, 2000). There is a weak but broad $n \rightarrow \pi^*$ transition that occurs around 220 nm, and a more intense $\pi \rightarrow \pi^*$ transition that occurs around 190 nm (Hammes and Hammes, 2005; Snatzke, 2000). In some cases, the presence of aromatic amino acid side chains can also contribute significantly to this spectral range. The various regular secondary structures in proteins create CD spectra with characteristic features in the far UV region. A graphical representation of this can be seen in Figure 1.6 (Kelly et al., 2005)

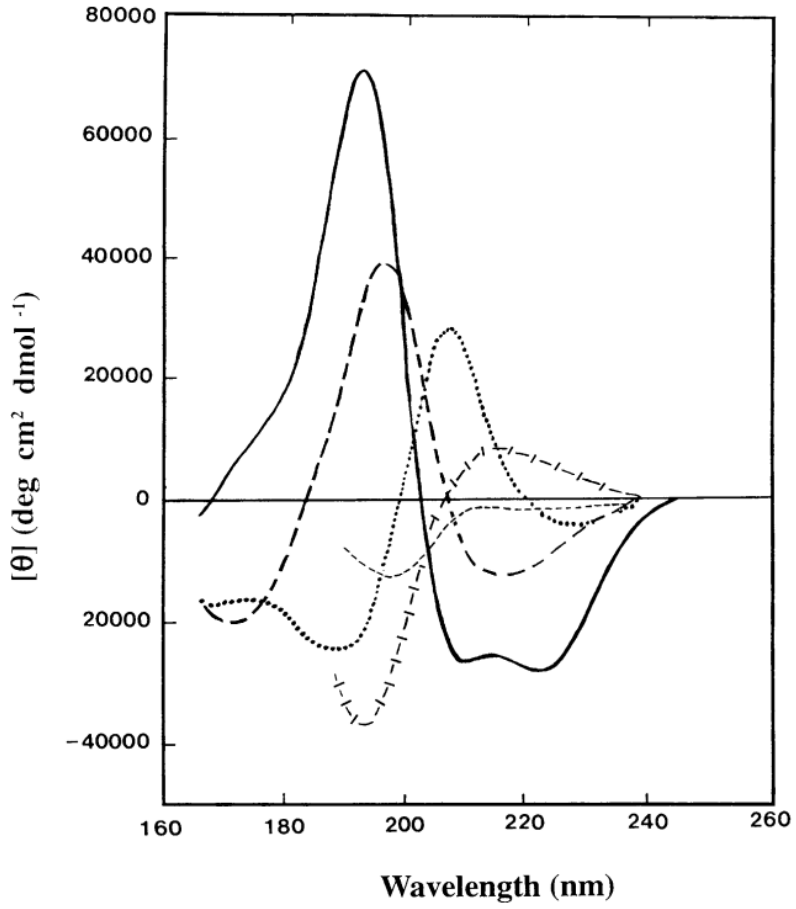


Figure 1.6 Examples of far-UV CD spectra associated with several types of secondary structure. Solid line, α-helix; long dashed line, antiparallel β-sheet; dotted line, type I β-turn; cross-dashed line, extended 3_1-helix or poly (Pro) II helix; short dashed line, irregular structure. The data are adapted from Kelly et al.(Kelly et al., 2005)

Processing CD spectra using certain algorithms can elucidate the structure and quantify the secondary structure components (Lobley et al., 2002; Miles et al., 2022). While the detailed workings of the algorithms are beyond the scope of this work, there are a few notable examples that are worth mentioning. These include the SELCON (self-consistent) algorithm (Sreerama and Woody, 1993), VARSLC (variable selection) (Manavalan and Johnson, 1987), CDSSTR (Johnson, 1999), and K2d (Moran, 2001). An online server called DICHROWEB has been developed and is hosted at Birkbeck College, University of London, UK (Lobley et al., 2002; Miles et al., 2022). This tool provides the option to

analyse the data by various algorithms with a choice of databases, including the neural networking method (CDNN) used throughout this thesis. The algorithms can estimate secondary structure through CD spectra based on a database of standard structures solved via X-ray crystallography(Whitmore and Wallace, 2008).

CD analysis requires a firm understanding of protein secondary structures, as databases with no matching proteins can lead to inaccurate results., The reported accuracies for CD are 97% for helices, 75% for β-sheet, 50% for turns, and 89% for other secondary structures (Manavalan and Johnson, 1987). The amount of material required can also be a limitation for more expensive materials as the analysis requires 20- 200 µl of a solution containing 1 to 50 µg/ml of the protein (Kelly et al., 2005). Excipients with absorbance in the far UV region can also interfere with the protein signal, therefore it is important to baseline buffer solutions prior to data analysis. Overall, CD spectroscopy provides a powerful platform for analysing protein secondary structures.

Second derivative UV

As previously mentioned, proteins can absorb at 280 nm due to an abundance of the fluorophores Phe, Tyr, and Trp. The absorbance spectra are composed of several broad peaks between 200 nm – 300 nm, the weakest of which do not overlap with the peptide bond absorbance. The fluorophores' weak absorbance bands can be attributed to the $\pi \rightarrow \pi^*$ transitions exhibited by the electrons of the aromatic rings. Derivatization significantly improves the resolution of protein signals in absorbance spectra, allowing for clearer separation from noise and other signals (Giese and French, 1955; Grum et al., 1972). Figure 1.7 illustrates how derivatization affects a Gaussian absorbance band, resulting in a more complex derivative spectrum.

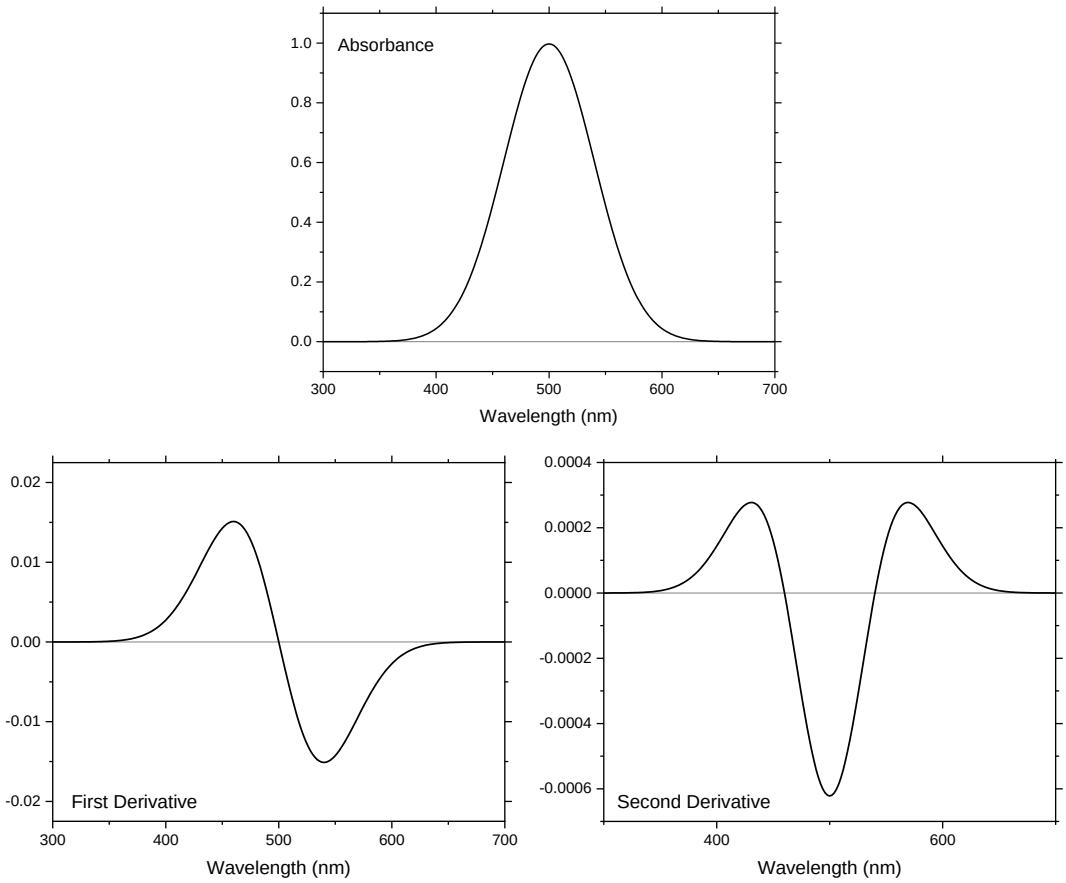


Figure 1.7 Example absorbance Gaussian band with first and second derivatives

Advancements in high-resolution UV absorbance spectroscopy have made it possible to precisely locate derivative peaks with a resolution of about 0.01 nm (Mach et al., 1991). This means that even small shifts in peak positions can be detected and analysed. The analysis of second-derivative UV spectra has been used to identify conformational changes in terms of alterations in aromatic side chain microenvironments (Balestrieri et al., 1978; Ichikawa and Terada, 1979, 1977; Levine and Marcia Federici, 1982). More recently, it has been used to create empirical phase diagrams to study protein stability under different solution conditions (Kueltzo et al., 2003).

1.2.3. Aggregation and oligomerization

1.2.3.1. Dynamic light scattering and Fluorescence Correlation

As the name suggests, dynamic light scattering (DLS), is a sizing technique based on the principles of Brownian motion. DLS instruments use a light source to illuminate a sample to measure fluctuations in the light scattering patterns. The velocities of particles under Brownian motion in solution are inversely proportional to their size. By analysing the patterns in the light scattering fluctuations, it is possible to determine the size distributions of particles in the sample. Obtaining the diffusion coefficient is possible via least-squares fitting of particle velocities. The number of observed particles can be calculated from the inverse of the autocorrelation function at $\tau = 0$.

Fluorescence correlation spectroscopy uses similar principles to DLS but relies on fluctuations in fluorescence signals detected by confocal microscopy instead. For FCS analysis, proteins require fluorescent labelling using a highly efficient dye with low photo-bleaching at working wavelengths (Mittag et al., 2018).

Protein aggregation occurs via several intermediates, nucleation sites and Brownian kinetics. As such, aggregates formed in solution will tend to follow natural distributions rather than discrete quantities. Therefore, it is more advantageous to use distribution fittings of the autocorrelation function instead of a discrete fit for a number of diffusion species (Mahler and Jiskoot, 2012).

1.2.3.2. Opalescence measurements

Turbidity is an optical kinetic method that helps to identify the presence of suspended particles in a solution. It is particularly useful in determining the formation of oligomers and aggregates. Aggregates scatter light, leading to a decrease in transmitted light and an

increase in absorbance or turbidity (Zhao et al., 2016). A particle size of at least 1/50th to 1/20th that of the incident light wavelength produces turbidity (Mahler and Jiskoot, 2012). The effect is usually measured at wavelengths > 300 nm to avoid the absorbance effects by aromatic amino acids at lower wavelengths. Turbidity (τ) which is defined as the total amount of light scattered in all directions by the sample, is calculated per the Eq. 1.5 (Berns, 2019).

$$Turbidity\ (\tau)\ =\ -\ln(\frac{I}{I_o})\ =\ -2.3\ OD \qquad \textit{Eq. 1.5}$$

Where OD the optical density is described by Eq. 1.6:

$$OD\ =\ A\ +\ S \qquad \textit{Eq. 1.6}$$

With *A* being the absorbance and *S* the total light scattered and is considered the indicator for aggregation. Thus, while measuring absorbance, it is important to correct for scattering effects using an exponential decay curve (Pignataro et al., 2020) as previously discussed in section 1.2.1.1.

A more practical approach to scattering measurements is the aggregation index (Ai), which provides an approximation of the aggregation state of the sample. The aggregation index is given by the following equation (Eq. 1.7),

$$\mathrm{Ai}\ =\ \frac{\mathrm{A}_{350}}{(\mathrm{A}_{280}\ -\ \mathrm{A}_{350})}\cdot 100 \qquad \textit{Eq. 1.7}$$

Where A_{350} and A_{280} are the absorbance values at 350 nm and 280 nm, respectively. An *Ai* value above 30 indicates heavy aggregation, between 30 – 3 is considered a lightly aggregated protein solution and below 3 is considered a clear sample (Gupta and Mahalakshmi, 2020, 2019; Herrera et al., 2016).

Turbidity measurement does not require protein labelling and is, therefore, advantageous. However, it has limitations as it cannot distinguish between the distinct types of

aggregates or detect oligomeric intermediates. Consequently, it should be used in conjunction with additional biophysical techniques to provide a more comprehensive understanding of protein behaviour.

1.2.3.3. Size exclusion chromatography

High-Performance Liquid Chromatography (HPLC) and Ultra-Performance Liquid Chromatography (UPLC) are highly standardised and well-understood analytical techniques for the separation, identification and quantification of small and large molecules. Chromatographic techniques involve the separation of different species in a solution based on their physical or chemical properties. Size-exclusion chromatography (SEC) relies on the particle size for separation, using the differential exclusion of pores in a bed of porous particles. Size exclusion chromatography is characterized by its gentle, non-adsorptive interaction with the sample, which preserves the biomolecular activity. SEC-HPLC is commonly used for its ability to separate antibody monomers from both aggregates and fragments (Diederich et al., 2011). HPLC instruments are coupled with different types of detectors, including UV diode array, fluorescence, or mass spectrometry.

Although an abundance of analytical techniques exists for the characterisation of all aspects of a protein species throughout the drug product development and manufacturing lifecycle, only a small number of spectroscopic and imaging techniques were discussed. The methods presented were chosen to facilitate an understanding of the analytical aspects of this thesis.

1.3. Statistical analysis for spectroscopic data

1.3.1. General regression

Linear regression is a well-known tool for analysing univariate, univariable or multivariable data. “Linear” models use a type of line equation to map out the relationship between variables, while “regression” refers to the grouping of the data around the mean (regression to the mean) (Heinisch, 1962; Seal, 1967).

To find the linear correlation of two random variables the Pearson’s coefficient (r) is used (*Eq. 1.8*). The relationship between the variables is measured as covariance (cov (x, y)) which indicates the degree of correlation between the variables, S_x and S_y represent the standard deviations of the variables (allows the measurement to be unitless).

$$r = \frac{cov(x, y)}{S_x - S_y} \qquad \textit{Eq. 1.8}$$

A value of r close to or equalling one indicates a direct positive linear relationship between the two variables, while a value of negative one indicates an indirect (negative) linear correlation. An (r) value of 0 does not imply that there is no correlation whatsoever only that there is not a directly linear one (Kirch, 2008).

To measure the degree of determination the coefficient of determination (R^2) value is used and is the square of the Pearson’s linear correlation coefficient. The value of R^2 represents the amount of variability in the dependent variable explained by the independent variable(s). R^2 is also a measure of predictive ability and a value close to 1 (closeness varies by discipline/natural variability within a population) is ideal. The adjusted R^2 value can also be used in cases of many variables and few observations (Draper and Smith, 1998).

1.3.2. Multivariate analysis

1.3.2.1. Introduction

Multivariate analysis (MVA) is a group of techniques often used when an individual in one or more samples is characterized using several measurements. As established the individuals are referred to as *observations*, while the actual measurements are called *variables* (Tsai, 2013).

Multivariate analysis examines datasets by analysing multiple variables simultaneously. This method helps to uncover the underlying structures and meanings present within the dataset, and it involves applying and interpreting various statistical techniques (Bhattacharya and Burman, 2016). When conducting analysis using a multivariate approach two factors are required to ensure meaningful results: I) a multidimensional data matrix and II) the need to preserve the complex structures present. Such structures are present when the variables are correlated so that only a full test may provide a complete understanding of the underlying structures within the dataset. It also follows that analysis interpretations through univariate and bivariate methods are unable to obtain all the desired information. The concomitant treatment of all the variables then reflects the reality of the addressed problem (Grimm and Yarnold, 1995).

1.3.2.2. Types of multivariate analysis

Multivariate techniques can be categorised according to their goals, and the relationships between the variables as presented in the dataset (Rencher and William, 2012). The first set of techniques is considered exploratory as they tend to generate hypotheses rather than test them. These techniques are often reductive, as they attempt to synthesise information through dimensionality reduction. For such descriptive techniques the process centres

around obtaining the optimal linear combinations of variables (Rencher and William, 2012). Where linear combinations are the result of a simple mathematical transformation. In linear algebra, the linear combination (as an expression or value) can be obtained from a set of vectors by multiplying each term by a constant and then adding all the results (Strang, 2016). In this way a linear combination of two or more variables is said to contain all the information of the original variables via their preserved variances. The two-fold benefit of linear combinations is their mathematical simplicity and proven ability to deal with the inherent structures studied in the natural sciences.

Examples of such techniques can be subdivided according to the type of input data into metric, and non-metric (Afifi et al., 2019). There are some examples of MVA handling non-metric data including correspondence analysis, log-linear models, and structural methods (Agresti, 2015). Examples of MVA of metric data include cluster analysis, factor analysis, and principal component analysis (PCA) (Wang, 2009). Some examples of which will be discussed in further detail in the following section.

On the other hand, to perform hypothesis testing multivariate techniques can be used while preserving the significance level and accounting for different intercorrelations of the variables. These techniques can be considered inferential and are extensions of univariate analysis. The primary advantage of these techniques is the ability to control the per-experiment error rate, as the significance level (α) remains constant no matter how many variables are tested simultaneously. As a result, the inferential techniques can be coupled with descriptive ones to provide a follow-up analysis. For example, as the statistical significance is preserved a descriptive technique can be used to find the variable (or combination of variables) that have led to a rejection and study their contributions to gain further insight (Rencher and William, 2012). Some examples of these techniques that

are commonly used include survival analysis, analysis of variance (ANOVA), and regression analyses (one of which, partial least squares regression is described below (Afifi et al., 2019).

The following is an overview of two MVA techniques that are highly representative of their respective goals (descriptive vs. inferential), as well as being highly relevant to this thesis.

1.3.2.3. Principal component analysis

As stated, PCA aims to determine important relationships between variables through the assessment of their linear combinations. To perform PCA we seek to maximise the variance of linear combinations of the variables. The first principal component is the linear combination with the maximum variance, or simply the dimension with the maximum spread or separation of samples. The second principal component will be the linear combination with the second highest variance, and so on. By combining variables we achieve our goal of reducing dimensionality while preserving the information within the variables truly responsible for the variance (or observed effect) within a sample (Rencher, 1998).

PCA is a one-sample technique and is applied to data with no groupings or stratifications. In contrast to regression analysis, PCA does not distinguish dependent and independent variables. Rather, it focuses on the core structure of a single sample of observations on a number of variables (Rencher, 1998; Rencher and William, 2012). To understand the outputs of PCA it is necessary to discuss the mathematical transformations involved. Although the specific equations are omitted (as they are outside the scope of the current

work) the following is a graphical interpretation of PCA calculations (a full outline is available in (Rencher and William, 2012)).

The following visualisations and data are adapted from the CRAN vignette by Harvey et al. (Harvey and Hanson, 2022). Table 1.2 shows a random data matrix of eight observations of two variables and Figure 1.8 shows a scatterplot visualisation of the data. The scatterplot shows a general trend of linear increase for the two variables.

Table 1.2 Example dataset for PCA analysis

Sample	*Variable a*	*Variable b*
1	2.66	3.32
2	-1.23	-5.15
3	0.25	-2.69
4	-3.86	-7.21
5	4.94	7.82
6	0.25	1.57
7	2.96	4.4
8	-3.96	-8.31

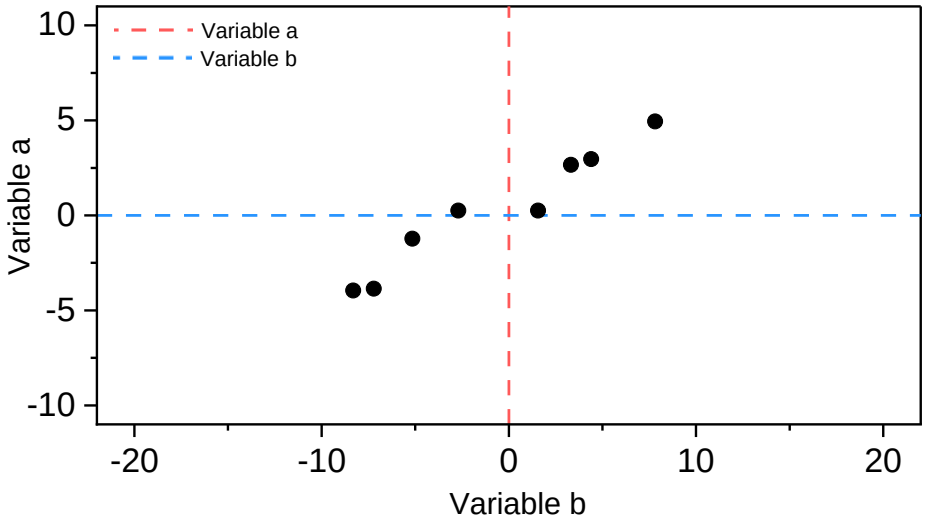


Figure 1.8 Scatterplot of variable a (red) and b (blue) for eight samples, The dashed lines are the original axes for the two variables.

To describe the degree of dispersion or spread within a number (n) of datapoints (x_i) we can use the variance equation (Eq. 1.9):

$$Variance = \frac{\sum(x_i - \overline{x})^2}{n - 1} \qquad \textit{Eq. 1.9}$$

Ideally the spread of data is representative of the changes in variable *a* with changes in variable *b* and thus representative of the phenomenon being observed. The variance can be conceptualised as the signal of interest. As such our goal with PCA is to optimise the signal to noise ratio, by maximising the amount of explained variance along a lower number of axis than the original data matrix (dimensionality reduction). For our hypothetical dataset the individual and total variances are calculated to be 10.29 (22.8%), and 34.84 (77.2%) for variables *a*, and *b* respectively. The total variance in the dataset is 45.13 (100% of the relative variance).

For any set of data the process of PCA would involve the singular value decomposition of linear combinations of the variables (Berry, 1992; Deprettere, 1989). The process can be visualised as a rotation of the (n) axes until an axis is found to explain the maximum amount of the total variance. This axis becomes the first principal component, then the process is repeated on an (n-1) dimensional plane orthogonal to the first. The process can be repeated infinitely for any number of variables.

For our demonstration of two linearly correlated variables this process can be simply performed once. For example, we can arbitrarily rotate both axes by 20° and project the points onto the new axes. The result of the rotation can be seen in Figure 1.9 with the solid lines representing the new axes (Axis 1 and Axis 2), and the dashed lines representing the original axes for variables *a* and *b*. The new axes can be thought of as "proposed principal components" and we can examine their effectiveness by calculating their explained

variance. We can see that the variance is now more equally distributed. However, both axes still play an important role in explaining the total variance.

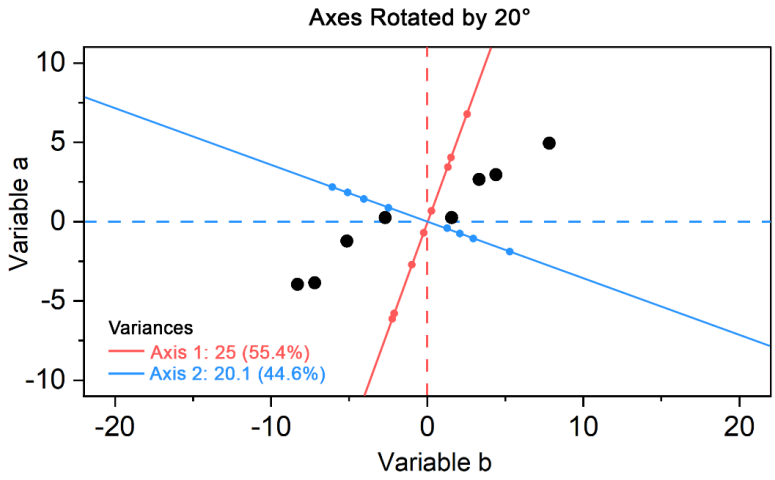


Figure 1.9 Rotating the original axes by 20° gives the axes shown as solid lines. The points in blue and red represent the projected datapoints of the variables onto the new axes.

After rotating the new axes by an additional 40°, the total rotation becomes 60°. This reveals that axis 1 is able to explain almost 99% of the total variance, as seen in Figure 1.10 Therefore, we can now consider Axis 1 to be the first principal component.

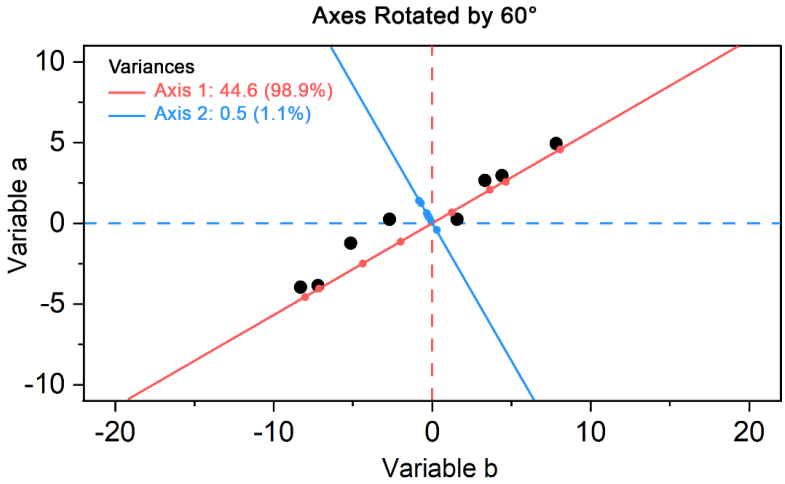


Figure 1.10 Rotating the original axes by 60° maximising the variance along one of the rotated axes. The line in red can now be considered the first principal component with the orthogonal blue line as the second.

Two major metric outputs of PCA are scores and loadings. The scores can now be simply understood to be the positions of the projections along our rotated principal components. While the loadings represent the cosine of the angle of rotation of the principal components relative to the original axes. The loadings are important for preserving information and allow us to reconstruct the original points from their principal components projections (scores). The loadings also indicate the relationship between the variable and the principal component. As rotations can range from 0° to 180° their cosines can only range from -1 to +1. Thus, a positive loading indicates that the presence of a variable contributes in some way to the principal component, while a negative loading implies the absence of the datapoint affecting the principal component. The loadings are often referred to as the "importance" of a variable to their principal component.

The principal components derived from the analysis can be used to identify patterns, relationships, and trends in the data (Jollife and Cadima, 2016). However, they are routinely used for analysing and reducing the complexity of high-dimensional data, to be used as inputs for other forms of analysis (Shlens, 2014).

1.3.2.1. Partial least squares regression

Partial least squares regression (PLSR), also known as projection to latent structures, is a popular inferential statistical method for predictive modelling (Höskuldsson, 1988). As with other predictive MVA methods, PLSR is a development of multiple linear regression (MLR) (Avkiran and Avkiran, 2018; Wold et al., 2006). As previously mentioned, mathematical regression is a method used to describe how a group of dependent variables (Y) behave in relation to a group of independent variables (X). When there are more

variables than observations or if the predictors are correlated, using ordinary MLR becomes difficult (Geladi and Kowalski, 1986).

PLSR is used when dealing with two sets of two-dimensional data. One set is the block of predictor variables, and the other set is the block of response variables. The purpose of PLSR is to determine the best regression coefficient between linear combinations of responses and linear combinations of the predictors. In this way, PLSR is similar to PCA in the use of linear combinations and singular value decomposition. Additionally, PLSR shows the correlation structure among variables for both blocks of variables (Geladi and Kowalski, 1986).

After a pilot model framework is established, the next step is the iterative improvement of the model parameters to increase the model's robustness (generalisability). Generalisation refers to the ability of a model to predict values outside of the tested set or interval. Cross-validation (CV) is a technique for the evaluation and testing of predictive models (Picard and Cook, 1984; Wold, 1978). However, it does not directly evaluate the performance of the global model of interest. Instead, it creates several local models with the same parameters as the global model and collects the performance results from these different models. This limits the cross-validation outcomes to the performance statistics primarily. Nonetheless, in most cases, this is acceptable for optimization purposes, assuming that the performance of the final model will be later assessed using an independent test set (Arlot and Celisse, 2010).

A typical cross-validation procedure entails dividing the dataset into two parts: the training dataset and the validation dataset. The former is usually the larger of the two datasets and used for model development and prediction of the validation dataset values. The cycle is repeated several times depending on the validation protocol. The results are

reported as the difference between the model's predicted results and the actual results in the validation dataset using known statistical identifiers (Arlot and Celisse, 2010) such as the coefficient of determination (R^2) for fit and values of root-mean-standard error for calibration (RMSEC), root-mean-standard error for cross-validation (RMSECV), and predicted residual error sum of squares (PRESS) for predictive ability. Common cross-validation procedures include hold-out, K-fold, and leave-one-out (LOOCV) cross-validations. The validation techniques can also be adapted to deal with complex datasets by coupling with stratification, nesting, and recursive analysis (Picard and Cook, 1984).

In hold-out cross-validation(Esbensen and Geladi, 2010), a dataset is divided into two subsets. The first subset accounts for approximately 80% of the data and is used for training the model. The second subset, constituting around 20% of the data, is used for verifying and validating the model results. This method is particularly useful for large datasets as it requires a single analysis to be run, thus saving both time and computational power. However, the hold-out method relies on the assumption of normal data distribution, and consequently may be compromised when dealing with datasets that exhibit more erratic distribution patterns.

K-fold cross-validation minimises the errors from the hold-out method and overcomes the informational bottlenecks of single testing (Geisser, 1975). To conduct K-fold testing for a dataset of n number of observations a K (number of folds) is arbitrarily decided ($K \mid K < n$). The dataset is divided into K equal parts and the training subset is set as (K -1) with the remainder subset used as the test set. The model is independently trained and validated k times using a new subset as a test set each time. K-fold is preferable to hold-out cross-validation as the training is performed on the entire dataset, thus avoiding issues caused by data distributions. The number of folds can also be increased to enhance the robustness

of the analysis. The contrasting downside of K-fold CV is the increased processing power and time required. Leave-one-out cross-validation (LOOCV) is the extreme case of K-fold CV where the number of iterations (K) is equal to the total number of observations (n). Although LOOCV improves the analysis resolution, the computational costs can be prohibitive for larger datasets (Picard and Cook, 1984).

Certain scenarios or datasets necessitate modifications to the above mentioned cross-validation methods. For instance, highly skewed datasets or datasets that include categorical stratification require the use of a stratified K-fold CV. This method involves splitting the data to ensure that each fold closely mirrors the strata distribution of within the original dataset. In addition, a specialised technique, known as repeated K-fold CV or repeated random sub-sampling CV, relies on the randomisation of selected observations for both testing and training.

The use of multiple MVA techniques in tandem is ubiquitous in the literature (Andrade et al., 2020; Kamga et al., 2013; Lee et al., 2012; Perera et al., 2021; Uarrota et al., 2014). One example that demonstrates the use of PCA and PLS is the publication by Lee et al. (Lee et al., 2012). The aim of the study was to understand the behaviours of soy hydrolysates in cell culture and their estimation through infrared spectroscopy and multivariate analysis. PCA was used to characterise the variability of soy hydrolysate lots captured via infra-red spectroscopy. The clustering of individual soy hydrolysate lots on PCA models constructed from bioassay data and infrared spectroscopy were compared. The PCA scores for each model showed clear delineation and grouping for soy hydrolysate lots obtained from different manufacturers. Although the scores were non-identical, the clustering was enough to conclude that infrared data correlated with lot-to-lot variability for soy hydrolysates detected via bioassay. Using this conclusion as a

starting point Lee et al. were able to construct PLS models to estimate the differences in behaviour for the soy hydrolysate lots in cell culture (dependant variable), from their infrared spectra (independent variables).

To conclude, MVA is a long-standing pillar of laboratory analysis that has seen many developments, algorithms, and validation tools over the years. MVA allows researchers to extract information from multifaceted and complex datasets in a robust manner.

1.4. Aims and objectives

The aim of this PhD project was two-fold:

- To explore analytical spectrophotometric techniques capable of measuring critical quality attributes of biopharmaceuticals reliably, reproducibly.
- To apply these techniques to a model protein solution for a holistic analysis of protein critical quality attributes, while maintaining sterility.

To achieve these aims the following objectives were established.

Chapter 2:

- To identify a method and parameters for the measurement of protein intrinsic fluorescence in-vial.
- To quantify the reconstitution time of a lyophilised protein formulation using the intrinsic fluorescence in-vial.

Chapter 3:

- To update the intrinsic fluorescence instrumental system in Chapter 2 for monitoring protein conformation, aggregation, and stability in-vial using fluorescence anisotropy.

Chapter 4:

- To relate individual aromatic amino acid fluorescence spectra to multivariate curve analysis deconvoluted protein fluorescence at 365~400 nm excitation.

Chapter 5:

- To corroborate the results of the stability analysis of protein formulations subjected to practical conditions of destabilisation using robust multivariate analysis

.

Chapter 2
UV excited fluorescence spectroscopy for accurate determination of reconstitution time of an in-vial lyophilised product

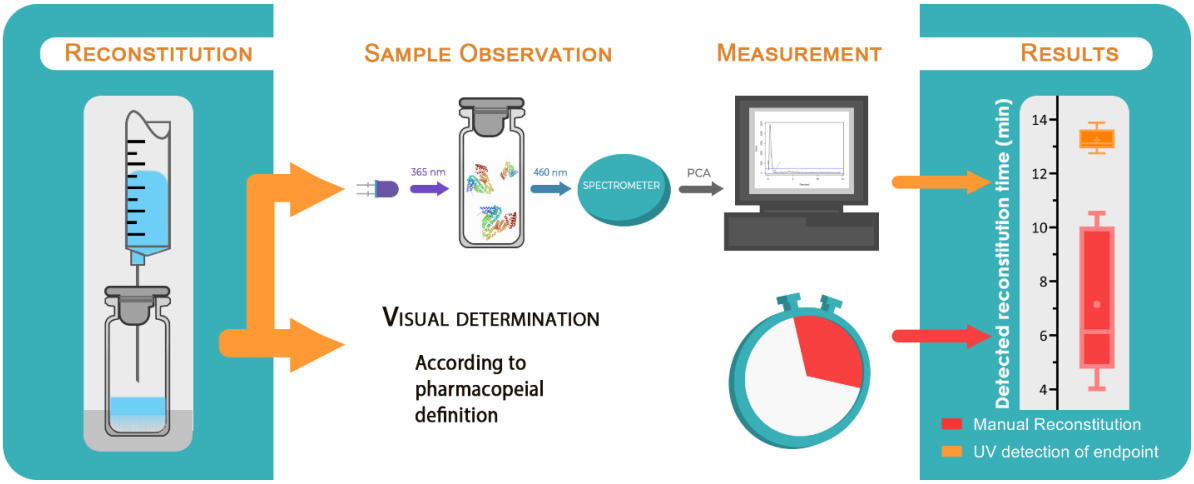
2. Chapter 2: UV excited fluorescence spectroscopy for accurate determination of reconstitution time of an in-vial lyophilized product

Published as:
ElKassas, K. and Chullipalliyalil, K., McAuliffe, M., Vucen, S., Crean, A., 2021. Fluorescence spectroscopy for the determination of reconstitution time of an in-vial lyophilised product. Int J Pharm.

2.1. Highlights

- By monitoring the fluorescence of a lyophilised protein during reconstitution the endpoint can be observed.
- Identification of an excitation wavelength of 365 nm enabled in vial monitoring of changes in fluorescence during reconstitution.
- Analysing the data using PCA allows the precise determination of the reconstitution endpoint.

2.2. Graphical abstract



2.3. Introduction

Many therapeutic protein formulations require stabilisation for long-term storage stability (Adamian and Liang, 2002; Edsall and McKenzie, 1983; Manning et al., 2010; Privalov and Gill, 1988; Rupley et al., 1983). The destabilisation of proteins in solution can be caused by interactions with the aqueous milieu and therefore can be minimised by freeze drying (Allison et al., 1998; Chansanroj and Thanawattanawanich, 2015; Deutscher, 1990; May, 2004; Snowman, 1997). The process of freeze-drying, also referred to as lyophilisation, is capable of reducing the moisture content of materials to below 3% (May, 2004). Lyophilisation is widely used in the processing of biopharmaceutical formulations to prepare some of the most common parenteral products on the market (V Gervasi et al., 2018).

During lyophilisation, the therapeutic protein solution is frozen to allow the extraction of water by sublimation below its triple point. The entire process proceeds below freezing temperatures thus avoiding any thermal degradation to thermolabile materials (Snowman, 1997). Lyophilisation converts liquid formulations to porous solid cakes. These solid formulations require solvation with a specific solvent, such as water for injection (WFI) or saline before they can be administered to the patient by injection or infusion. The process for rehydrating the lyophilised cake is termed “constitution” or “reconstitution”. The overall process is a combination of multiple physical processes occurring simultaneously, including wetting, disintegration, and dissolution. The United States Pharmacopoeia defines the criteria for completeness of constitution as *‘solid dissolves completely; leaving no visible residue as undissolved matter, or the constituted solution is not significantly less clear than an equal volume of the diluent or purified water present in a similar vessel and examined similarly’* (Awotwe-Otoo et al., 2013, 2012; Beech et al., 2015; Kulkarni et al.,

2018; The United States pharmacopeia, 2014). The time required to reach the clear state is considered an important critical quality attribute for lyophilised parenteral products (The United States pharmacopeia, 2014), as obtaining a clear injectable solution rapidly and consistently is important for patient convenience and safety.

The conventional method for measuring reconstitution time is by manual, visual confirmation of completed reconstitution as outlined in the previous definition. In this method, the test for reconstitution time is carried out by the addition of the reconstitution fluid via a needle and syringe through the stopper of the vial, then allowing the lyophilised cake to completely dissipate into solution, either by gentle agitation/swirling or by leaving the vial on a steady surface for a certain period of time. The endpoint is determined by visual observation of a clear solution that meets pharmacopoeia guidelines (Awotwe-Otoo et al., 2013, 2012; Beech et al., 2015; Kulkarni et al., 2018).

A number of recent studies have focused on characterizing, measuring, and optimizing the reconstitution of lyophilised protein formulations (Beech et al., 2015; Cao et al., 2013; Kulkarni et al., 2018; Werk et al., 2015). Due to the potential subjectivity of the visual confirmation methodology, the reported reconstitution times are highly variable, which is recognised as a key limitation of visual reconstitution time measurement. The visual method depends on the observation of the exact moment at which the solution goes from murky/opaque to completely clear which makes it difficult to quantify the reconstitution time accurately. Furthermore, exceedingly small, but visible or even sub-visible particles may persist in the solution after the determined reconstitution time when using only visual detection to identify the endpoint of reconstitution.

A recent attempt to circumvent the limitations of visual reconstitution determination by Werk et al. was to establish an automated, reproducible, impedance-based method for the evaluation of reconstitution time for a high concentration formulation of a model BSA protein (Werk et al., 2015). It was concluded that the equipment-based technique provided significant advantages in terms of standardization, reproducibility, and consistency, as user-caused variabilities are excluded.

Spectroscopic techniques can be used for the structural elucidation of molecules as well as monitoring physicochemical processes in real-time. Chandran et al. proposed a system for the classification of the reconstitution times of lyophilised formulations, which involved measuring the UV absorbances of different reconstituting samples in a 96-well plate (Chandran et al., 2013). A UV-vis plate reader was used to measure the absorbance and transmittance of the samples after addition of the reconstituting fluid. The study concluded a direct correlation between the relative reconstitution times of the manual and instrumental methods.

To our knowledge there is a dearth of studies detailing the use of fluorescence spectroscopy for the measurement of reconstitution time of lyophilised products. However, in an associated application, ElKhabaz et al. used real-time UV fluorescence spectroscopy to monitor drug release from an amorphous solid dispersion, where they were able to relate drug concentration with the intensity of the fluorescence at a specified wavelength (Elkhabaz et al., 2019). Fluorescence spectroscopy has an advantage over other spectroscopic techniques such as UV absorption, near IR absorption and Raman spectroscopy, in the fact that it has higher limits of detection (LOD) (Lakowicz, 1999; Patonay et al., 2014). UV-excited intrinsic fluorescence has proven capable of achieving

higher LOD in real-time in situ process monitoring in biopharmaceutical, pharmaceutical and dairy sectors (Bhartia et al., 2008; Cappius et al., 2015; Chullipalliyalil et al., 2019; Hug et al., 2014). Excitation in the ultraviolet wavelength range ensures the generation of "auto-fluorescence or intrinsic fluorescence" from the target molecule, thus removing any need for additional reagents.

The aim of this study was to investigate deep UV excited fluorescence spectroscopy as a method to quantify the physical endpoint of reconstitution of freeze-dried protein formulations. By analysing the total changes in the spectra over time, after the addition of the reconstitution fluid, and observing the point at which all changes cease, it was hypothesized that the reconstitution endpoint could be determined more precisely than the current manual, visual method. One important focus of this study was the ability to measure the reconstitution process in-vial. Removal of the cake from the vial can be difficult without fragmenting the cake structure and exposing the cake to atmospheric water vapour which can induce crystallinity and lead to inaccurate reconstitution time measurements. The primary challenge to overcome is the absorption of certain wavelengths of UV light by the walls of the vials. It was necessary to choose appropriate fluorescence excitation parameters that would not confound the results due to any interactions with the borosilicate glass of the vial. Hence the determination of an ideal wavelength for excitation became a significant part of this study.

2.4. Materials and methods

2.4.1. Materials and formulation preparation

All raw materials were obtained from Sigma-Aldrich, Ireland. Sucrose stock solutions (20% w/v) were prepared using de-ionised water and filtered under aseptic conditions using

0.2 µm syringe filters. Bovine serum albumin (BSA) was used as a model protein. The sucrose stock solution was added to the protein and diluted to reach the desired concentrations. Protein formulations studied are detailed in Table 2.1.

Table 2.1 Composition of freeze-dried formulations studied.

Formulation	BSA		Sucrose	
	%(w/v)	*mg/ml*	*% (w/v)*	*mg/ml*
F0S	0	0	20	200
F01	1	10	19	190
F05	5	50	15	150
F10	10	100	10	100
F15	15	150	5	50

Formulations were aliquoted at 2 ml pre-lyo volume into 10ml glass vials (FIOLAX® Clear neutral glass Type I Class B, Schott, Germany) under aseptic conditions and partially stoppered (20mm grey stoppers, Adelphi Healthcare Packaging, UK). Vials were placed at the centre of the shelf, in a hexagonal pattern, in an AdVantage Pro laboratory scale freeze-dryer (SP Scientific, Stone Ridge, *New York, USA*). Empty vials of the same size were used to completely fill the shelf, surrounding the formulation vials in order to reduce the variability of freeze-drying conditions due to the radiation effect from the walls of the freeze-drier (Rambhatla et al., 2005). Process parameters applied for the lyocycle performed are reported in Table 2.2. After completion of the lyocycles, the lyo-chamber was opened, and the vials were manually stoppered.

Table 2.2 Lyophilisation process parameters for the lyocycle used to prepare lyophilised samples.

	Shelf Temperature (°C)	*Ramp rate (°C)*	*Hold Time (min)*	*Vacuum (mbar)*
Loading	5	0.5	30	~700
Freezing	-40	0.5	300	~700
Primary drying	-30	0.3	2800	0.133
Secondary drying	25	0.3	240	0.133
Storage	5	NA	NA	800

2.4.2. Determination of reconstitution endpoint visually

Vials were un-stoppered and reconstitution fluid (2 ml of de-ionised water) was pipetted onto the cake. Timing of the reconstitution process began when all liquid was expelled from the pipette. Vials were left on the benchtop without agitation or disturbance. The reconstituting samples were observed throughout the reconstitution process. Reconstitution endpoint was determined visually when the solution was observed to be clear of all solid particulate material. An example of the appearance of lyophilised formulations during reconstitution can be seen in Figure 2.1.

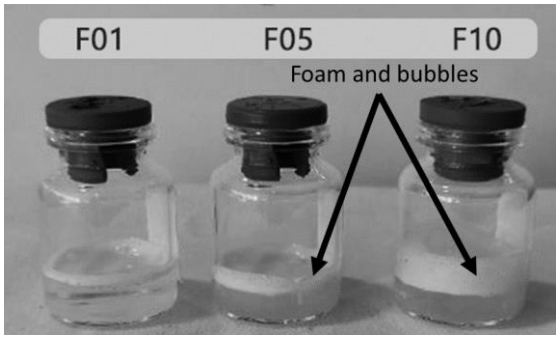


Figure 2.1 Lyophilised formulations after the addition of the reconstitution fluid (2 ml deionised water) showcasing the tendency of the lyophilised BSA/Sucrose formulations to foam upon reconstitution. F01, F05 and F10 represents 10mg/ml, 50 mg/ml and 100 mg/ml BSA respectively.

2.4.3. UV-Fluorescence reconstitution time measurement

The 3D plots of an excitation-emission matrix of a 15 % BSA and 5% sucrose solution after reconstitution obtained using a Perkin-Elmer LS-55 fluorescence spectrometer are shown in Figure 2.3. The measurements were performed with the solutions in a quartz cuvette. The stronger emission of BSA at 340 nm, as previously reported (Moriyama et al., 1996; R Core Team, 2019), can be observed for excitation wavelengths below 300 nm. As the excitation moves to longer wavelengths, the emission maximum of BSA is observed at 460 nm. The maximum efficiency for the BSA emission at 340 nm is when excited at 303

nm and for the emission at 460 nm was achieved when excited at 369 nm. An excitation-emission matrix of 5% sucrose solution is shown in Figure 2.2 (b). For excitation between 270-320 nm, the emission maximum of sucrose solution is at 348 nm, and a weak tail between 540-580 nm can also be observed.

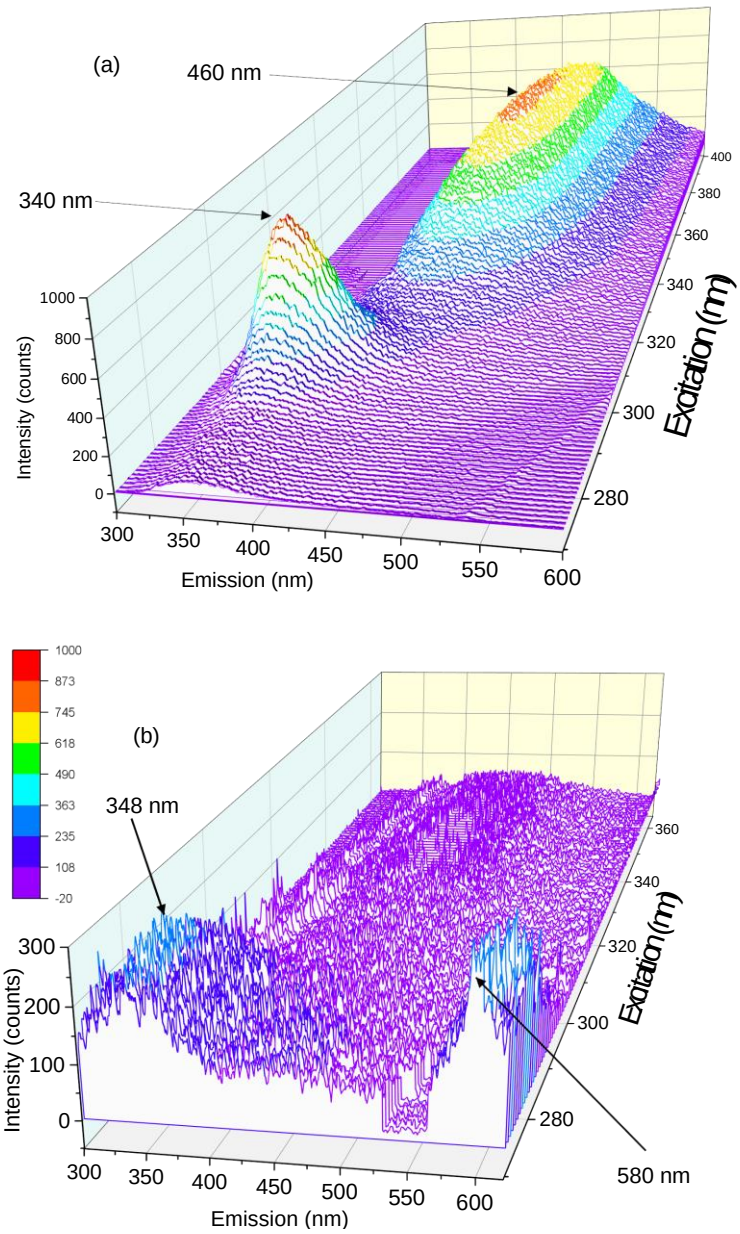


Figure 2.2 3D plots of the emission-excitation matrix fluorescence spectra. Each plot in the x-axis correspond to an emission spectrum at a particular wavelength of excitation (a) BSA solution (BSA 15% w/v, sucrose 5% w/v), for the emissions at 340 nm and 460 nm, the maximum efficiency is when the excitation is at 303 nm and 367 nm, respectively. (b) Sucrose 5% w/v solution, the emission spectra were observed at excitation between 270-300 nm.

Emission intensities in the regions 420-450 were also observed in Figure 2.2 (a). These intensities varied depending on composition and shifted as reconstitution proceeded. The phenomena underlying these changes in intensity were of interest and are explored further on in this thesis.

For in-vial measurements, the broadband absorption spectrum (250 nm – 1000 nm) of the lyophilisation vial had less than 60% transmission for wavelengths shorter than 320 nm (Figure 2.3). Hence, excitation wavelengths shorter than 320 nm could not produce sufficient fluorescence for the detection of BSA placed in-vial at lower concentrations. A broadband source would not have enough intensity to generate fluorescence and also interferes with both the emission at 340 nm and 460 nm as can be observed in Figure 2.3. Therefore, a 365 nm LED was used for the in-vial measurement set up, where the emission at 460 nm had maximum efficiency (Figure 2.3). The collection of the fluorescence was made from the bottom of the vial to avoid any loss of light due to the curvature of the vial.

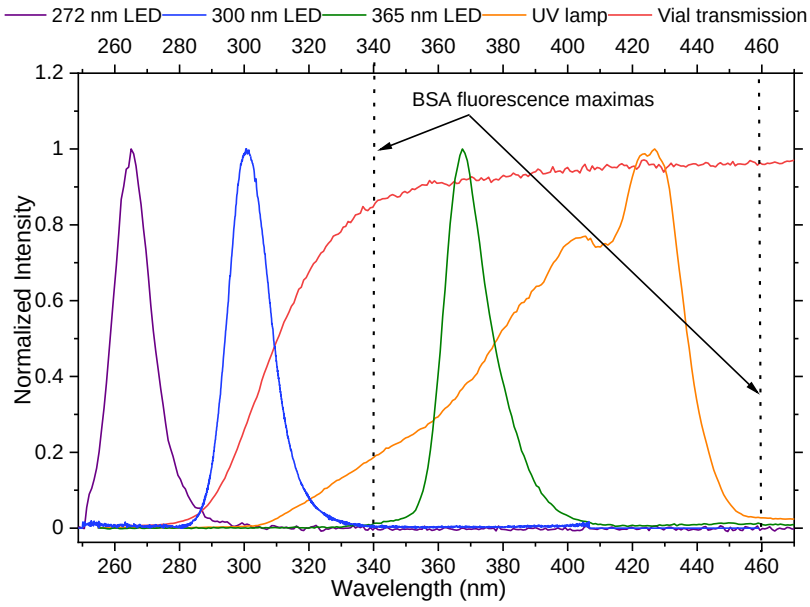


Figure 2.3 Normalised transmission spectra of an empty lyophilisation glass vial (in red) and different light sources in comparison. Spectra of LEDs at 272 nm, 300 nm, 365 nm and a standard UV lamp are shown. Wavelengths below 300 nm have less than 50% transmission. The position of two fluorescence maxima of BSA are also shown by dotted lines.

The schematic of the instrumental set up is shown in Figure 2.4. Vials (10 ml) were placed in a 3D printed sample holder to avoid any stray light. The excitation source was a UV LED (LED at 365 nm, Drive current= 200 mA, Drive voltage = 20 V, Radiant flux≈100-20 mw) which is end-connected to a Thorlabs UM 22−600 UV optical fibre. A focusable lens collimator with a PCX lens and a refocusing assembly was used at the end of the fibre, to focus the excitation light into the sample vial. The LED is operated at constant current for all the measurements. The fluorescence generated from the sample is collected via the flat bottom of the vial to an achromatic lens and onto a fibre. The fibre end is supplied to an Ocean Optics QE pro spectrometer with monochromator arranged in Czerny-Turner configuration (185-1100 nm) and a Hamamatsu S7030 back-thinned FFT-CCD on the detector side.

Data collection was started before adding 2 ml of de-ionised water to initiate reconstitution in a completely dark room, the dark box was capped. A total of approx. 1800 spectra were acquired for each sample analysed, with an acquisition time of 500 ms per spectrum. Measurement for individual samples took a maximum of approx. 15 min.

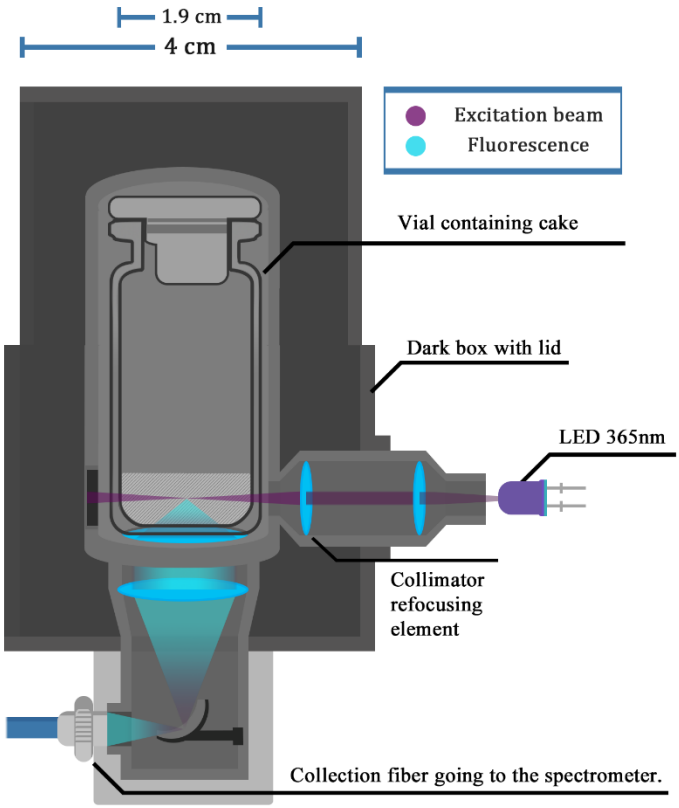


Figure 2.4 I*n-vial fluorescence detection set up showing a side view schematic of the sample holder details: The fluorescence collection from the bottom of the vial maximises information gain and minimises distortion of the light by the curved glass vial walls.*

2.4.4. Data pre-processing

Spectra were collected and analysed using Spectragryph® optical spectroscopy software [24] using the built-in baselining tool. Spectra exported were limited to the relevant wavelength range (420-700 nm) to avoid interference from the light source excitation signal. A Savitzky-Golay smoothing was applied to base-lined spectral data, and principal component analysis was conducted using the in-built "stats" module of (Rstudio v1.2.5001.0 (R Core Team, 2019) utilising packages; "BBmisc" (Bischl et al., 2017), "prospectr" (Stevens and Ramirez-Lopez, 2013), "BreakoutDetection" (James et al., 2014)).

2.4.5. Determination of reconstitution endpoint using fluorescence data

Three methods were explored with regards to reconstitution endpoint selection, each with a different level of user input required. A profile of the PC1 score for individual fluorescence spectra corresponding to each time point was used to determine the reconstitution endpoint (Figure 2.5).

2.4.5.1. Manual method

Reconstitution endpoint was determined by the user as the point after which the curve showed no major deviations or changes. To aid this potentially subjective measurement, the boundaries at ± 7.5 % of the average 1st derivative PC1 score of the equilibrium state values (last 300 seconds of a curve), Figure 2.5.

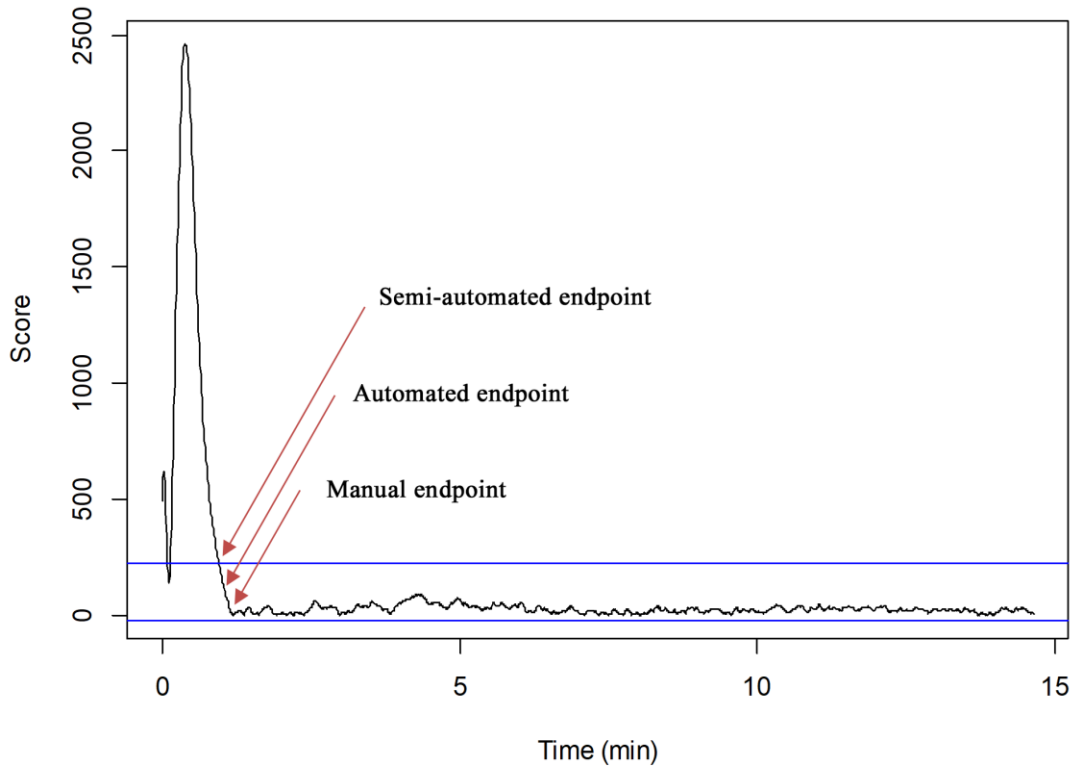


Figure 2.5 First derivative plot of PC-1 Score over time obtained by PCA of the fluorescence spectra

between 420-700 nm obtained during the reconstitution of a F01 sample (10 mg/ml BSA and 190 mg/ml sucrose) demonstrating the endpoint determination methods used. Blue lines represent ±7.5% variability in the signal. Endpoint was determined manually at indicated position (red).

2.4.5.2. **Threshold Variability method**

The endpoint was defined as the first time point after which the first derivative of the first principal component scores (PC-1) stabilised. The profile was considered to stabilise when the values were below a ± 7.5% variability limit determined from the values acquired in the last 300 seconds of monitoring. The user can determine this by observation of the last point to intersect with the upper horizontal line indicating ± 7.5% variability indicted in Figure 2.5.

2.4.5.3. **Application of Breakout Detection algorithm**

The algorithm employed was developed by Twitter to automatically detect breakouts in a time series of data (James et al., 2014). Breakouts refer to a change in one steady state to another: segmented shifts in the mean and/or gradual ramp up/down. E-Divisive with medians (EDM) is the underlying algorithm (Kejariwal et al., 2013). Breakout Detection was selected as an automated way to determine reconstitution endpoint from the time data series. It has previously been used to analyse time-based data series in a wide range of sectors, public health surveillance data (Wiemken et al., 2018), electricity theft (Messinis and Hatziargyriou, 2018) and stock prices (Alostad and Davulcu, 2015). A minimum of 30 consecutive data points was chosen to indicate a breakout by default. R version 3.3.1 (R Foundation for Statistical Computing, Vienna Austria) was used for all analyses.

Automated method utilizes the “Breakout Detection” library. At Beta values of 0.0125 the function maps the statistically significant breakout values in a times series (James et al., 2014), the last of which was taken to be the endpoint of reconstitution.

2.5. Results and discussion

The instrumental set-up was designed to take advantage of the inherent fluorescence of protein content of the lyophilised formulations. As with most proteins, BSA has been known to be easily detectable using fluorescence (Sharma and Kalonia, 2003)**.** The instrumental design was optimised to track changes in BSA intrinsic fluorescence during reconstitution in-vial. A key challenge encountered was the absorption of UV wavelengths < 300 nm by the wall of the vial, as the maximum efficiency of BSA intrinsic fluorescence at 340 nm is achieved when excited in the region of 300 nm. To minimise vial wall interference, alternative excitation parameters were identified at 365 nm which generate a weaker BSA intrinsic fluorescence signal around 460 nm. To maximise fluorescence detection a UV LED at 365 nm excitation source was employed and the fluorescence collection point was located at the base of the vial to avoid any loss of light due to vial curvature.

2.5.1. Spectral responses during reconstitution

UV excited fluorescence spectroscopy was investigated with the objective of measuring the reconstitution time of lyophilised protein formulations in-vial. It was hypothesised that spectral information captured in real-time during the reconstitution process would enable the endpoint of reconstitution to be accurately identified based on two basic premises. Firstly, during reconstitution, physicochemical changes occur during the process up to the endpoint (Sharma and Kalonia, 2003). Secondly, changes in the system as reconstitution

occurs could be detected by monitoring the changes in the fluorescence of emitted spectra (Rupley et al., 1983).

During reconstitution, the cake structure breaks up into suspending particulates. These particulates scatter photons from the light source elastically, which results in the initial detection of the Rayleigh scattered light source at 365 nm (Figure 2.6(a)). As the floating particulate matter dissipates into solution the scattering is diminished, following the translucency of the reconstituted sample as it goes from opaque to clear.

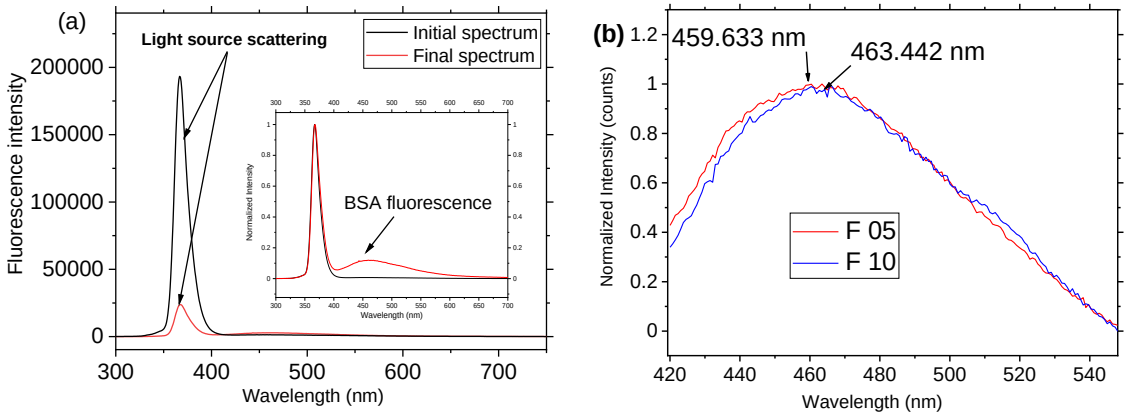


Figure 2.6 a) Plot showing the initial fluorescence spectra at time 0 (black) and last fluorescence spectra at 15 min (red) acquired during the reconstitution of lyophilised formulation F05 (50 mg/ml BSA and 150 mg/ml sucrose) in-vial. LED excitation at 365 nm. Decreasing opacity causes reduction of scattering and decreases peak intensity at 365 nm. Increasing solubilisation of BSA increases peak intensity at 428 nm (BSA fluorescence peak). The inset shows the same spectra after normalization. (b) Red Shift of fluorescence with the increase in concentration of BSA from 50 mg/ml (F05) to 100 mg/ml (F10).

Energy transfers among the C=C and C=O double bonds which are favoured in high-concentration BSA, as well as the proximity of the amine groups (e.g., lysine, arginine, etc.), resulting in a red shift in the observed fluorescence (Perucho et al., 2015). A fluorescence emission maximum shift of ~ 4 nm was observed as the concentration of BSA increased from 50 mg/ml to 100 mg/ml (Figure 6b).

The fluorescence spectra captured during the reconstitution of the four lyophilised formulations with 0, 10, 50, and 100 mg/ml BSA can be seen below in Figure 2.9. The

changes in spectra with time show interesting features due to their differences in component ratios (BSA: Sucrose), time scale, and underlying physical phenomena.

Aromatic amino acids present in BSA such as tryptophan, tyrosine and phenylalanine are known to be a source for intrinsic protein fluorescence in the UV-blue wavelength region. Their spectral distribution varies depending on the energy transfer processes that can happen according to the molecular environment. BSA has two tryptophan residues that produce intrinsic blue fluorescence: Trp-134 and Trp-213. This tryptophan emission at 340 nm (Figure 2.2) dominates BSA fluorescence spectra in the UV region, but only when excited at or below 303 nm and can shift with the polarity of the medium (Moriyama et al., 1996). When excited at 365 nm, an intense emission can be observed at 463 nm for the 50 and 100 mg/ml BSA formulations. An example of BSA fluorescence in the solid and liquid states can be seen in Figure 2.7.

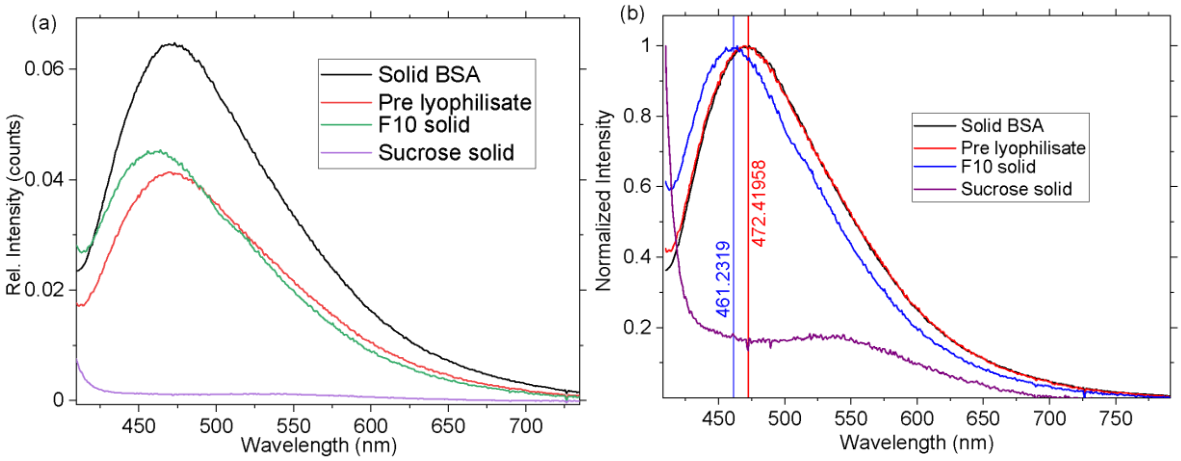


Figure 2.7 (a) Solid state fluorescence spectra of (i) BSA, (ii) pre-lyophilisation mixture, (iii) Post lyophilisation (F 10 used in the study) and sucrose. (b) Normalized spectra. All the spectra are measured using the modified set up described in the chapter.

The light beam from the exciting LED (365 nm) passes through the centre of the vial and intrinsic BSA fluorescence is generated and captured by the collection lenses (Figure 2.4, set up diagram). For samples with higher concentrations, the fluorescence emitted may be

quenched to a greater extent than the lower concentration solutions prior to detection. A comparison of BSA fluorescence for equivalent concentrations is seen in Figure 2.8.

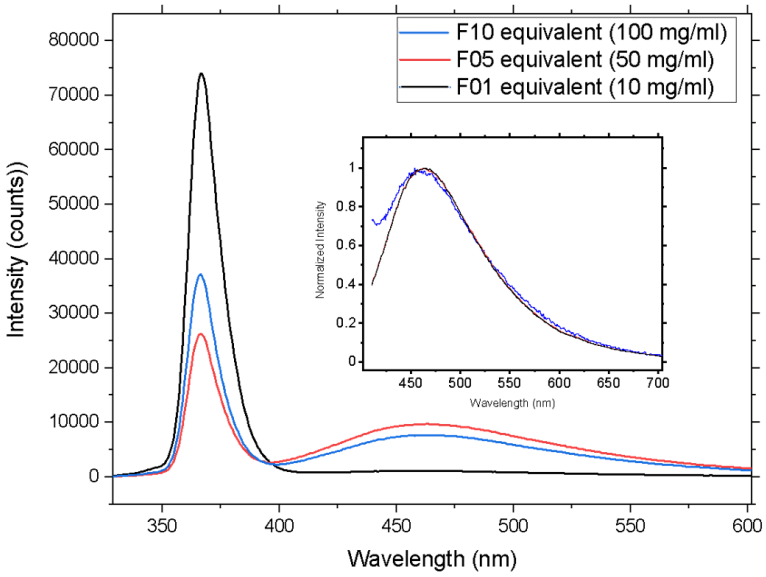


Figure 2.8 BSA Fluorescence response variation with concentration. As can be seen, there is only a little shift in the fluorescence maxima for the range used in the study between F01 to F10. There is a difference in the source intensity detected in the spectrometer for each concentration, indicating difference in the light absorbed and converted into fluorescence in each case.

The F10 formulation concentration quenching results in a fluorescence that is overall slightly lower than F05. This effect can be observed in the emission spectra of an F10 (100 mg/ml BSA and 100 mg/ml sucrose) formulation captured over the reconstitution period. Fluorescence increases in intensity for a short period, decreases, then slowly rises again to a maximum intensity (Figure 2.9(a)). The maximum intensity of the F10 formulation is lower than that of the F05 formulation (50 mg/ml BSA and 150 mg/ml sucrose) possibly due to re-absorption by BSA in the sample (Figure 2.9(b)).

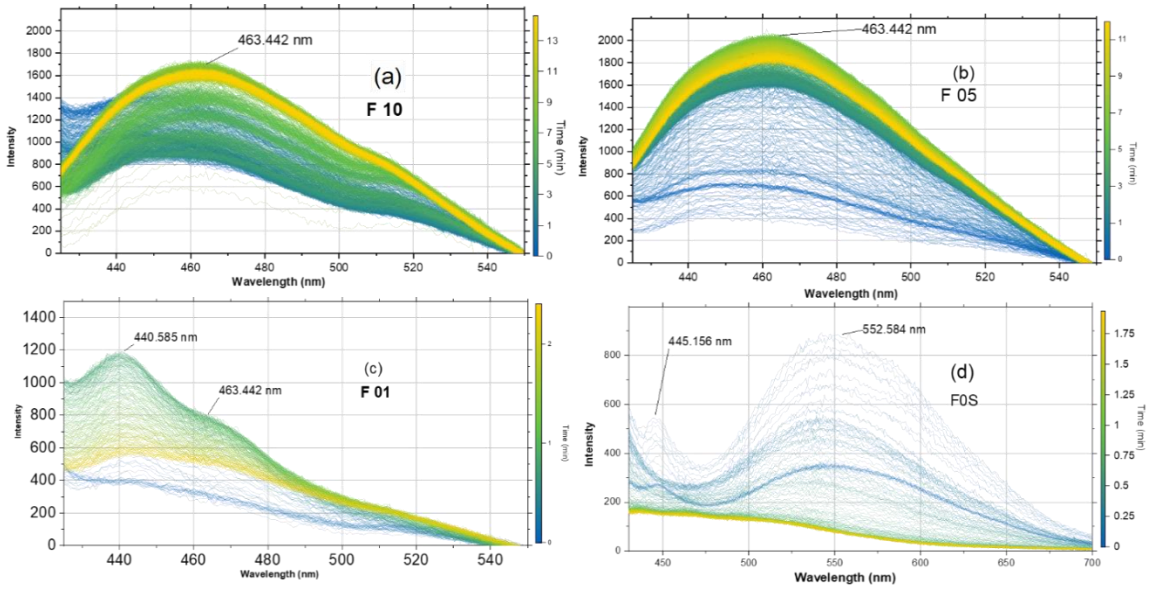


Figure 2.9 Raw fluorescence spectra acquired during the reconstitution of lyophilised BSA formulations (time shown on colour gradient), a) F10 (100 mg/ml BSA and 100 mg/ml sucrose), b) F05 (50 mg/ml BSA and 150 mg/ml sucrose), c) F01 (10 mg/ml BSA and 190 mg/ml sucrose), and d) F0S (200 mg/ml sucrose). The main emission maxima for each formulation are also shown.

Figure 2.9 (c) and (d) show spectra for formulations F01 (10 mg/ml BSA and 190 mg/ml sucrose) and F0S (200 mg/ml sucrose), respectively. These formulations had a significantly higher proportion of sucrose to BSA compared to formulations F05 and F10, and hence an emission from sucrose was observed around 550 nm. Even though the efficiency of this emission is low when excited at 365 nm (as observed from sucrose the excitation-emission spectra in Figure 2.2), the high-powered LED and the high collection efficiency of the set up enables the detection of sucrose. Formulation F01 had a significantly higher proportion of sucrose to BSA compared to formulations F05 and F10.

Figure 2.10shows the variation in maximum intensity at the main emission maxima shown in Figure 2.6, namely at 438 nm, 463 nm, and 550 nm. The emission wavelengths reported for tyrosine detection are around 420-450 nm (Perucho et al., 2015). An emission in this region was observed for the 100 mg/ml BSA formulation at early time points but cannot be clearly observed in Figure 2.9(a). The emission intensities at 438 nm, and 463 nm are

attributed to emission from the BSA (Figure 2.9(a) and Figure 2.9(b)), while 550 nm emissions were attributed to sucrose (Figure 2.9(d)).

A rough estimate of the reconstitution time can be determined from the plots shown in Figure 2.10; the time that the emission intensity stabilizes, as seen for all formulations. BSA is shown to be the rate-limiting factor in the dissolution of the lyophilised formulations, hence a low BSA concentration results in a fast dissolution. For the formulations with 10 mg/ml BSA or no BSA (F01 and F0S), dissolution occurs within 100 s to 40 s respectively. It can be observed that as the formulations become more concentrated with BSA, the reconstitution of all components of the formulation are retarded. Indicating that the reconstitutions of the formulation components are not independent from one another.

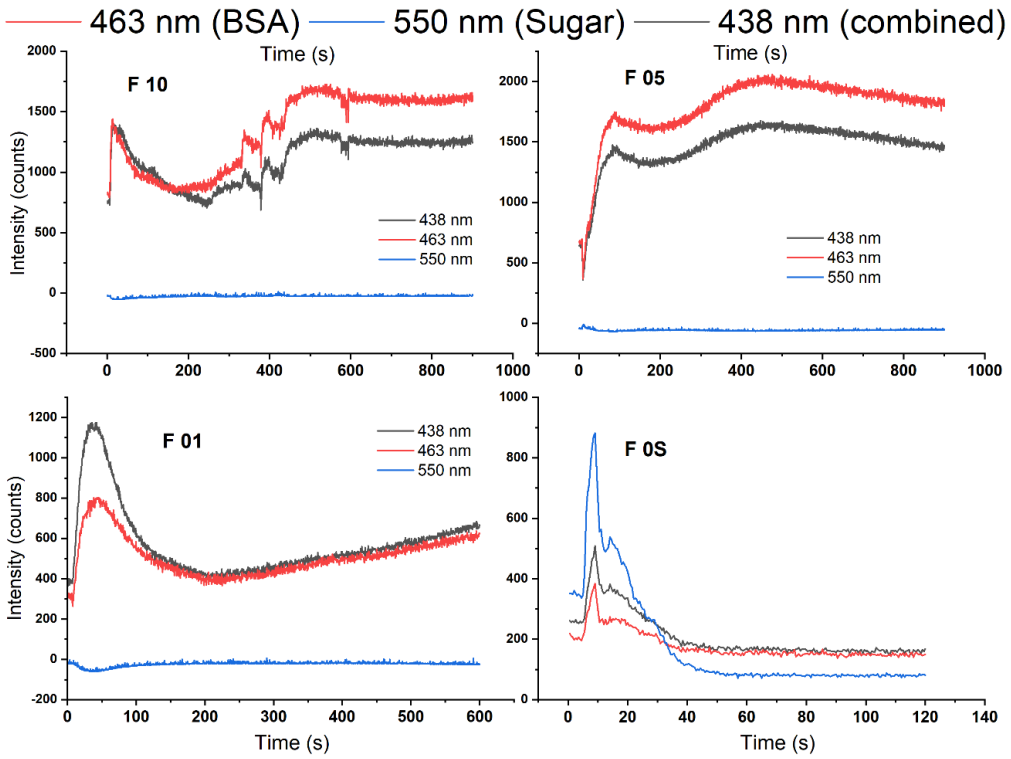


Figure 2.10 Variation of the maximum intensity of the emission from BSA collected during the reconstitution of the four formulations. The intensity at 438 nm has contributions from both BSA and sugar, intensity at the higher concentration solution exhibits a lower plateau due to fluorescence reabsorption.

2.5.1.1. PCA for accurate endpoint determination

For a more precise determination of the reconstitution time, it was necessary to analyse the data further to observe the effect of BSA concentration on the spectra acquired, peak height changes are not directly comparable between different formulations or may result in inaccurate reconstitution rate interpretations. Consequently, performing an analysis of the underlying patterns in the spectra using principal component analysis was performed to ensure that only significant changes are represented in the results. The spectra were collected in sequence over time, each spectrum representing the same solution at a different point in time. The average percentage variability described by the PC-1 eigenvalues was from 88 % up to 96% (individual values) for formulations F01, F05 and F10 respectively. Therefore, it was concluded that the PC1 score was capable of physically describing the changes in solution.

Plotting the PC-1 scores over time revealed a reproducible pattern of changes for formulation (F01) during reconstitution. The PC-1 curves were derivatised to facilitate the determination of the endpoint. The plots can be seen in Figure 2.11 (a) and (b) respectively. The endpoint was determined when the derivatised PC1 score remained within certain boundaries (15%) about the baseline value, for a continuous period (300 sec) as previously outlined in the methods section.

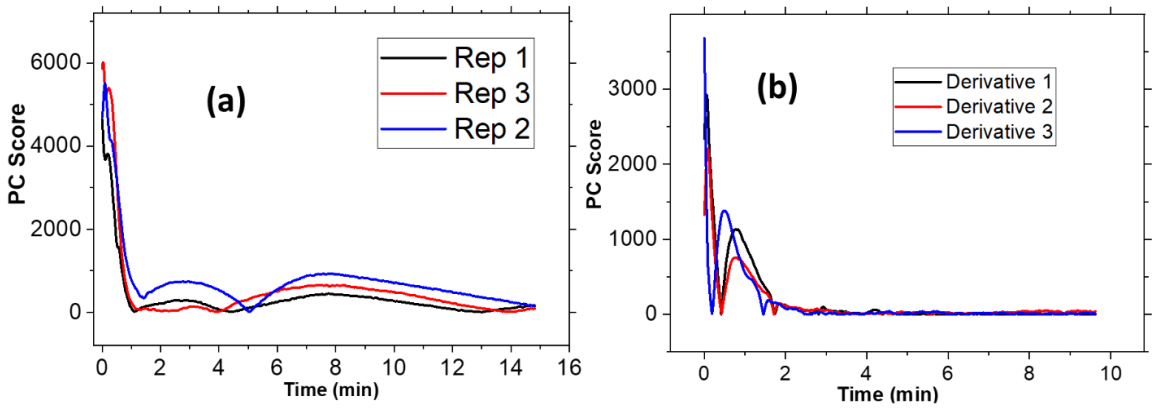


Figure 2.11 (a) Profiles showing changes in PC-1 scores of fluorescence spectra with time for replicate samples of formulation F01 (10 mg/ml BSA and 190 mg/ml sucrose) during reconstitution, (b) profiles of the 1st derivatives of PC-1 scores with respect to time over the same period (n=3).

Figure 2.12 shows the PC-1 profiles of all the tested formulations included in the reconstitution time calculations below. Samples used for a pilot study (namely Derivatives 4 and 5 in Figure 2.12) are also shown. The initial setup resulted in baseline drift over time, and revealed that the initial observation time of 600 seconds was insufficient for higher concentration formulations (F05 and F10). Derivative 4 and 5 results informed later experiments and the baseline drift was corrected via detector calibration.

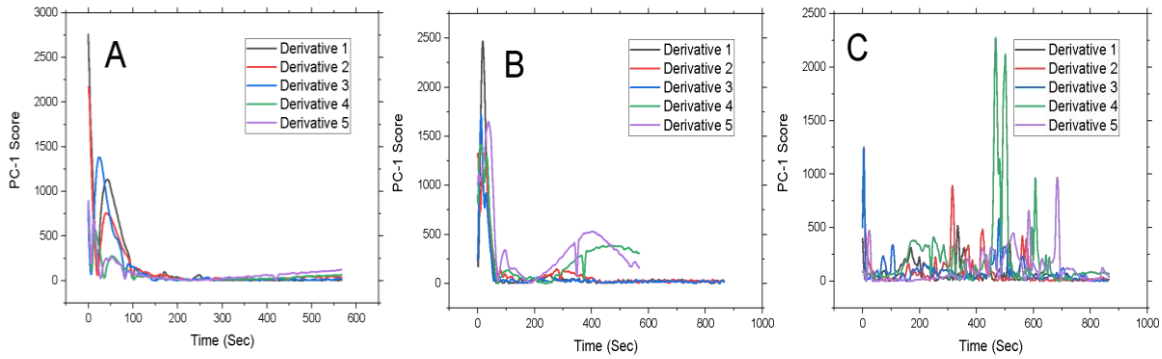


Figure 2.12 PC-1 1st derivative plots for all replicates (n=5) of a specific formulation (F01, F05, F10 respectively). Note the baseline drift in the second part of the observation which is more apparent in figures A and B which occurred due to setup variables. The drift was excluded from the reconstitution time calculation and corrected in later experiments.

2.5.1.2. Reconstitution endpoint determination

As described in the methods section three approaches were used to determine the reconstitution endpoint from the 1st derivatives of PC-1 scores with respect to time. These methods included visually determining when a steady state is reached (Manual Method), calculating the time point when variability in response obtained was below ± 7.5% variability determined from the values of the last 300 sec (Threshold Variability) and application of the Breakout detection algorithm (Breakout Detection). Reconstitution endpoints determined using transformed fluorescence data were compared to the traditional visual observation method. The reconstitution endpoints determined for the lyophilised formulations using the different methods are shown in Figure 2.13.

Formulations F01 (10 mg/ml BSA and 190 mg/ml sucrose) and F05 (50 mg/ml BSA and 150 mg/ml sucrose) show reconstitution times (~2 min) that are independent of differences in BSA concentration. However, the formulation with a higher BSA concentration, F10 (100 mg/ml BSA and 100 mg sucrose) took significantly longer (~14 min) to reach complete reconstitution (Sane et al., 2020). The reconstitution end point determined by visual observation for F10 was 7.18 ±2.47 min. The instrumental methods determined a narrower range of reconstitution times for replicate samples. The instrumentally measured reconstitution times are on average longer than those obtained by visual observation, which is attributed to the presence of sub-visible particles and/or human error. For formulations F01 and F05, the reconstitution times determined were similar for the three endpoint identification techniques employed to analyse fluorescence data with results within a 30-second window. For formulations, F10 reconstitution endpoints significantly differed

between the three techniques ($p < 0.025$) with all results being within 3 minutes of one another.

Overall, the reconstitution times detected instrumentally were less variable than the visually determined values. The instrumental methodology presented has higher time resolution than traditional visual reconstitution; a collection of spectra takes place every 500 milliseconds, which results in 1800 spectra for a 15-minute period of observation. The great number of spectra allows the determination of the exact time of reconstitution completion with far greater precision than is possible with the naked human eye.

The greatest variability was observed between instrument measurement techniques for the high-concentration formulation, containing 100 mg/ml BSA. Literature available on the reconstitution times of lyophilised high-concentration protein formulations shows that these formulations exhibit long and variable reconstitution properties. This is also corroborated by the results in Figure 2.12 (C). For example, at concentrations of 100 mg/ml BSA the reported reconstitution times ranged between 6 and 30 minutes (Cao et al., 2013). These variable and prolonged reconstitution times agree with those reported in this study (Figure 2.13). Longer average reconstitution times were determined using the instrumental technique compared to the visual method. These differences can be attributed to fluorescence operating on a molecular level (average radius 90 angstrom (Shire, 2015)), rather than a visible particulate one (30 microns) and the opaque nature of these formulations during constitution.

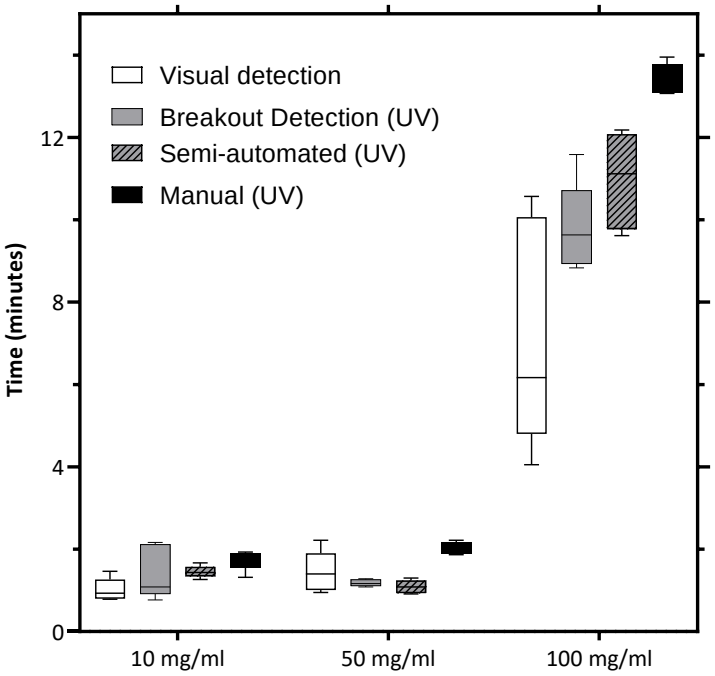


Figure 2.13 Comparison of reconstitution endpoints determined visually versus those obtained by analysis of fluorescence data. The x axis indicates the concentration of BSA in each lyophilised formulation F01 (10 mg/ml BSA and 190 mg/ml sucrose), F05 (50 mg/ml BSA and 150 mg/ml sucrose) and F10 (100 mg/ml BSA and 100 mg/ml sucrose). Y-error bars indicate standard deviation (n=5).

Of the three analysis techniques investigated to pinpoint the reconstitution endpoint from the fluorescence data, the manual method showed the least variability. However, as the manual method requires a trained operator to recognise the events occurring during the reconstitution and determine the endpoint according to their interpretation of the statistical data, this leaves room for end-user error. Both the Threshold Variability and Breakout Detection methods detect reconstitution endpoints without operator influence and are therefore less prone to subjectivity and error. Threshold Variability is a relatively simple approach compared to the more complex Breakout Detection algorithm. While the underlying statistics are more complex the open-access nature of the algorithm facilitates its use. To our knowledge this is the first time the Breakout Detection algorithm has been reported for pharmaceutical analysis and its full potential compared to other data analytics approaches has yet to be fully explored.

The method provided a precise technique to measure the reconstitution endpoint of both high (100 mg/ml) and low (10 mg/ml) protein concentration in lyophilised formulations. High protein concentration formulations are of interest due to the increased interest in monoclonal antibody products designed for self-administration in the form of subcutaneous injections. These formulations mandate low administration volumes (<1.5 ml) and therefore high concentrations of the protein (>100 mg/ml) (Awotwe-Otoo et al., 2013; V Gervasi et al., 2018). The advantages of using an instrumental method to observe the reconstitution process rather than depending on the visual acumen of the analyst are many. Firstly, for accurate quantification of the reconstitution time it is required by the analyst to observe (using the naked eye) the exact moment when a solution goes from turbid to clear. The endpoint is not only difficult to precisely and consistently pinpoint but is subjective to the analyst's judgement on how to apply the worded definition to a real-world phenomenon. It is also possible that for higher-concentration lyophilised protein formulations, visible or sub-visible particles can remain out of the solution without being detected by the analyst or administrator. It has long been known that the limit for visual resolution by the naked eye is around 1 arcmin or 1/60° (Campbell and Green, 1965; Radner and Benesch, 2019), which could be calculated to an equivalent resolving capability of 30 microns for an adult. Furthermore, visual detection is complicated by the tendency of lyophilised protein formulations to foam during reconstitution and cake particulate matter becoming entrapped in the foam bubbles, unable to reach the dissolution media for a time (Kulkarni et al., 2020, 2018) as presented in Figure 2.1. It can therefore be said that the detection of the reconstitution endpoint by visual observation is inherently limited by a human analyst's eyesight's natural limit of resolution.

Another advantage of an instrumental technique is the resolution of inconsistencies in reported reconstitution times throughout the literature as has been discussed by Cao et al. (Cao et al., 2013). Inconsistencies in the reported reconstitution times of protein formulations were attributed to the different reconstitution procedures used by different analysts to achieve reconstitution.

The UV excited fluorescence spectroscopy method for reconstitution endpoint detection presented in this study provides an easy-to-use, precise, and reproducible method, which could be potentially used in both lab and clinical environments.

2.6. Conclusions

The UV-excited fluorescence spectroscopy methodology presented provides a precise and reproducible method for in-vial measurement of the reconstitution time of lyophilised protein formulations. An excitation wavelength of 365 nm was identified to capture intrinsic BSA fluorescence during in-vial measurements, minimising interference due to the vial wall. Changes in formulation intrinsic fluorescence during reconstitution were captured by principal component analysis of spectra acquired. Reconstitution endpoints determined using this fluorescence methodology were less variable, and at high concentrations more prolonged than those determined visually. The instrumental methodology presented offers a number of advantages over the traditional visual methodology and could potentially be employed to support formulation design and quality control of lyophilised formulation, as well as to confirm complete formulation reconstitution in clinical settings.

Chapter 3
In-vial detection of protein denaturation using intrinsic fluorescence anisotropy

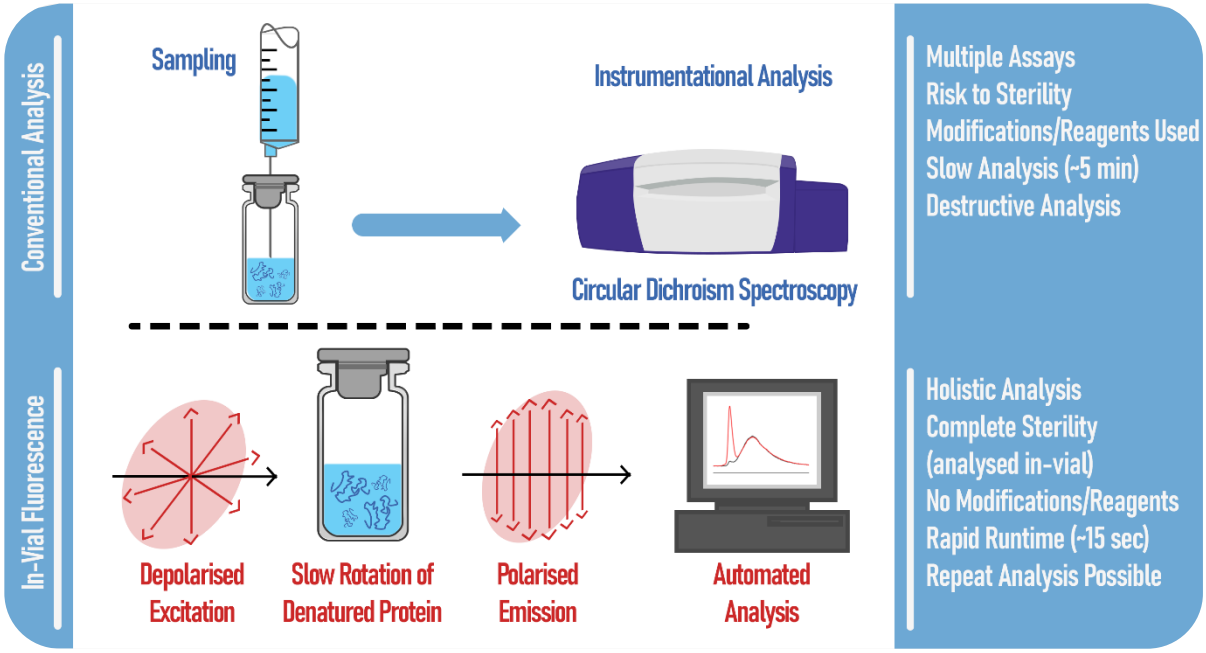
3. Chapter 3: In-vial detection of protein denaturation using intrinsic fluorescence anisotropy

Published as:
Chullipalliyalil, K. and Elkassas, K., McAuliffe, M.A.P., Vucen, S., Crean, A., 2023. In-Vial Detection of Protein Denaturation Using Intrinsic Fluorescence Anisotropy. Anal Chem 95, 2774–2782.

3.1. Highlights

- Fluorescence anisotropy was used to measure the stability of proteins using a bespoke apparatus for in-vial analysis.
- A stability study was conducted and a model protein was degraded to produce aggregated and unfolded protein samples.
- Fluorescence anisotropy was used as an indicator for unfolding while scattering was used to measure aggregation.

3.2. Graphical Abstract



3.3. Introduction

The majority of biopharmaceutical products, particularly vaccines, still rely on refrigeration or freezing for storage and transport to maintain stability. Due to poor conditions in some countries, maintaining the cold chain has become difficult, especially after COVID 19 (Brenner, 2015; Khan et al., 2022; Tirivangani et al., 2021). Instances of disruption could be undetectable, leading to deleterious effects such as immunogenic reactions or inefficacy upon administration due to product instability (Matthias et al., 2007). Therefore, an essential part of the cold chain supply process is establishing the protein structure and stability state. As degradation can follow multiple pathways, often concurrently, a combination of lab-based techniques is employed for routine assessment of the different aspects of protein stability.

Techniques employed include SDS-PAGE for fragmentation, ion exchange chromatography, isoelectric focusing, reverse-phase HPLC for chemical degradation, and size-exclusion HPLC for aggregation (Federici, 1994). These techniques require the removal of the sample from its storage vial. Regulatory authorities have stringent requirements on sample removal from storage vials since it may breach the parenteral drug sterility (FDA, 2005; Davis, 2008; Polen, 2006). In addition, a limitation of some of these techniques is the need for sample manipulation such as dilution into a suitable concentration range or exposure to chromatography phases. These manipulations can alter the conformational state of the protein sampled (Wang et al., 2010). A closed vial measurement would allow measurement of the protein conformational status and denaturation levels without breaking the seal and without changing the local environment. Such a technique could provide biopharmaceutical stability data along the product's journey from manufacturer to patient. An ideal technique would perform non-destructive

product analysis, be portable, easy-to-use to enable use in underdeveloped countries with high clinical need. One possible solution to developing such a technique to monitor a sample's degradation would be to use a sensitive technique such as intrinsic fluorescence polarisation.

Fluorescence polarisation happens when a molecule excited by linearly polarised light emits light preserving the polarisation direction (Jain et al., 2011; Lakowicz, 1999; Mann and Krull, 2003). If linearly polarised light is used to excite the molecule, it will preferentially select molecules with an absorption transition moment parallel to the direction of light polarisation (Figure 3.1). The light thus emitted by each molecule has its polarization in a plane parallel to that of the absorption transition moment. It has been well established that this tendency can be used to measure structural changes in large molecules (Grossman, 1984; Mann and Krull, 2003; Poklar et al., 1994). If the measurements are done in solution, the molecules probed are free to rotate. Hence, the polarisation of the emitted light depends on the fluorescence lifetime of the molecule and its extent of rotation in the solution. The molecule's rotation depends on the solution viscosity, its absolute temperature, molecular volume and gas constant. If the viscosity and temperature of the solution are constant, then the molecular volume is the critical variable determining the rotational speed difference. Hence if polarization is monitored, the kinetics of proteins which involve a volume change, such as protein folding and unfolding, can be monitored in real-time. Typically, the anisotropy is measured using a fluorophore as a label, such as laser dyes (fluorescein or rhodamine), by linking them to peptides or nucleic acids. Intrinsic fluorescence allows measurement without needing fluorescent labelling or modification of the protein structure.

Utilizing intrinsic fluorescence anisotropy of fluorophores like tryptophan enables a broader view of protein stability to understand the conformational structure of a protein in solution. Intrinsic fluorescence has been used before to study protein denaturation(Vlasova and Saletsky, 2009) for open lab samples. Single-Molecule Förster Resonance Energy Transfer (smFRET) and Fluorescence Correlation Spectroscopy (FCS) have been used before to understand the denaturation level, but not for in-vial solutions (Chen and Rhoades, 2008). The unavailability of suitable excitation sources has also prevented the use of intrinsic fluorescence for in-vial measurements until recently (K. ElKassas et al., 2021).

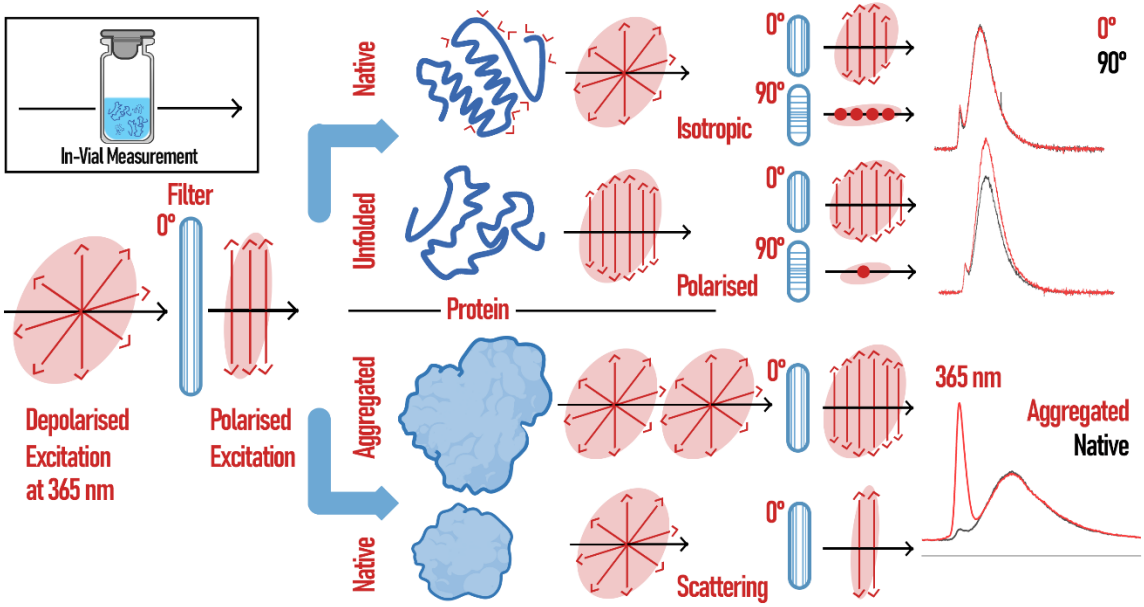


Figure 3.1 Principle of intrinsic fluorescence polarisation: A significant difference in the total intensity is detected between the orthogonal polarization directions if the luminescence emitted from the molecule retains the polarization.

The study detailed in this chapter aims to investigate intrinsic fluorescence polarisation as a method for developing handheld devices that can quantify the denaturation/unfolding of in-vial protein formulations. This chapter building on the results in Chapter 2 in which in-vial measurements for identifying accurate protein re-constitution times before using the

excitation at the absorption tail of bovine serum albumin (BSA) at 365 nm. In this chapter the instrumental setup developed in Chapter 2 is modified to a luminescence polarisation set up for identifying the level of denaturation in a closed vial. The setup can be further developed into a handheld system that can judge the level of denaturation among a group of samples by just switching the sample, which would form the commercial aspect of the work. Samples of solutions of a model protein, bovine serum albumin (BSA), with different denaturation levels are measured with the setup and compared with circular dichroism measurements. Since size changes caused by aggregation can also bring slower rotation and thus an increase in anisotropy, it is also essential to check if the contribution from aggregation can be understood using the same technique.

Hence two sets of samples were used in the study: (1) one set, prepared by inducing stress by varying temperatures (and varying aggregation) and (2) a second set by adding a surfactant (with low or no aggregation) prepared by diluting the protein samples with a surfactant (SDS driven). The first set of samples will have contributions from agglomerated and unfolded samples, while the second one will solely have variations originating due to folding. Recent advances in biopharmaceuticals have led to an increase in the ubiquity of high-concentration protein formulations. A recent publication by Gervasi et al.(V. Gervasi et al., 2018) outlines the current state of monoclonal antibody (Mab) parenteral formulations on the market. According to the metanalysis, 50% of the market's Mab parenteral formulations are formulated at 50- 200 mg/ml concentrations. The high protein concentrations allow sub-cutaneous administration and thereby reduce the required volumes. Hence, this study investigated a high protein concentration of 100 mg/ml.

3.4. Materials and methods

3.4.1. Sample preparation

Bovine serum Albumin (BSA) supplied by Sigma-Aldrich, Ireland was used to prepare a stock solution (20% w/v) using sodium phosphate buffer (pH 7.2) at 100 mM and 40 mM NaCl and filtered under aseptic conditions using a 0.2 µm syringe filter. Samples were individually diluted with sodium phosphate buffer to a 100 mg/ml protein concentration. A final volume of 2 ml was aliquoted into 10ml glass vials (FIOLAX® Clear neutral glass Type I Class B, Schott, Germany) under aseptic conditions. Vials were fully stoppered and crimped (20mm grey stoppers, Adelphi Healthcare Packaging, UK).

3.4.2. Temperature driven aggregation

The samples were incubated at 75°C for increasing periods of time (1, 2, 4 hours) to induce aggregation. Incubation time was selected to maximise aggregation while minimising unfolding and chemical degradation of the protein (1). Following incubation, samples were cooled under ambient conditions then stored at 6°C prior to analysis.

Table 3.1 Temperature-driven aggregation samples used in the study

Sample	*Temperature (˚C)*	*BSA concentration (mg/ml)*	*Incubation time (hr)*	*Number of samples*
BSA.ref	NA	100	0	3x3
1hr	75	100	1	3x3
2hr	75	100	2	3x3
4hr	75	100	4	3x3

3.4.3. Surfactant-driven unfolding

A 690 mM sodium dodecyl-sulphate (SDS) stock solution was prepared. The SDS stock solution was added to the protein vials during preparation and diluted with phosphate buffer

to the concentrations of 25, 50 and 75 mM. Samples were incubated at 6°C for 12 hours prior to all measurements.

Table 3.2 Chemically driven unfolding samples used in the study

Sample	*SDS Concentration (mM)*	*BSA concentration (mg/ml)*	*Incubation time (hr)*	*Number of samples*
BSA ref	0	100	12	3x3
S25	25	100	12	3x3
S50	50	100	12	3x3
S75	75	100	12	3x3

3.4.4. Size exclusion chromatography

Samples were diluted 100-fold prior to analysis. Analysis was performed using a HPLC (Agilent Technologies 1260 Infinity) equipped with a UPLC detector. The column was a Biosep-S-3000. Samples were eluted using sodium phosphate 100 mM at pH 7.2 and flow rate 1.0 ml/min. Injection volume was set to 100 ml and the total run time was 15 minutes. Signal was detected at 280 nm. All samples were measured in triplicate.

3.4.5. Circular dichroism

Samples were diluted 100-fold (to approx. 1 mg/ml BSA) in 100 mM phosphate buffer (pH 7.2). Optical activity was measured with a Chirascan-plus spectrophotometer using a 0.1 mm pathlength quartz cuvette. Spectra were recorded in triplicate from 260 nm to 190 nm in 0.5 nm intervals and a 1.0 nm bandwidth. Protein conformational information was quantified using circular dichroism neural networking (CDNN) software to assess the degree of unfolding.

3.4.6. Fluorescence polarisation measurements

Samples were measured using the bespoke apparatus (Figure 3.2), with all measurements performed in-vial without breaking the crimped seal. The details of the fluorescence emission used for detecting denaturation in this study is that which is discussed in detail in Chapter 2. Additional details on the fluorescence emission including in vial fluorescence response and lifetime of the emission used in the study are given in the Appendix section S3. The excitation source is a diode with an emission wavelength of 365 nm, and the emission for BSA is obtained at ≈460 nm. The excitation beam passes through the sides of the vial, and collection of the fluorescence is via the bottom of the vial to reduce the loss of light induced by the curvature of the vial. The fluorescence spectra were generated by keeping the initial polariser constant at 0° polarisation (polariser in Figure 3.2) and the second polariser (shown as an analyser in Figure 3.2) in two orthogonal polarisation directions (0° and 90° polarisation directions). It must be noted that the collection plane is perpendicular to the excitation plane; hence, the polariser and analyser are orthogonal to each other. This would mean that the analyser on the collection side is to be kept at 90° to the angle of the polariser for them to have similar polarisation directions. Hereafter, the configuration with polariser at 0° and analyser at 90° is considered as the horizontal polarisation direction on the detection side, and the other configuration (polariser and analyser at 0° polarisation directions) is considered to be the vertical polarisation direction. A description of the optical geometry of the setup is explained in the Appendix section S3.2. The polarisers are chosen according to their transmission in the UV range. A Glan-Laser alpha-BBO polarizer with maximum transmission between 210-450 nm (GLB10-UV, Thorlabs, MgF_2 SLAR coating) was used as a polariser. Another Glan-Laser alpha-

BBO polariser with maximum transmission at 405 nm (GLB10-405, Thorlabs) was used as an analyser on the collection side. The fluorescence was collected using an Ocean Insight QE pro spectrometer via an optical fibre (UM 22-600 custom patch cable, Thorlabs). Following their collection, the difference between the area under the curve for each spectrum and its oppositely polarised partner was calculated. The polariser transmission spectra, the vial transmission spectrum and the absorption spectrum of BSA are shown in Figure 3.3. The reference measurement of an empty vial at horizontal and vertical polarisation directions is also shown. The fluorescence obtained from 10% BSA solution in water is also included as a reference.

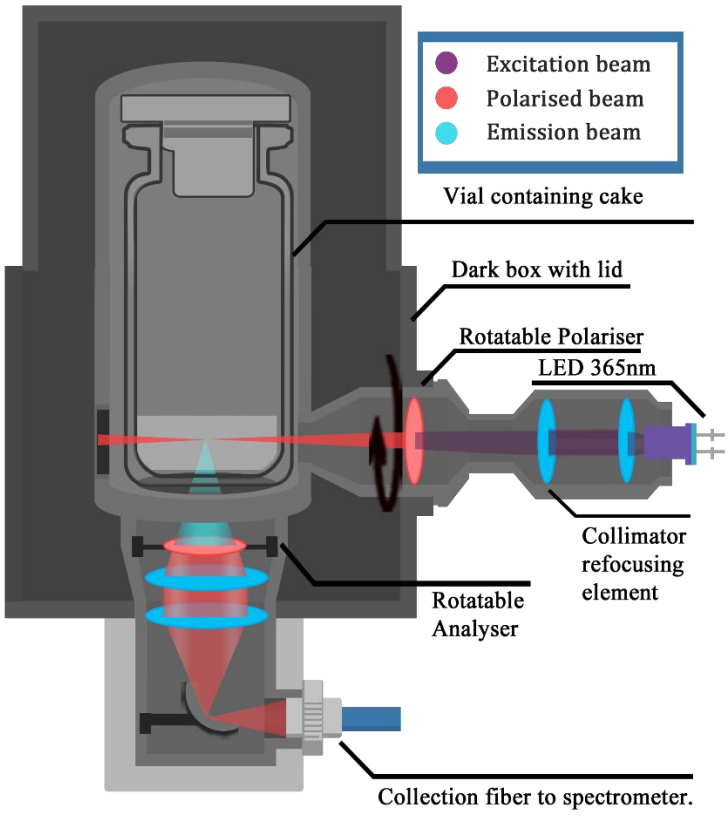


Figure 3.2 A simple schematic of the bespoke setup used for measurement: Light from the 365 nm UV LED is focused onto the sample via a collimator-refocusing unit. The light passing through the polariser produces fluorescence isotopically and the fluorescence is detected via the bottom of the vial through an analyser. The light collected is fed to the spectrometer via a fibre.

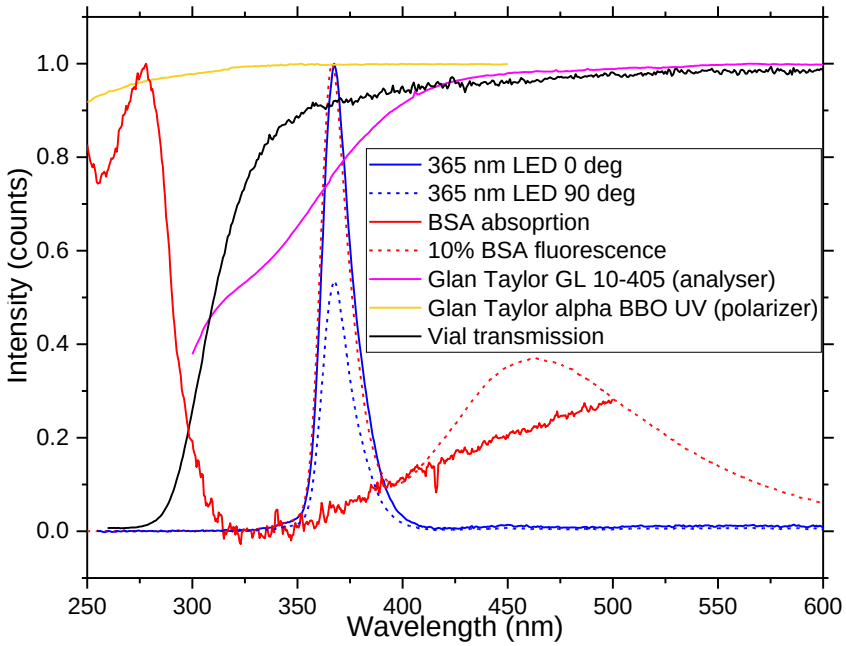


Figure 3.3 Transmission spectra of the closed quartz vial, polariser, analyser, and the emission spectrum of the source LED at 365. The LED spectra at orthogonal polarisation directions shown are obtained using the setup by measuring the empty vial and recording the corresponding reflection and Rayleigh scattering. The absorption spectra and emission spectra upon excitation at 365 nm for BSA are also indicated in dotted lines. All spectra are normalized or scaled to enable a comparison between them. Further details are provided in the Appendix section S3.1 regarding the choice of excitation wavelength.

The possible LED sources that can be used in the deep UV are shown in the Appendix 0section S3.1, explaining the choice of excitation wavelength. Lower laser drive currents in the range of 30-40 mA were used in the study to eliminate any possible non-linear behaviour from the closed vial. The vial transmission at 365 nm for different power for diode drive currents is shown in the Appendix section S3.1. Exposure to deep UV radiation can induce denaturation in proteins; hence, the denaturant effect of the UV-LED excitation source was also characterized for the same power levels using a standard photodiode power sensor (Thorlabs S120 VC). The characterization proved that minimal effect is produced upon UV exposure for the power levels used in the study. The details of the characterization are also provided in the Appendix section S3.3.

The degree of anisotropy resulting from the depolarisation of the emitted light is described by fluorescence anisotropy, *r*, and is given by Eq. 3.1.

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} = \frac{\Delta F}{I_{\parallel} + 2I_{\perp}} \qquad \textit{Eq. 3.1}$$

Where $I_{\parallel}$ is the emission intensity in the direction of polarisation of the excitation light, and $I_{\perp}$ is the emission intensity in a perpendicular direction. ΔF is the difference in fluorescence intensities in orthogonal directions. The denominator is proportional to the total intensity and is used for normalization.

3.5. Results and discussion

The experiments were designed to demonstrate the capability of our in-vial intrinsic fluorescence polarisation set up as a standard analytical technique for protein stability analysis. Variations in the fluorescence of replicate samples can result due to slight variations in individual compositions or pH. Considering this variability, the shift in fluorescence for the individual samples used in the study was characterized and was found to be minimal. Still, this factor is accounted for, and the area under the curve is considered a measure of fluorescence intensities rather than the intensity value at the emission maximum. A comparison between the fluorescence spectra of individual samples is shown in Figure 3.4 to demonstrate the minimal shift in fluorescence at different levels of temperature driven and SDS driven denaturation.

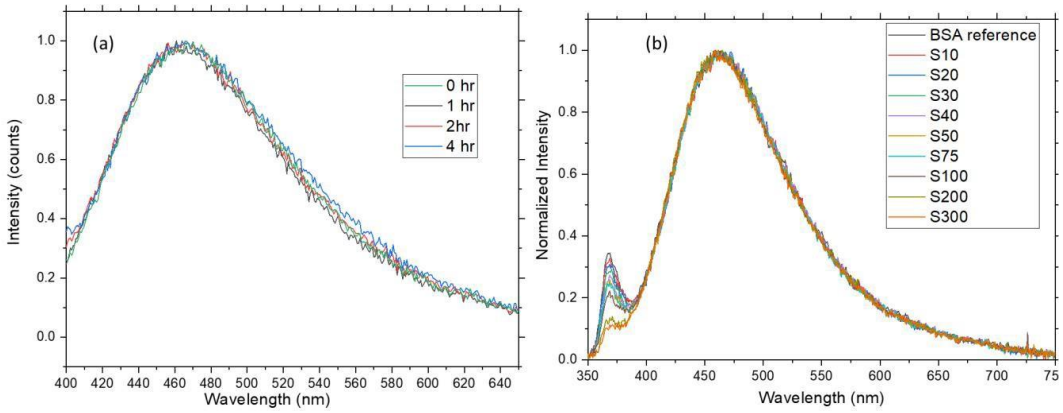


Figure 3.4 Normalized fluorescence spectrum between samples of different levels of denaturation: (a) Thermally denatured samples (b) Chemically denatured samples. Samples labelled S refer to the SDS concentration in mM. It can be observed that even though the samples are in different vials, it is attempted to keep them chemically consistent, which is evident from the minimal fluorescence shift between the samples.

Figure 3.5 (a) shows the fluorescence spectra plotted for different concentrations of BSA using the closed vial setup. Figure 3.5 (b) shows the intensity at maximum emission (λ_{emi}= 462 nm) plotted against the concentration. At concentrations above 80 mg/ml, there is significant fluorescence reabsorption from the sample, which causes the detected intensity to drop. At lower concentrations, the source absorbed is too low, leading to the detection of higher Rayleigh scattered light from the source.

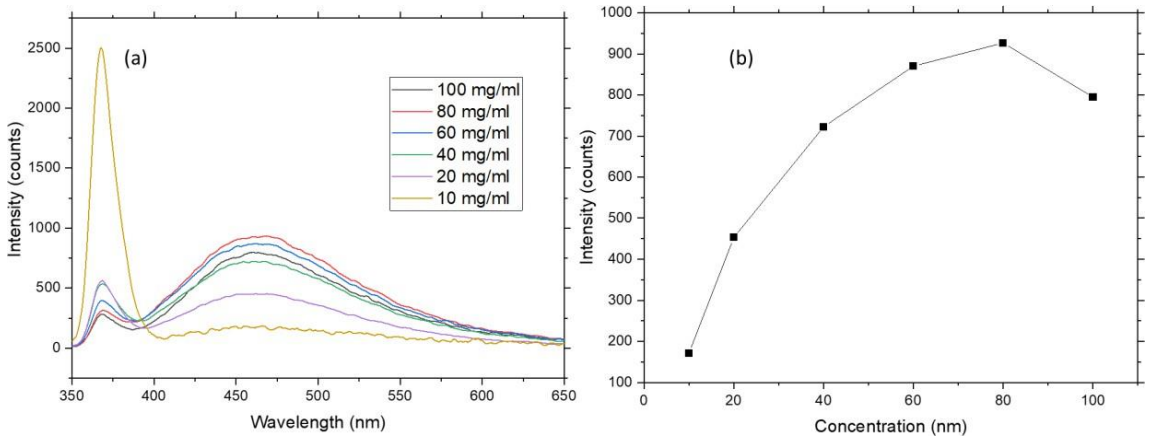


Figure 3.5 (a) Fluorescence spectra at different concentrations of BSA (b) Change in fluorescence intensity with concentration: At higher concentrations, the reabsorption from BSA causes the intensity to drop.

3.5.1. Thermally stressed samples

Figure 3.6 shows the raw data fluorescence spectra of thermally stressed samples at 100 mg/ml concentration of BSA mentioned in Table 3.1. Only one set of samples is shown in the plot for ease of demonstration. Figure 3.6(a) shows the reference measurement for the same concentration as the denatured samples. If the samples had the same denaturation level, then the emission intensities in horizontal and vertical polarisations for them would have been the same as in the case of the BSA reference sample. The spectra have two contributions in information: (1) the Rayleigh scattering source peak denoting agglomeration and (2) the total fluorescence intensity denoting denaturation. The scattering part is as expected as the reference scattering curves in Figure 3.3- the analyser transmits only half of the Rayleigh scattered light's intensity when placed orthogonally to the polariser.

If the molecules are small and freely rotating in solution, then the fluorescence generated is isotropic, and hence the same intensity is detected in both polarisation directions. On the contrary, if a sample with the same concentration has undergone thermal denaturation (Figure 3.6(b)), the agglomeration resulting from the thermally induced stress will increase the Rayleigh scattering. The aggregation can occur through multiple mechanisms, unfolding intermediates, or linkages between protein moieties. All aggregation reactions are accelerated by excess kinetic energies imparted by the increased temperatures(Wang et al., 2010). The increase in size caused by the protein unfolding (typically between 50-100 nm for BSA) (Park et al., 2018) can also cause an increase in Rayleigh scattering intensity detected, but this would be significantly less compared to the contribution from agglomeration. The fluorescence detected in the vertical direction (solid line) for different

denaturation levels shows a significant difference if the thermal stress induced is higher. In Figure 3.6(b), a higher difference in fluorescence intensity is observed between the orthogonal polarization directions as the time the thermal stress is induced increases. All the measurements are done on triplicate samples (n=3), and the repeatability of the results is ensured.

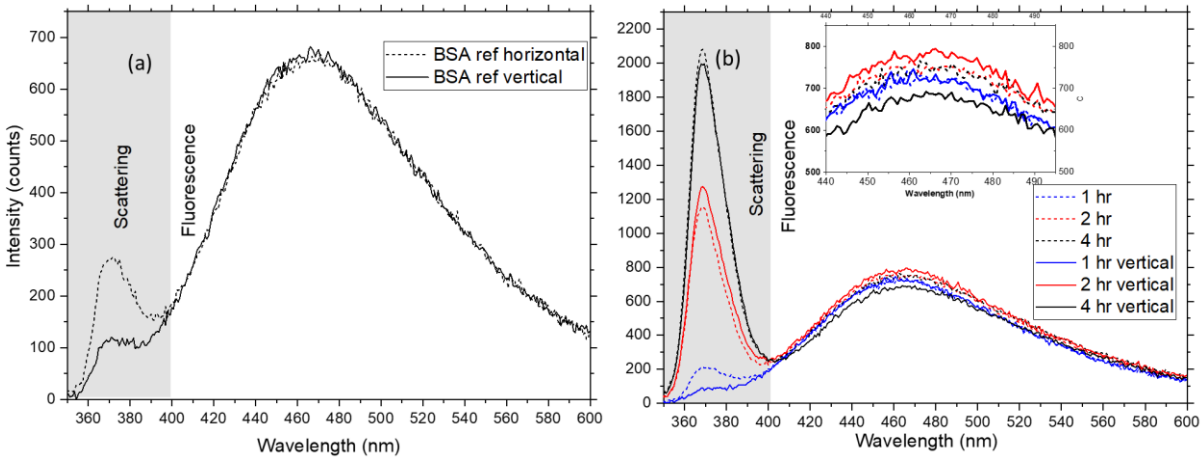


Figure 3.6 Raw data (a) Reference spectrum measured for the sample set 1 (b) Fluorescence spectra for orthogonal polarisation directions for the thermally stressed sample set 1. The dotted lines are horizontal polarization, and the solid lines are vertical polarisation directions. The inset shows the same intensities zoomed. All sample sets are at 100 mg/ml concentration. The samples are kept at 75 C for periods of 1 hrs to 4 hrs.

The size exclusion chromatography (SEC) results shown indicate the formation of high molecular weight aggregates at 2-4 hours of thermal stress 75 ˚C for a 100 mg/ml BSA solution. The scattering obtained after calculating the area under the curve for both vertical and horizontal directions for all samples are shown in Figure 3.7. Both results follow the previously reported trends(Canet et al., 2001; Grigolato and Arosio, 2019).

Circular dichroism (CD) measurements of the same samples confirm the observations on the fluorescence side of the spectrum, Figure 3.8. The CD spectra obtained after the analysis of 100 mg/ml BSA samples indicate a change in molar ellipticity for all conformationally relevant peaks (222 nm, 209 nm, and 189 nm). This may be explained by

the increase in random coil percentages in the samples(Kelly et al., 2005). Secondary structure analysis show an initial increase of alpha-helical content for samples stressed at 75°C for 1 hour. However, after 2 hours, there is a recovery in α-content with a slight increase in random coil percentages, Figure 3.8(b).

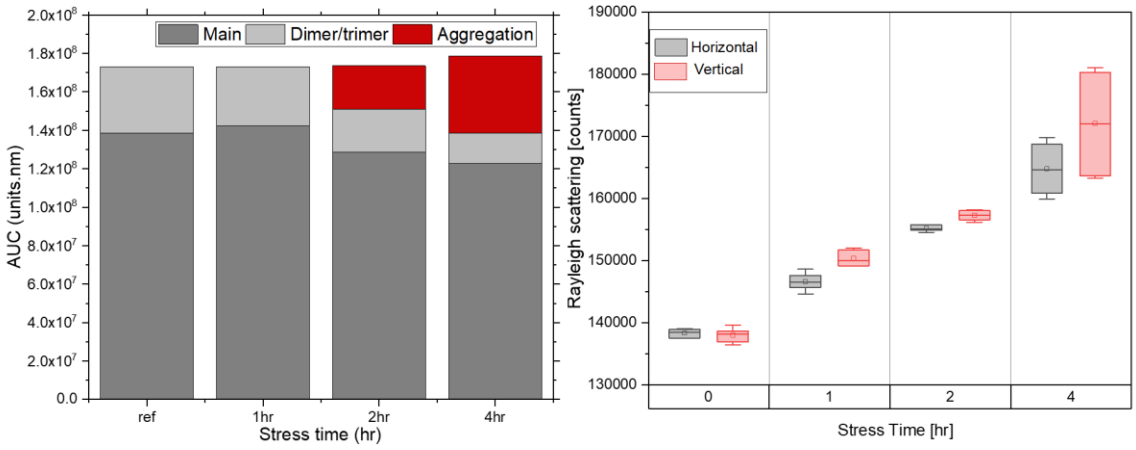


Figure 3.7 (a) Size exclusion chromatography peak areas for 100 mg/ml BSA under different temperature-induced stress levels. Each sample is kept at 75 ˚C for periods of 1, 2 and 4 hours. The decrease in monomer peak area is accompanied by the appearance and increase of aggregation peak that elutes slightly earlier than the main peak conglomerate (b) area under the curve for heat (75C) stressed samples, indicating an increase in the aggregation. As the sample aggregates the particle size increases and thus the Rayleigh scattering of the excitation beam becomes more intense, leading to increased detection at 365 nm. All the measurements are done in triplicates (n=3).

The difference in fluorescence emission intensity between the horizontal and vertical directions, $\Delta F = I_{\parallel} - I_{\perp}$, is proportional to the anisotropy of the sample. The emission fluorescence intensity is found by integrating the area under the curve for each fluorescence emission after normalization to the Rayleigh scattering peak. ΔF obtained is compared to the CD secondary structural analysis, and a good correlation can be observed between both. The trend in CD is a sharp decrease followed by recovery after 4h of stress at 75°C for the α-helix. A correlation analysis was performed for the data obtained from both CD secondary structural analysis and photoluminescence analyses. The data shown in Table 3.3 indicates a strong correlation between the photoluminescence results and both β-content and random coil values obtained via CD, as well as an inverse relationship between

the α-helix content and the fluorescence area. The statistical analysis is in accordance with the literature and suggests that the β-content correlation is most likely.

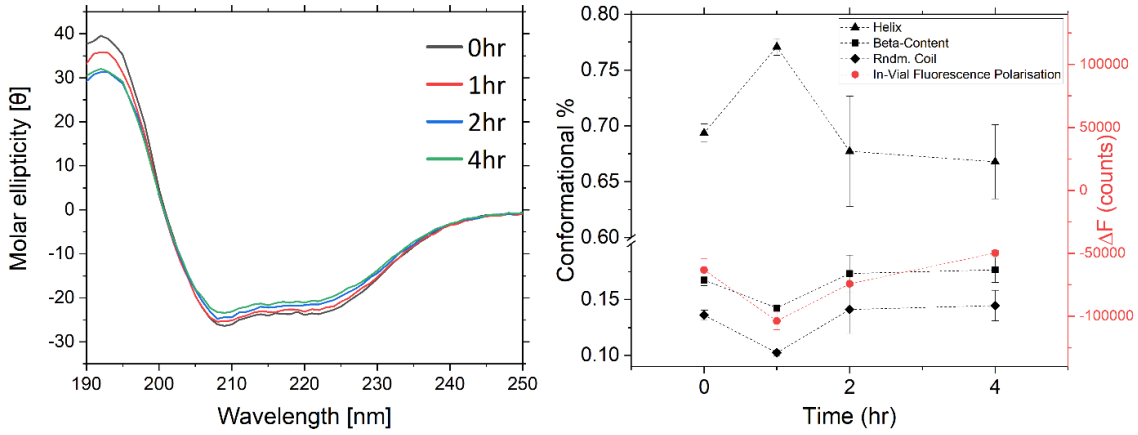


Figure 3.8 (a) Molar ellipticity values for 100 mg/ml BSA samples thermally stressed at 75 ˚C for 1, 2, and 4 hours. (b) Comparison between CDNN and fluorescence analysis. Although there is a sharp spike at 1 hr the BSA structure refolds for the remaining observation period. Fluorescence measurements show a similar trend to Beta-sheet content as measured using CD. (n=3 for all)

Table 3.3 Correlation matrix for CD secondary structural analysis of thermally stressed 100 mg/ml BSA vs. the analysis of photoluminescence technique described above.

	Helix	*Beta-Content*	*Rndm. Coil*	*Photoluminescence*
Helix	1			
Beta-Content	-0.99981	1		
Rndm. Coil	-0.99885	0.997723	1	
Photoluminescence	-0.91799	0.916233	0.921403	1

3.5.2. Chemically stressed samples

Sodium dodecyl sulphate (SDS) is an anionic surfactant capable of denaturing proteins through binding interactions with specific sites. The binding process proceeds over multiple steps, with denaturation occurring after the saturation of specific binding sites. The surfactant molecules act as traditional ligands at lower molar ratios, stabilizing the structure. Due to the functional nature of bovine serum albumin (BSA) as a plasma transporter of amphiphilic molecules, its structure includes twelve sulfonate half-ester binding sites (Otzen, 2011). After saturation, at higher molar ratios (above 1:12 for BSA),

protein unfolding proceeds. At intermediate SDS concentrations (10 mM~100 mM), micellar binding to protein monomer structures leads to a change in the Trp environment and an apparent decline in fluorescence emission (Otzen, 2002).

Figure 3.9 shows the fluorescence spectra in orthogonal polarisation directions for chemically stressed samples. Spectra for only one set of samples are shown for ease of interpretation. As for the thermally stressed samples, one would expect the fluorescence emission of the reference sample to be the same in both polarisation directions since the fluorescence emission is isotropic (Figure 3.9(a)). However, as the concentration of the surfactant increases, the fluorescence detected in the vertical polarisation direction decreases as the fluorescence preserves polarisation upon slow rotation of the proteins. A significant change in emission intensities can be observed between the orthogonal polarization directions if the concentration of the surfactant is increased further (plots for surfactant concentration up to 300 mM are shown in the Appendix section S3.4).

The lower concentrations of the surfactant (up to 75 mM of SDS) are used for the data analysis to check if the technique has enough sensitivity for comparison with CD. As can be seen in Figure 3.9(b), differences in intensities in orthogonal directions can be observed in the raw data. Even though the change in intensity seems minimal from the raw data, it is enough for comparison with CD if the area under the curve is integrated (Figure 3.10). The agglomeration, in this case, seems to be minimal since the difference in intensity in scattering seems linear compared to the reference.

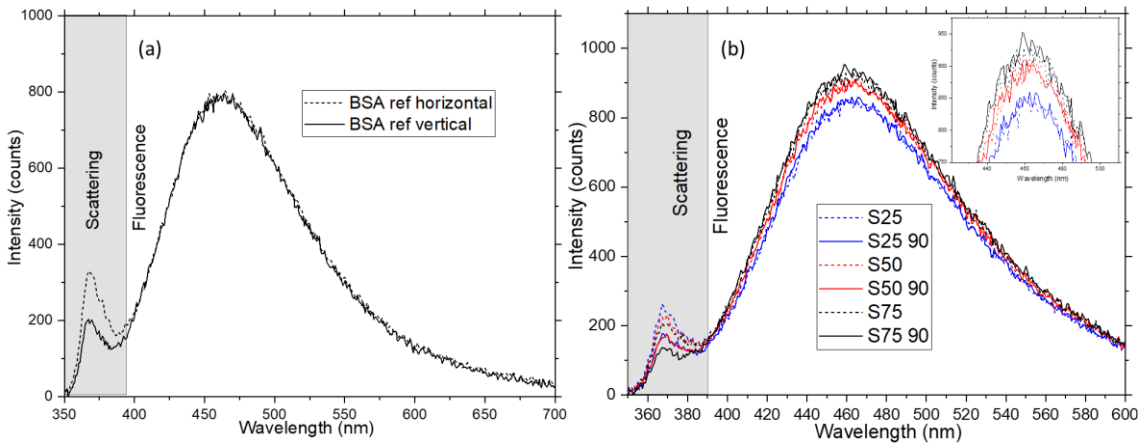


Figure 3.9 Raw data (a) Reference spectrum measured for the sample set 1 (b) Fluorescence spectra for orthogonal polarisation directions for the chemically stressed sample set 1. The dotted lines are horizontal polarization, and the solid lines are vertical polarisation directions. The inset shows the same intensities zoomed. All sample sets are at 100 mg/ml concentration.

Studies by (Xu and Keiderling, 2004) indicate that secondary structure is initially lost upon adding SDS at concentrations below the critical micellar concentrations (CMC). The electrostatic interactions caused by the amphiphilic moieties are the leading cause for unfolding. Upon reaching CMC, any further addition of SDS (up to a maximum of 20 mM SDS) does not cause additional unfolding but instead leads to the coiling of the protein in and around the micellar structure. Another study by Winogradoff et al. indicates that the interaction of the protein with SDS micelles takes shape in the form of a necklace-and-beads structure that alters the protein conformation. Further attachment of SDS micelles leads to global structural changes. The change in pH will also bring changes in fluorescence emission, but since all the samples were ensured to have the same pH (7.2), this variable can be neglected (Winogradoff et al., 2020).

All the reports stated above indicate many possible unfolded states upon adding SDS for concentrations above 25 mM. A large number of possible unfolded states(Otzen et al., 1999) make it impossible to determine the relationship between unfolding and SDS concentration quantitatively. However, CD secondary structural analysis reveal a

qualitative relationship between the protein conformational states detected on both instruments (Figure 3.10). The changes in structure do not occur consistently due to the numerous mechanisms and sites of attack on the protein. Since the fluorescence anisotropy induced and the difference in fluorescence detected is also a collective phenomenon, it would indicate only an increase or decrease in denaturation. This is satisfactory output for a handheld instrument that identifies if denaturation has occurs.

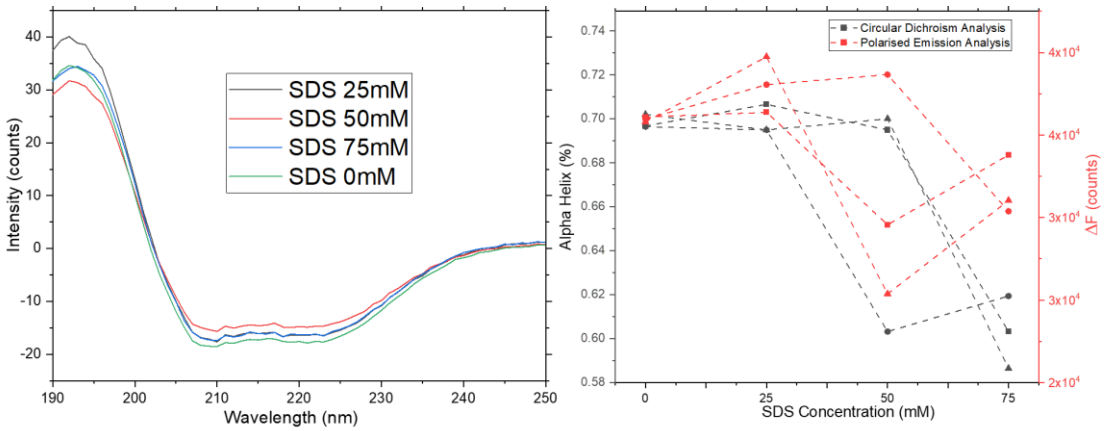


Figure 3.10 (a) Molar ellipticity values for 100 mg/ml BSA samples chemically stressed at 25, 50 and 100 mM SDS concentrations. (b) Comparison between CDNN and fluorescence: The primary structure of BSA 100 mg/ml samples as obtained CDNN vs the conformational changes detected by fluorescence polarisation. SDS unfolding proceeds via electrostatic and hydrophobic interactions on multiple states on the protein structure, leading to various unfolded states (n=3 for all).

3.5.3. Changes in polarisation

The changes in the polarization of the fluorescence due to rotational diffusion are due to the rotation of the BSA molecule in general and the rotation of the chromophore (tryptophan) relative to its surroundings owing to dipole-orientational relaxation of the chromophore after excitation. The contribution from the rotation of the domain containing the tryptophan group with respect to the nearby surroundings is considered to be negligible(Vlasova and Saletsky, 2009). Figure 3.11 shows *r*, the degree of polarisation

calculated from the intensities in orthogonal directions for thermally and chemically stressed samples.

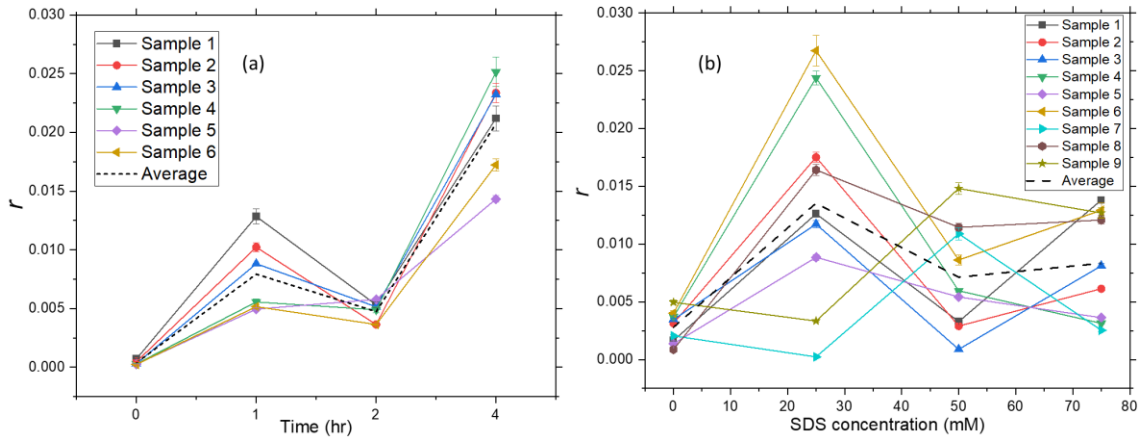


Figure 3.11 Degree of anisotropy measured for (a) thermally stressed samples and (b) chemically stressed samples. The data shown for (a) is for 6 sample sets, and (b) is 9 sample sets. The dotted black curve shows the average response from all the samples. The error bars show the variations occurring for the repetition of the measurements for individual samples.

For thermally treated samples, the *r* values indicate the average response from both the beta and α-helical content in the sample. A better match of *r* value trend is with the β-content, indicative of the presence of higher β-content with time. The increase in anisotropy for chemically treated samples before their descent is much higher than for thermally treated samples. This apparent increase in SDS-treated samples is most likely due to the unfolding intermediates causing an aggregation chain reaction, a phenomenon well discussed in the literature(Otzen et al., 1999; Winogradoff et al., 2020; Xu and Keiderling, 2004). The unfolding process by SDS-treatment proceeds stochastically leading to the variability seen in the results (Grigolato and Arosio, 2019). The pattern of evolution of tryptophan fluorescence with increasing SDS concentration differs below and above the isoelectric point, pI, of BSA (pI= 4.9 for BSA)(Kopac et al., 2008; Yohannes et al., 2010). The pH for the samples chosen in this study is above the pI value (pH=7.5). Studies indicate that the anisotropy *r* for pH values higher than pI would increase to a maximum value initially after which it drops down and reach a value after which further addition of SDS would not have

an affect indicating a complete denaturation of BSA molecules. The same trend can be observed for the *r* values obtained using the in-vial fluorescence anisotropy measurements.

3.6. Conclusion

The intrinsic fluorescence of a protein has been used to study the denaturation of an in-vial sample caused by two different methods: (1) thermal and (2) chemical. The method designed, discussed, and used in this research has shown to identify key stability indicators for in-vial biopharmaceutical samples. The protein's denaturation level can be monitored by analysing fluorescence intensities at orthogonal polarization directions. In addition, the protein aggregation can be monitored by observing the scattering effects on the spectrum. The wavelength and intensity for excitation are chosen at which both the vial and sample would have a minimal unwanted effect like non-linear behaviour or UV denaturation. The results demonstrate significant differences in the measurements obtained from the reference BSA solutions after preparation and the denatured samples.

The non-invasive technique can be used to further develop handheld devices which can provide an approach to monitor biopharmaceutical product stability from the point of manufacture to patient administration. The technique is customisable according to the user-which means in addition to handheld systems, the same technique can be used to develop in-line systems in pharma and biopharma industrial settings for rapid, non-destructive, 100% quality control inspection of product release, and providing a product quality overview during stability screening.

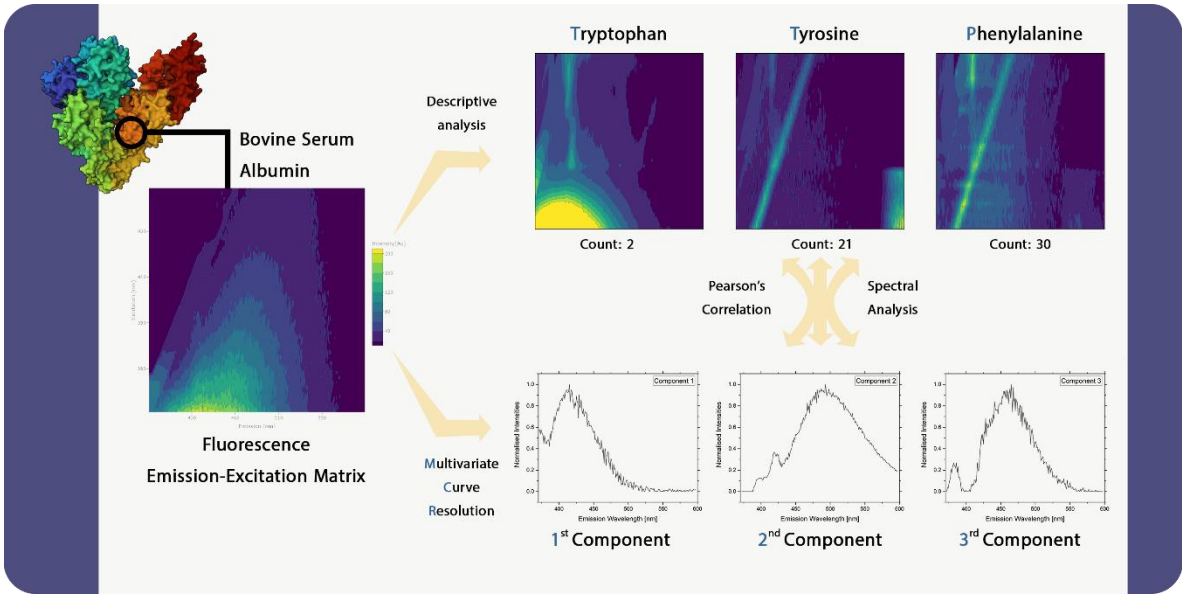
Chapter 4
Explorations of Intrinsic Protein Fluorescence using MCR-ALS

4. Chapter 4: Explorations of Intrinsic Protein Fluorescence using MCR-ALS

4.1. Highlights

- By analysing a 3D emission excitation matrix for 5 proteins we described the fluorescence due to excitations in the region of 365 nm ~ 400 nm.
- Using multivariate curve resolution we were able to relate protein fluorescence spectra with those of the emitting amino acids.
- Multivariate curve resolution estimated concentrations correlated with the relative abundance of amino acid fluorophores in the protein structure.

4.2. Graphical Abstract



4.3. Introduction

Fluorescence analysis of protein solutions provides valuable insights that can be leveraged in practical applications for manufacturing, transport, and storage. The three aromatic amino acids tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe) abundant in protein structures act as fluorophores which emit intense light upon exposure to UV and visible light (Lakowicz, 1999). The information contained within the fluorescent emissions is an indicator of the amino acid's electromagnetic environment, which can affect protein folding and stability (dos Santos Rodrigues et al., 2023; Hellmann and Schneider, 2019; Michalet et al., 2006). It is of interest to monitor proteins without disturbing the solution, ideally within a sealed glass vial to preserve purity, sterility, and stability (Atkinson et al., 2002). Previous protein fluorescence emission studies focused primarily on UV excitation at 270-280 nm (Klijn and Hubbuch, 2021) barring some recent exceptions (Wang et al., 2019) including previous work by the authors (Chullipalliyalil et al., 2023; ElKassas et al., 2021). Excitation in the absorption tail of the protein's fluorophores, around 365 nm, is more suitable for in-vial analysis as conventional glass vials transmit light at this wavelength. Although these excitations have only recently been explored for potential applications, further investigation is needed to fully understand the photophysics involved. Protein intrinsic fluorescence is primarily a result of a mixture of emissions of the three aromatic amino acids Trp, Tyr, and Phe present within their structures (Lakowicz, 1999). The process of separating component spectra from a total sum spectrum is termed deconvolution and is common practice for spectral techniques where band overlapping is unavoidable (Elliott, 1987). Mathematical approaches for signal separation have been essential in developing analytical capabilities (Tooke, 1988; Tooke and James, 1985). These approaches include but are not limited to derivatives, linear deconvolution, and

Fourier deconvolution (Jansson, 1984). More recent developments have also borne significant advancements in mathematical signal processing, with the advent and popularity of multivariate data analysis (MVA) techniques (Bayat et al., 2020).

Among MVA methods, multivariate curve resolution (MCR) is a family of methods developed for the exploration of complex spectral profiles, mixture separation, and concentration (contribution) calculations for overlapping spectral features (De Juan et al., 2014). Analytical measurements, especially those based on spectroscopic methods, are well-suited for analysis by MCR. This is because the underlying analytical model is a bilinear model of pure signal contributions. As a result, MCR is a natural fit for fluorescence measurements since it accurately mimics the structure of the analytical measurement (Olivieri and Escandar, 2014; Parastar and Tauler, 2014).

Joan et al. compared various MCR algorithms and discussed the advantages of using the Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) algorithm. MCR-ALS is an iterative algorithm that optimizes the model by applying constraints at each cycle. These constraints can be categorized as (1) mathematical conditions, (2) natural constraints, and (3) process constraints (De Juan et al., 2014).

Excitation-emission matrices (EEMs) may allow us to describe general features in the intrinsic fluorescence of the studied proteins (Malik et al., 2015). Fluorescence EEMs also fit the paradigm of a bilinear data matrix, and therefore are ideal for use with MCR (Olivieri and Escandar, 2014). In this study, we used EEMs measuring the intrinsic fluorescence of the three amino acids (Trp, Tyr, and Phe) and five model proteins to further understand the emissions presenting at 450 nm upon excitation at 365 nm. Our basic hypothesis was that the deconvolution of the protein EEMs using MCR-ALS analysis would identify

component spectra loadings and contributions and that these components would relate to the fluorescence contributions of the individual aromatic amino acids.

4.4. Materials and methods

4.4.1. Sample preparation

All protein and amino acid raw materials were supplied by Sigma-Aldrich, Ireland. Bovine serum albumin (BSA), pepsin, trypsin, and lysozyme solutions were prepared at 1% w/v using 10 mM sodium phosphate buffer pH 6.7. Insulin solution 1% w/v was prepared using 0.01M sodium hydroxide. All protein solutions were filtered under aseptic conditions using a 0.2 µm syringe filter. Samples were individually diluted with sodium phosphate buffer to a 1 mg/ml protein concentration. Trp, Tyr, Phe samples were prepared at 1 mg/ml (0.1% w/v) using deionised water with a final pH of 6.8. The molecular weights and amino acid counts for the chosen proteins is shown in Table 4.1 below.

Table 4.1 Molecular weights and aromatic amino acid residue counts for proteins studied. Values were obtained via the UniProt Database (UniProt, 2023)

	Mol.Wt (KDA)	AA residues	Trp	Tyr	Phe
BSA	66	607	3	21	30
Pepsin	41	385	6	17	16
Trypsin	25	246	4	10	6
Lysozyme	16.2	147	6	3	4
Insulin	6	53	0	7	3

4.4.2. Spectroscopy

4.4.2.1. Excitation emission matrices

Fluorescence EEMs were captured using an Agilent Cary Eclipse Microplate Reader (CA, United States). The excitation range was 250 nm – 450 nm with 2 nm increments and a 5 nm excitation slit. Emission was captured from 250 nm – 800 nm at 1600 nm/min with a 2 nm data interval. Solutions of the 5 proteins (BSA, lysozyme, pepsin, trypsin, insulin) were measured in triplicate resulting in a total of 15 EEMs. The amino acid samples were similarly measured in triplicate resulting in a total of 9 EEMs.

4.4.2.2. Data analysis

EEMs were processed using Rstudio to remove first- and second-order scattering effects using a nearest-neighbour interpolation (interpolated data was excluded from multivariate analysis). Spectra were not normalised or smoothed in any way to preserve the maximum amount of information.

MCR is a technique that breaks down the data matrix (X) into three components: concentration profiles (C), spectra profiles (ST), and residuals (E). The data matrix contains one spectrum per line from the specific protein EEM. The concentration profiles represent the variation among the spectra, while the spectra profiles represent the variation among

the emission intensities at each excitation wavelength. The residuals are the part of the data that is not explained by the model. The relationship between the model components can be represented in the following equation Eq. 4.1:

$$X = C \cdot ST + E \qquad \textit{Eq. 4.1}$$

By using MCR, it is possible to extract the spectra fingerprint of each component in a mixture. In this study, the individual components of interest are the individual amino acid fluorophores which generate the overall protein spectral profile.

MCR-ALS analysis was conducted using Rstudio (RStudio,PBC., MA, USA) and the mdatools package. The MCR-ALS algorithm and functions of mdatools are discussed in detail by the author in their original publication (Kucheryavskiy, 2020). The algorithm was used with a constraint of non-negativity applied according to the fast combinatorial non-negative least squares solver (FC-NNLS)(Van Benthem and Keenan, 2004) and spectral estimates from the MCR-purity function were used (Jaumot et al., 2005). The number of MCR components was set to three to match the number of aromatic amino acids expected to contribute to the overall protein fluorescence. The choice of matching the number of MCR components to the number of fluorophores was made to facilitate the simplest physical interpretation of the results (Motegi et al., 2015).

Values of purity (concentration) and spectral loadings (component spectra) were obtained for different proteins (BSA, pepsin, lysozyme, trypsin) via MCR-ALS modelling of their EEMs. Pearson correlation analysis was performed between the MCR loadings spectrum of the four proteins measured in triplicate (n=12) and the average spectrum of the reference amino acid. Insulin was excluded from the correlation analysis due to its lack of a Tr p component and possessing only 2 fluorophores. Spectral loadings and the reference

spectrum for each amino acid were compared to determine which emissions resulted from each of the three fluorophores. Spectral analysis was also performed to support the statistical analysis. All correlations were considered significant at an α of 0.01.

To confirm the analysis, MCR concentration (contributions) values were compared to the fluorophore counts using a Pearson correlation. For the four protein solutions studied in triplicate, emission spectra at 3 wavelengths nearest the excitation maxima (362 nm, 364 nm, 366 nm) were selected for component analysis (n = 36). The numbers of fluorophores in each protein were compared to a matrix of fluorophore frequencies in the proteins. All correlations were considered significant at an α of 0.01.

4.5. Results and discussion

4.5.1. EEMs for fluorophores and proteins

The condensed nature of the information captured in an EEM and the multitude of variables that can affect protein intrinsic fluorescence prevent us from drawing direct relationships between the amino acids and the spectrophotometric features in the proteins. Therefore, a qualitative and descriptive approach to show the relationships between contributing amino acid fluorophores and the resulting protein EEM was taken for this section. The 3-dimensional EEMs captured for solutions of the three aromatic amino acids responsible for the intrinsic fluorescence in proteins are shown in Figure 4.1.

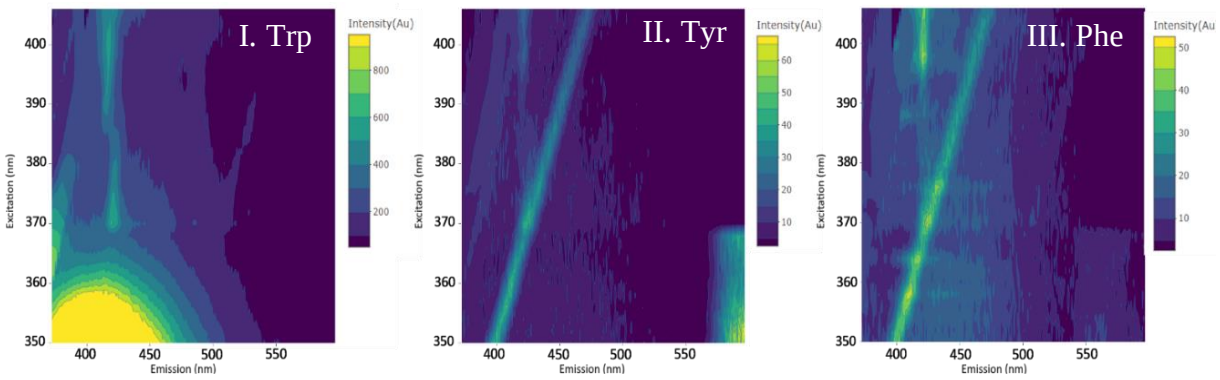


Figure 4.1 Individual amino acid emission excitation matrices for tryptophan, tyrosine, and phenylalanine at 370 nm – 450 nm excitation/ 365 nm – 600 nm emissions. Note intensity scales are different for each protein spectra due to differences in maximum intensity. The scales are coloured according to emission intensity as a percentage of maximum emissions with navy for the lowest intensities and yellow for the highest.

Figure 4.2 shows the EEMs for the five proteins used in this study (BSA, pepsin, lysozyme, trypsin, insulin). By comparing EEMs we are able to demonstrate the changes that can occur to the fluorophore spectrum when translated into the protein structure.

Due to differences in maximum emissions, it is important to note the differing scales of the z-axis for the plots shown in Figure 4.1 and Figure 4.2. Trp emissions are by far the most intense with the most striking characteristic of Trp emission being the significant peak around 340 nm - 360 nm excitation/400 nm – 450 nm emissions, which fades gradually up to excitations of 370 nm. The overwhelming majority of the emissions seen in BSA and pepsin emissions can be attributed to Trp (Agro' et al., 1974) with Tyr and Phe less evident. Conversely, despite the abundance of Trp moieties in both Lysozyme and Trypsin, the main Trp 350 nm - 360 nm excitation/ 370 nm – 450 nm emission is notably not evident. Minor features of Trp emissions can be seen in the EEMs of lysozyme and trypsin between 370 nm - 410 nm excitations/420 nm emissions. Comparatively, the 370 nm - 410 nm excitations/420 nm emissions peak is completely absent in the insulin EEM, which is expected due to insulin's lack of Trp moieties. The relative intensity reduction in Trp

emissions as seen in Figure 4.2 between lysozyme and trypsin when compared to BSA and pepsin is most likely due to the inaccessibility of the Trp moieties or their otherwise engagement in bonding, reducing their emissions (Burstein et al., 1973; Formoso and Forster, 1975). The relative molecular weights of the proteins studied and the total amino acid counts inevitably play a role in the intensity of the emissions as they affect the electrochemical environment of the fluorescing amino acids and increase the amount of light reaching the detector through scattering (Wyatt, 1993).

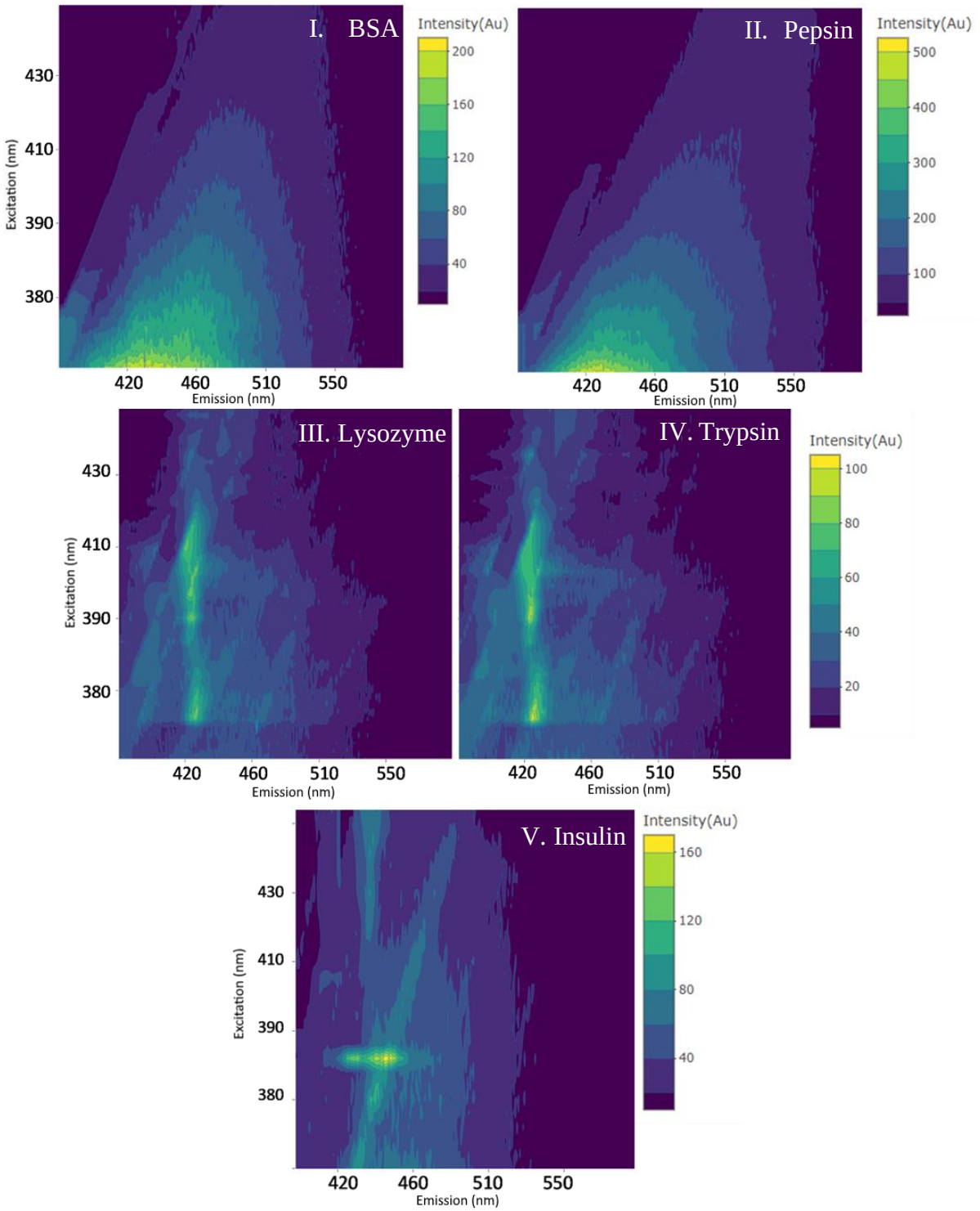


Figure 4.2 3D emission excitation matrices of 5 proteins (BSA, pepsin, lysozyme, trypsin, insulin). Emissions were captured at 10 mg/ml for all proteins at 370 nm – 450 nm excitation/ 365 nm – 600 nm emissions. Note intensity scales are different for each protein spectra due to differences in maximum intensity. The scales are coloured according to emission intensity as a percentage of maximum emissions with navy for the lowest intensities and yellow for the highest. As Lysozyme and Trypsin showed similar emissions their scales are identical.

Considering the Tyr and Phe EEMs we are able to see great similarities in the emission profiles. The characteristic features include a) the gradually red shifting peak at 400 nm – 460 nm of emissions, b) the single peak around 400 nm excitation/430 nm emission, and c) the secondary peak at 360 nm excitation/550 nm-600 nm emission. Despite the intense secondary emission of Trp at 370 nm excitation/ 420 nm emission, the faint remnants of Tyr and Phe peak between 400 nm and 460 nm can be seen in the EEMs of lysozyme and trypsin. Insulin is naturally more interesting in this regard, due to its lack of Trp residues. While the red shifting peak at 400 nm – 460 nm emission is more apparent in insulin, another wide emission of Phe can be seen between 375 nm – 380 nm excitation/420 nm – 450 nm emission. The overall emissions are higher in intensity for Tyr over Phe. However, the literature shows that Tyr is highly liable to concentration quenching and as such appears weaker than what the quantum yields would imply (Chrysochoos, 1974; Feitelson, 1964). The quenching effect is further compounded by interactions with the solvent media (in this case water) for both Tyr and Phe(McGuire and Feldman, 1973).

The emissions of aromatic fluorophores are affected by various factors such as interactions with the surrounding environment, self-quenching, and pH (Chrysochoos, 1974; Feitelson, 1964; Yang and Honig, 1993). The presence of the fluorophores in combination can also lead to interactions, like the quenching of one fluorophore by another, or in the context of the protein structure such as interactions with the protein backbone, the electrochemical interactions between different residues, or conformational bonding (Edelhoch, 1967; Feitelson, 1964). Due to these complex interactions, it is difficult to establish solid and quantitative relationships between fluorescence spectra and specific characteristics of a studied molecule. The mathematical deconvolution of the fluorescence emission spectra of

amino acids within a protein will also be limited by the changes to the aromatic amino acid fluorescence caused by changes in the electronic environment, bonding, and steric effects (refer to Chapter 1). Therefore, it was important to control for concentration quenching effects, pH, and conformational effects in protein samples to minimise non-bilinear behaviours in the fluorescent emissions (De Juan et al., 2014).

4.5.2. MCR-ALS analysis

As previously discussed MCR-ALS uses an ordinary least squares calculation to resolve the concentrations using estimates of the spectra and then using the resolved concentrations to calculate the spectral loadings (Malik et al., 2015). As such evaluating the quality of the estimates is dependent both on the loadings and scores. MCR is similar to other multivariate models (Afifi et al., 2019), with the added benefit of having physical meanings for both the loadings and the scores (De Juan et al., 2014).

MCR models were developed independently for each protein, using the fluorescence EEM to predict the fluorophore concentration in each protein. For all the models the explained variance was within the range of 96.87% ± 2.05 for three components.

4.5.2.1. Comparing fluorophore spectra to MCR spectral loadings

Comparing the average fluorescence of individual amino acid fluorophores to the loadings of each MCR component using a Pearson correlation we found that each fluorophore showed the greatest correlation with a different component for each protein model. However, in all of the models, components had a single component which had a higher degree of correlation with one or more amino acids, indicating that the amino acid fluorescence is interdependently linked (see Table 4.2). Individual correlation results for

each protein including p-values are shown in the addenda tables Supplementary Table 4.1~5. Subsequently, the Pearson correlation coefficient for the MCR loadings of component 1 showed the highest affinity to multiple amino acids (Trp and Phe)(Figure 4.3a) which also carried the highest proportion of explained variance in the models at 59.49% ± 9.81.

For the BSA model components 1 and 2, Trp and Tyr both show high correlations. In the case of component 1 Trp is the higher correlation. For component 2 both Trp and Tyr show equal correlation. Similarly, component 3 shows the highest correlation for Phe. Figure 4.3 shows the relationship between captured amino acid spectra and the MCR-resolved spectra loading for all tested proteins.

Table 4.2 Correlation coefficients for the fluorophore spectra and loadings of the MCR component for all resolved spectra. Numbers in parentheses are the explained variances for each component.

	BSA 1 (44.19%)	*BSA 2 (32.28%)*	*BSA 3 (22.72%)*
Trp	0.96	-0.49	0.27
Tyr	0.70	-0.59	0.08
Phe	0.92	-0.21	0.55
	Pepsin 1 (49.18%)	Pepsin 2 (20.42%)	Pepsin 3 (29.82%)
Trp	0.89	-0.82	0.18
Tyr	0.64	-0.79	0.03
Phe	0.99	-0.71	0.50
	Lysozyme 1 (43.71%)	Lysozyme 2 (38.23%)	Lysozyme 3 (14.75%)
Trp	0.71	0.53	0.61
Tyr	0.52	0.46	0.13
Phe	0.66	0.77	0.54
	Trypsin 1 (54.00%)	Trypsin 2 (26.00%)	Trypsin 3 (16.14%)
Trp	0.67	0.51	0.62
Tyr	0.55	0.41	0.12
Phe	0.66	0.75	0.51

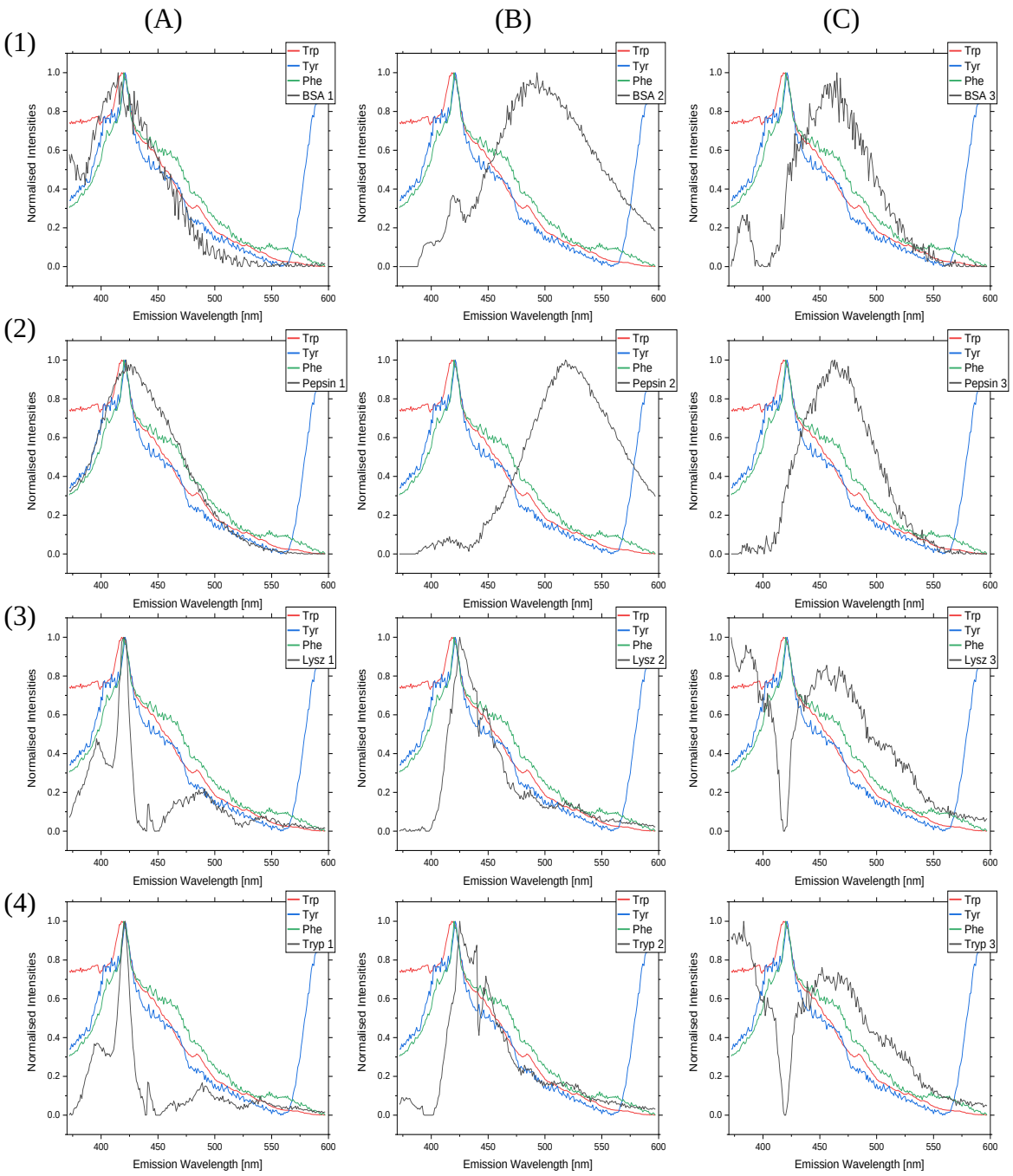


Figure 4.3 Normalised MCR components for BSA, pepsin, lysozyme and trypsin, compared to their amino acid reference spectrum. The black line represents the MCR resolved spectral loadings for each protein (1. BSA, 2.pepsin, 3.lysozyme, 4.trypsin) showing the first (a), second (b) and third (c) components. The red, blue, and green lines represent the fluorescent of 1mg/ml solutions of Trp, Tyr, and Phe respectively.

Comparing the spectra visually, the lack of Trp emission between 425 nm and 450 nm in component-1 lysozyme (Figure 4.3 (3a)) is most likely due to the quenching of tyrosine by tryptophan (further explored below). The BSA and pepsin however follow similar trends with slight lambda max shifts (4-5 nm), for all components which is possibly explained by the difference in secondary structure between the two proteins. The EEMs of BSA and pepsin in Figure 4.2 indicate a peak shift for pepsin over BSA. Database values indicate that BSA is mostly composed of α-helices (63%) in the native state, while pepsin is primarily composed of β-sheets (58%) (The UniProt Consortium, 2023).

MCR-resolved component 2 peak maxima, attributed primarily to Trp and Tyr shift dramatically (~50 nm-100 nm) from the reference Tyr spectrum for both BSA and pepsin (Figure 4.3b), possibly due to the relative weakness of the Tyr signal in the presence of Trp (Edelhoch, 1967). The peak shifts also explain the negative correlation values of Tyr in BSA and Pepsin shown in Table 4.2. The BSA and pepsin resolved component 2 also lack the detail present in all reference peaks, lacking the subpeaks at 450 nm to 500 nm. Conversely, the lysozyme and trypsin MCR-resolved component 2 closely follow the reference Tyr and Phe spectra. The detail absent from component 2 in BSA and pepsin and present for lysozyme and trypsin is possibly due to the aforementioned quenching of the tryptophan signal in the latter pair, which can also be seen in Figure 4.2.

Similarly, MCR-resolved component 3 peaks in pepsin and BSA show a lack of detailed features observed in the reference Phe spectra. However, the lambda max shift for all three proteins' resolved component 3 is similar (~75 nm from the reference Phe spectra) within a 10 nm range. The quenching of Phe by both Trp and Tyr could be likely, as both amino acids absorb in the region of 400 nm ~ 450 nm as demonstrated in Figure 4.2. The

secondary peak in trypsin and lysozyme at 425 nm excitation/ 425 nm emission is likely to be represented in this component. This peak can be linked to all three amino acid EEMs in Figure 4.1 but is most apparent in Phe. The structural differences in the spectral loadings for the proteins is likely due to differences in their secondary structures. Beta sheet concentrations for BSA, pepsin, lysozyme, and trypsin are reported to be 18%, 58%, 38%, and 37% respectively (The UniProt Consortium, 2023).

Overall, the MCR-resolved spectral loadings show strong to moderate correlations with their respective amino acid fluorescence spectra. Although the spectral loadings were able to predict some of the spectral characteristics of fluorophores using the overall fluorescence signal only, the results stand to be improved by analysing the results in the context of the quantum yields of the fluorophores at any given excitation wavelength. The separation of the resolved spectra could also be improved through contextualisation via more varied protein structures.

4.5.2.2. Comparing fluorophore abundance and MCR contributions

MCR contributions are indicative of the concentrations of the resolved components in a mixture. After establishing the fluorophore–component associations using the spectral loadings, the contributions for the MCR model were correlated with the abundance of each fluorophore in the 5 proteins used in this study. An example of the data matrix used for the correlation of amino acid counts with MCR-calculated contributions at 364 nm excitation can be seen in Table 4.3.

Table 4.3 Amino acid counts in each protein and the contribution scores of the MCR model at 364 nm excitation for each protein. The matrix is not readily parsed, thus it requires a correlation calculation.

	Trp	*Tyr*	*Phe*	*Comp-1*	*Comp-2*	*Comp-3*
BSA	2	21	30	71.7	35.6	43.9
Pepsin	5	17	16	18.9	19.8	0.5
Lysozyme	6	3	4	7.3	11.6	7.9
Trypsin	4	10	6	10.0	12.7	9.7

To facilitate the drawing of linear relationships between the components and the number of amino acids, the data matrix was analysed using a correlation matrix. The results for the correlations are shown in Table 4.4 below. A subset of spectra representing the previously studied 365 nm excitation was chosen for their relevance (Chapters 2 and 3), as well as the presence of strong fluorescence signals for all the aromatic amino acids at this excitation wavelength.

Table 4.4 Correlation coefficients for individual amino acid fluorophore counts and MCR-ALS calculated contributions. The shown fluorophore-component associations were matched according to their spectral loading.

Excitation	*Tryptophan*			Tyrosine			Phenylalanine		
(nm)	C1	C2	C3	C1	C2	C3	C1	C2	C3
362	-0.888	-0.896	-0.885	0.725	0.688	0.563	0.915	0.886	0.785
364	-0.884	-0.851	-0.878	0.797	0.873	0.566	0.957	0.992	0.792
366	-0.847	-0.734	-0.878	0.861	0.948	0.560	0.990	0.984	0.786

The inverse relationship between the Trp contributions and the other amino acids is indicative of the quenching behaviour of Trp. The changes in contribution correlation are indicative of the relative representation of the emissions at a particular wavelength range. For example Tyr correlations with components 1 and 2 increase over the wavelength range. The behaviour of Phe is similar for all the measured components. The increase in correlation for Tyr and Phe could be due to the decrease in quenching as Trp absorbance decreases between 360 nm – 370 nm as seen in Figure 4.1. As MCR analyses the variance

in the spectral dataset, Trp effects can dominate the variance in the range of 360 nm – 370 nm as their fluorescence intensities are much higher than Tyr and Phe.

The overall negative correlations of the Trp contributions are likely due to the decreasing quantum yield of tryptophan at the measured excitation range, however, quantum yield quantification is required to confirm.

4.6. Conclusions

The EEMs for the three fluorescent amino acids (Trp, Tyr, and Phe) were compared to the EEMs of five proteins (BSA, pepsin, lysozyme, trypsin, and insulin) at excitation wavelengths at 360 nm and higher. The analysis explained the prominent fluorescence features in the previously unstudied range of emissions and excitations for proteins.

The analysis of the EEMs for the four proteins was essential when interpreting the results of the MCR-ALS analysis. Distinctive features of amino acid fluorescence were distributed amongst the MCR components. The MCR model provided correlations between the abundance of fluorescence from the amino acids in each of the four proteins and MCR component contributions. The observed differences in correlation provided insight into the conformational differences between the proteins, the relative differences between the quantum yields of the fluorophores at any given excitation wavelength, and the relative signal-to-noise ratios for each of the fluorophores in the mixture.

The MCR models examined in this research were able to deconvolute a complex protein fluorescence signal and show correlations to the relative proportions of the aromatic amino acids contributing to the intrinsic fluorescence. The work provides a detailed description of fluorescence features for the lesser utilised excitations beyond 360 nm. The ability to break down complex fluorescence signals demonstrated offers great potential to simplify the process of identification of fluorescent components from the spectra of a mixture. The use of multivariate techniques demonstrated in this research provides possible applications in in-line industrial applications and laboratory-based protein analyses.

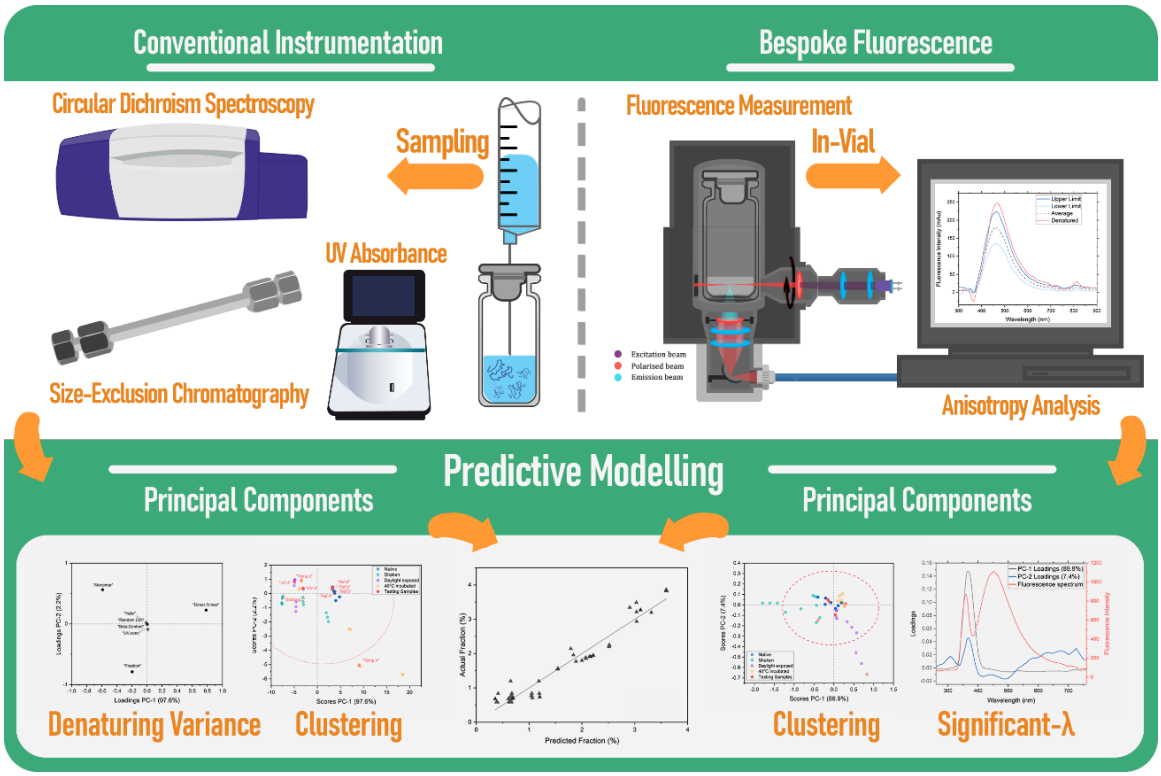
Chapter 5
Multi-variate analysis of fluorescence spectra for the determination of in-vial protein stability

5. Chapter 5: Multi-variate analysis of fluorescence spectra for the determination of in-vial protein stability

5.1. Highlights

- A stability study using BSA as a model protein and simulated realistic stress conditions was conducted and samples were characterised using standard laboratory techniques.
- Data analysis was enhanced with a combination of PCA and several constructed PLS models.

5.2. Graphical abstract



5.3. Introduction

The denaturation of protein therapeutics during transport and storage is known to lead to inefficacy or even deleterious immunogenic reactions (Deutscher, 1990). Proteins are reliant on their conformation and structure to achieve functionality (Varadhan, 1990). Processes imparting energy to the protein or resulting in enthalpic changes can lead to the destabilisation and loss of functionality (Ahern and Manning, 1992; Borzova et al., 2016; Deutscher, 1990; Hageman, 1988; Privalov and Gill, 1988; Yang and Honig, 1993). It is a regulatory requirement to provide evidence of therapeutic protein stability throughout its production and shelf life up to the point of administration (Kroger et al., 2022), hence, predictive testing of the stability during all stages of the drug's lifecycle is essential.

Several techniques are available for the high-throughput analysis of biopharmaceuticals during the manufacturing process (Razinkovet al., 2013). The utilization of online and in-line analysis enables the characterization of protein solutions without compromising the aseptic conditions during processing (Liu et al., 2022). An example of current technologies reported by Liu et al. relies on the sampling of materials on the manufacturing line to be analysed through an HPLC system coupled with MS or UV detectors (Liu et al., 2022). This automated system represents the state-of-the-art in online monitoring technologies and is capable of fully automated sampling and measurement of bioreactor products powered by a Python interface and specialised equipment. The system was capable of sampling at preset intervals or time points.

Another study by Harris et al. discussed real-time monitoring during the downstream processing within biomanufacturing (Harris et al., 2021). The study measured protein concentration using index of refraction (IOR) sensors and evaluated the effects of measurement parameters on accuracy and precision. This technology was presented as a

strong candidate for improving continuous manufacture through in-line analysis of biopharmaceuticals. However, a gap in constant and real-time monitoring of protein stability, and conformation remains to be addressed.

Due to the complexity of the datasets produced during the analysis of protein therapeutic stability, there has been a growing need for maximising the information that can be gleaned from analytical outputs. One such approach that provides insights into the relationships between multiple variables and helps to identify patterns and trends that may not be visible through other methods is multivariate analysis (Afifi et al., 2019).

Multivariate analysis is a well-known tool for the handling of multifaceted and complex datasets. Principal component analysis (PCA) is a data analysis technique used for dimensionality reduction by projecting the data onto a new low-dimensional space called principal component space. Therefore, the multitude of data points in a spectrophotometric data set could be characterised by a small set of new variables called principal components (Jollife and Cadima, 2016; Wang, 2009). The scores of the original data points on the newly created principal components can help explain the variance in the datasets without much loss of information (Wold, 1966).

Partial least squares (PLS) is a widely used chemometric multivariate analysis technique used to construct and quantify the correlations between predictors and variables. Although initially developed for econometrics, it has become increasingly popular in other fields too (Hui and Triandis, 1985). Unlike other techniques that focus on explaining covariance between items, PLS focuses on explaining variance in predictions of latent constructs. PLS produces parameter estimates that maximise explained variance, hence it is mostly used for prediction rather than explanation of relationships between items (Wold, 1966).

Multivariate techniques pair well with spectroscopic data, which can capture the subtle changes in protein structure and function but can be information-dense, and difficult to analyse. For these reasons, multivariate techniques have been used in tandem to provide overviews and valuable insights into sources of protein variance (Lee et al., 2012) and stability (Andrade et al., 2020; Uarrota et al., 2014).

In previous chapters, protein in-vial characterisation using fluorescence anisotropy was explored. The method described in Chapter 3 demonstrated the ability to distinguish freshly prepared bovine serum albumin samples from aggregated and unfolded samples utilising fluorescence anisotropy.

The objective of this work is to extend the application of fluorescence anisotropy characterisation to protein solutions in less controlled, more real-world environments. PCA was used to identify the patterns of denaturation of protein under the stability study conditions captured by fluorescence anisotropy and the secondary techniques (size exclusion chromatography (SEC), circular dichroism (CD), and UV absorbance). A PLS model was constructed to analyse the changes in protein stability markers (aggregation, unfolding and content) as a function of fluorescence behaviour. Moreover, a holistic image of protein sample stability was obtained by building a comprehensive protein in-vial analysis toolkit, using spectra as inputs.

5.4. Materials and methods

5.4.1. Sample preparation

Bovine serum albumin (BSA) was supplied by Sigma-Aldrich, Ireland. BSA samples were prepared at 25 mg/ml using a sodium phosphate buffer at pH 6.8 and 100 mM concentration. A total of 54 vials of BSA were prepared with 2 ml of solution aliquoted

per vial. The vials were split into a set of 9 reference samples which were measured immediately, and 3 sets of 15 vials for each stability condition. Prior to measurement, all proteins were filtered under aseptic conditions using a 0.2 µm syringe filter. Samples were individually diluted with sodium phosphate buffer to a 1 mg/ml protein concentration for CD, SEC, and UV measurements.

5.4.2. Secondary techniques

Size-exclusion chromatography, UV-spectroscopy, and circular dichroism analysis of a model protein (BSA) are described in section 3.4 of Chapter 3.

5.4.3. Stability study

To build and validate the PLS model a test set of samples was developed. BSA samples were stressed under conditions of ambient light, temperature and shaking simulating the stresses normally faced during the transport and storage of protein drugs.

BSA samples referred to as ‘daylight-exposed’ samples were left at room temperature (20°c) in direct sunlight for 1, 2, 7, 14 and 21 days. BSA samples referred to as ‘shaking’ samples were left on a rocking platform ((Biosan MR-1 Mini Rocker @ 30 Osc/min) for 1, 2, 7, 14 and 21 days, while 'temperature’ samples were left in a 40°C oven also for 1, 2, 7, 14 and 21 days. Reference samples were samples freshly prepared immediately prior to measurement.

All samples were tested for stability indicators using SEC for aggregation, UV spectroscopy for protein concentration, and circular dichroism neural networking analysis (CDNN) for BSA secondary structure. Fluorescence anisotropy was measured for all samples at each time point.

In total, the 54 vials for BSA prepared were subjected to the different storage conditions as described in Table 5.1 and generated a dataset of 72 fluorescence spectra.

Table 5.1 Sample count for each condition of the stability study.

Sample type	**Number of timepoints**	**Number of replicate vials**	**Number of spectra**	**Total spectra**
Reference	1	9	3	27
Temperature	5	3	1	15
Shaking	5	3	1	15
Daylight	5	3	1	15
Total	NA	18	6	72

5.4.4. Data processing

5.4.4.1. Pre-processing

The stability of BSA solutions was measured using the fluorescence anisotropy technique demonstrated in Chapter 3. The analysis involved normalisation to eliminate scattering effects on fluorescence intensity[1], followed by subtraction of the areas under the curves for each sample's two spectra (one at each polarisation angle). The normalisation approaches included the previously used single wavelength normalisation, henceforth referred to as lambda max. A standard normal variate (SNV) approach to normalisation was also attempted. The resultant average anisotropy for the 6 freshly prepared reference BSA samples (measured in triplicate) was used as the baseline for comparison. The pre-processing and anisotropy calculations are illustrated in Figure 5.1.

[1] The fluorescence intensity is normalised within polarisation paired spectra for a single sample. The effects of concentration, aggregation, and size increase due to unfolding are preserved within the data and were tested for between separate samples.

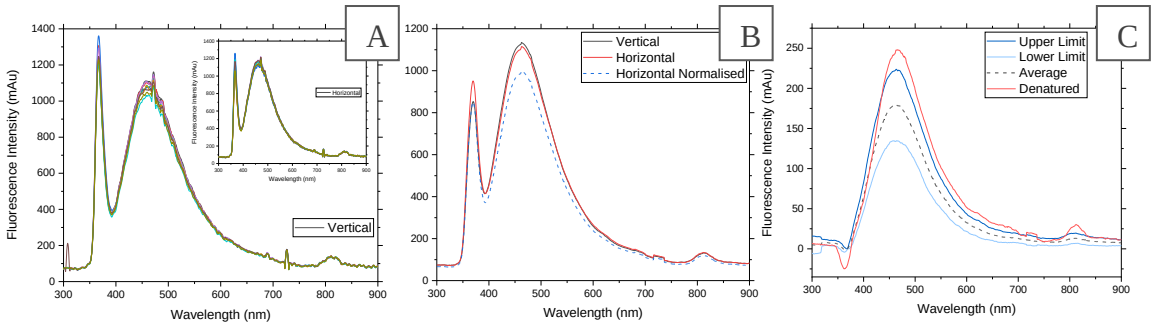


Figure 5.1 Processing of spectra to obtain a reference batch standard set. (A) The raw fluorescence spectra for both light polarisations for 9 samples of 25 mg/ml BSA. (B) The process of normalisation for a single sample, the horizontal spectrum (red) is normalised at 365 nm (emission wavelength) to eliminate scattering differences between the two polarisations. (C) The difference between both polarisations for 9 BSA samples as an average (black dotted) as well as the limits one standard deviation of the samples at each wavelength (upper limit in dark blue and lower limit in light blue resp.) a denatured sample (red) lying outside the limits.

5.4.5. Multivariate modelling

The fluorescence spectral database (n=72) and the results of the secondary analytical techniques were separately analysed using PCA. The PCA loadings for the fluorescence spectra were used to indicate the wavelengths of emission with the highest impact on the variance within the data set. The scores were compared to the PCA-modelled secondary technique data to observe variance patterns during the stability study.

Using a PLS model, the data set of 54 samples (9 vials for reference samples and 15 vials x 3 for stability study) was used to construct and test the relationship between the fluorescence spectra and the protein stability markers. The final time points for each stability condition were removed from the dataset to avoid overweighing outliers and corruption of the model (Xie et al., 2022). Validation of the PLS model was conducted using a “leave-one-out” cross-validation due to the relatively small size of the dataset (Liland et al., 2020).

The PLS dataset was organised with the sample conditions (reference/thermal/shaken/daylight exposed) as observations, the analytical (secondary

technique data) as dependent variables, and the fluorescence spectra as predictors. Several models were built to find the optimum method for analysing the data. R^2 values were used as a metric for the goodness of fit (closer to one is better), while the Root Mean Square Error of cross-validation (RMSECV) and the root mean Prediction Residual Error Sum of Squares (PRESS) were used as a measure of the prediction ability of the model (minimum value).

This study involved developing various models to analyse different response variables, including SEC-determined dimer/trimer content (high molecular weight species (HMW)), monomer content, degraded fractions (low molecular weight species (LMW)), UV-measured concentration, and CD-determined alpha helix, beta sheet, and random coil content. For each response, independent PLS models were created. In addition, semi-independent models were developed by grouping multiple responses for aggregation markers concentration, HMW species, LMW fraction, and a separate model for conformation markers (alpha, beta, random coil). To eliminate the effects of perfect collinearity observed in previous models, a single conformation response model was developed, which utilized the aggregation markers and beta content only as responses.

5.5. Results and discussion

5.5.1. Secondary Data Analyses:

5.5.1.1. Protein Concentration

The BSA concentration after 40°C incubation, daylight exposure (UV), and shaking, as indicated by UV spectroscopy is shown in Figure 5.2.

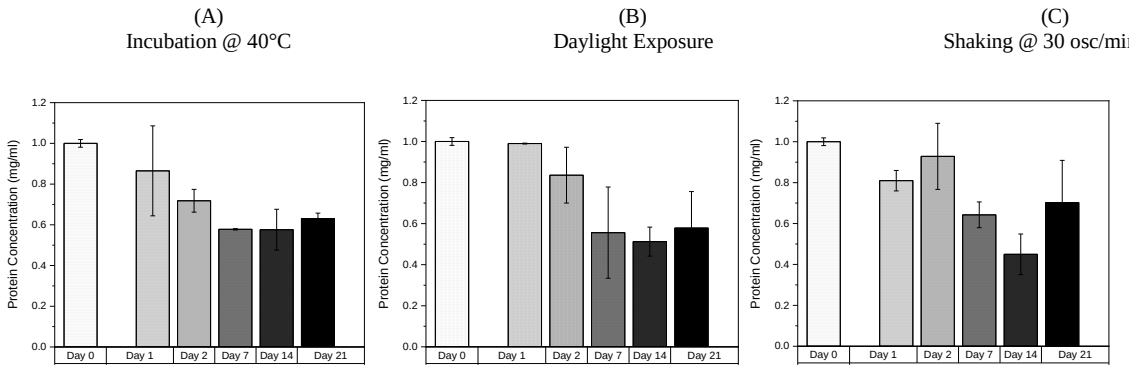


Figure 5.2 BSA concentration determined using UV spectroscopy for BSA samples (A) incubated at 40°c (normalised for scattering) (B) exposed to daylight and (C) Shaken at 30 osc/min. Samples were diluted from 25 mg/ml to a target concentration of 1 mg/ml to avoid concentration effects on CD, SEC, and UV. The BSA signal at 280 nm generally decreases over the observation period, with a plateau at days 7-14.

The general motif hints towards a decrease in monomer concentration. As discussed by Mahler et al., the kinetic degradation of protein takes the form of an increased susceptibility to proteolysis in the presence of oxygen radicals and under UV irradiation which are conditions likely to be present in the daylight (UV) exposed samples (Mahler et al., 2009). Due to the large variabilities in UV absorbance, it is difficult to directly compare the different conditions of degradation.

5.5.1.2. Aggregation

In its native state, a BSA solution contains a proportion of reversibly formed dimers, and trimers. In combination with oligomers of higher orders, dimers and trimers were considered for this study as HMW species to simplify the measurement of proportional aggregation. The development of HMW species began at different stages for samples under each of the stressor conditions.

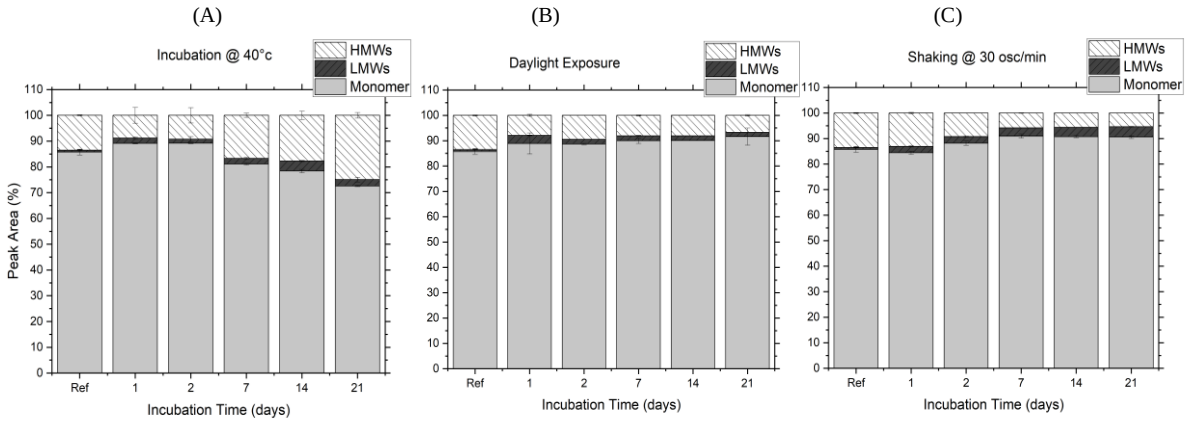


Figure 5.3 SEC area under the curve analysis for 25 mg/ml BSA stressed under "real-like" conditions for the specified number of days. BSA samples (A) incubated at 40°C (normalised for scattering) (B) exposed to daylight (C) Shaken at 30 osc/min. Visible aggregation can be seen, accompanied by an increase in the aggregate peak from day 7 onwards. Temperature incubation caused an increase in aggregation despite the loss in total concentration, while daylight and shaking caused decreases in proportional aggregates.

Figure 5.3 shows the SEC proportional content of HMW, LMW, and monomers. BSA solutions shaken at 30 osc/min, as well as, protein samples exposed to daylight exhibited an overall loss in HMW species proportion. Shaken samples' decreased HMW formation was accompanied by the development of later-eluting LMWs.

The kinetics of thermally-induced aggregation are well understood (Borzova et al., 2016; Chi et al., 2003; Weiss et al., 2009) and the patterns of BSA aggregation seen in Figure 5.3 are typical of high-temperature aggregation. Thermal incubation at 40°C caused insignificant changes to BSA aggregation in the first two days. However, an increase in aggregation peak area was seen after 7 days up to a maximum on the last point of observation (day 21).

5.5.1.3. Unfolding

CDNN results presented in Figure 5.4 show minimal changes after 1 day for all stress conditions, with an increase in variability and deviation starting at day 2. Daylight

exposure and shaking caused the highest increases in protein unfold variability, while incubation at 40°C caused a consistent unfold and refold effect. The increase in variability for both daylight exposure, and shaking, is most likely due to the energy of unfolding resulting from the collision of protein molecules in solution (Lapidus, 2017; Otzen, 2002).

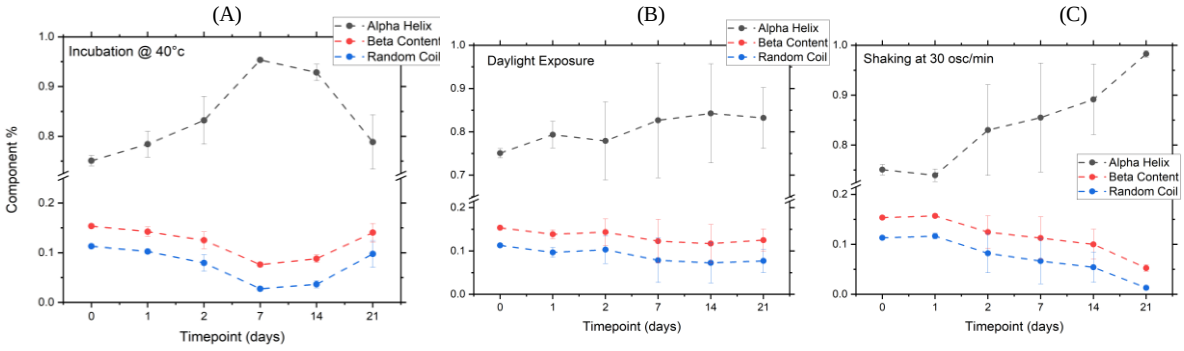


Figure 5.4 CDNN analysis of 25 mg/ml BSA samples A) incubated at 40°c, B) kept in daylight or C) shaken at 30 osc/min. A non-significant change in structure is seen on day 2, followed by the most significant change on day 7, and a slight refolding.

5.5.1.4. General patterns of degradation

BSA solutions stored under the three stress conditions showed degradation patterns consistent with the literature (Borzova et al., 2016; Michnik et al., 2008). Despite the differences in aggregation occurring on days 7, 14, and 21, a decrease in apparent concentration is still observed under all stress conditions.

Daylight exposure caused a loss in the proportion of HMW species, most likely due to an energy-selective effect under UV radiation. A higher-order species on the aggregation pathway have less energy and are therefore likely to be more stable under UV irradiation (Kaylor et al., 2005; Stefani, 2008). The study by Kaylor et al. suggests that the formation of intermolecular bonds leads to a higher stability of high-order aggregated species over lower-order species (Kaylor et al., 2005). The reduction in HMW species proportion continued over the entire observation period (21 days). Furthermore, a minor increase in

the LMW proportions and variance was observed over time. The proportional increase in monomer should also be considered with the changes in concentration seen in Figure 5.2. The larger standard deviation seen on days one and two is most likely a sampling variability due to the heterogeneous nature of the solution. Heterogeneity is to be expected due to the increased viscosity of a protein solution, and the kinetic nature of HMW and LMW species' formation (Grigolato and Arosio, 2019; Mahler et al., 2009). Due to the compounding factors of aggregation leading to increased scattering, as well as the aforementioned sampling difficulties, it is more accurate to consider the results in Figure 5.2 as images of sample absorbance, rather than absolute measurements of concentration. However, the figure indicates a clear change in the BSA solutions under the destabilising conditions for all formulations.

It is difficult to differentiate clearly between any single time point for multiple stressors using one technique. Relying on multiple techniques requires the analyst to parse multiple factors acting synergistically or antagonistically. Therefore, to enhance the clarity of the analysis PCA was used to analyse the fluorescence data in tandem with absorbance, aggregation, and unfolding results, and can be seen in Figure 5.6.

5.5.2. Multivariate analysis

Due to the large number of variables (spectra recorded at 1023 individual wavelengths) used for the spectral analysis, a multivariate approach was chosen to elucidate the relationships between spectra and sample properties as outlined by the secondary techniques.

5.5.2.1. Principal component analysis

Principal component analysis was used to show the feasibility of studying the relationships between the fluorescence spectra and the stability indicators using a multivariate model. The analysis was used to assess how the measured fluorescence related to protein degradation markers established via the secondary techniques. Analysing the loadings and scores allowed us to show significant effects in the datasets and relate effects to the variances in the fluorescence spectra.

Figure 5.5 (a) demonstrates the loadings for a PCA model built using the data from stability analysis via SEC, UV spectroscopy, and CDNN. While Figure 5.5 (b) demonstrates the variance captured by the fluorescent spectroscopy loadings in comparison to the reference spectrum.

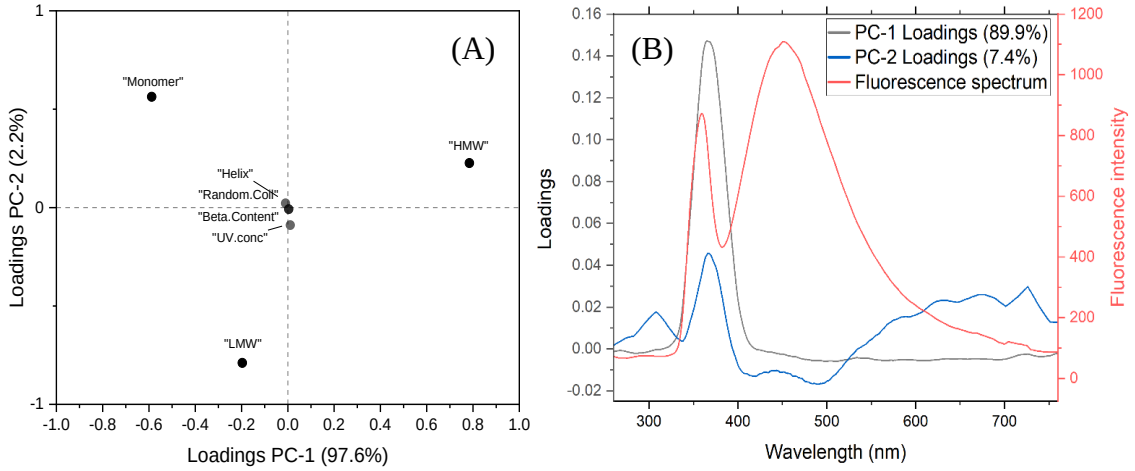


Figure 5.5 PCA model loadings for a model based on the secondary method analysis (left [a]) and the fluorescence spectra (right[b]) captured for a 25 mg/ml BSA solution under simulated stress conditions. The instrumental loadings (left) demonstrate the larger variance of aggregation effects over conformational changes. The fluorescence loadings (right,) for PC-1 (black) and PC-2 (blue) as compared to an average of the fluorescence spectra (red), show higher loadings at the 365 nm peak which is the part of the spectrum related to scattering and aggregation.

The conformational loadings (secondary structure) closer to the origin have a lesser effect on the model due to lower variance in the dataset. By comparison, the loadings of

monomer, HMW, and LMW proportions are furthest from the origin indicating a higher variance in the dataset, and as such, have the highest effect on the PCA model. As datasets are centred and scaled prior to multivariate analysis the higher loadings for aggregation indicators hint towards a real, comparatively more apparent effect, in the stressed samples. Similarly, the loadings for a PCA model for the spectral dataset of the same sample show the highest deviation from the origin at 365 nm~390 nm which is the area of the spectrum related to scattering and therefore aggregation (Chullipalliyalil et al., 2023). The loadings point towards aggregation being the main effect for both the stability data (analytical instruments) and the fluorescence spectra.

The PCA score plots seen in Figure 5.6 show the clustering of the reference samples in both the stability indicators and their corresponding fluorescence spectra (Figure 5.6 (a) & (b) respectively). Stressed samples show less clear delineations between clusters due to overlap in aggregation and unfolding effects between different stress conditions.

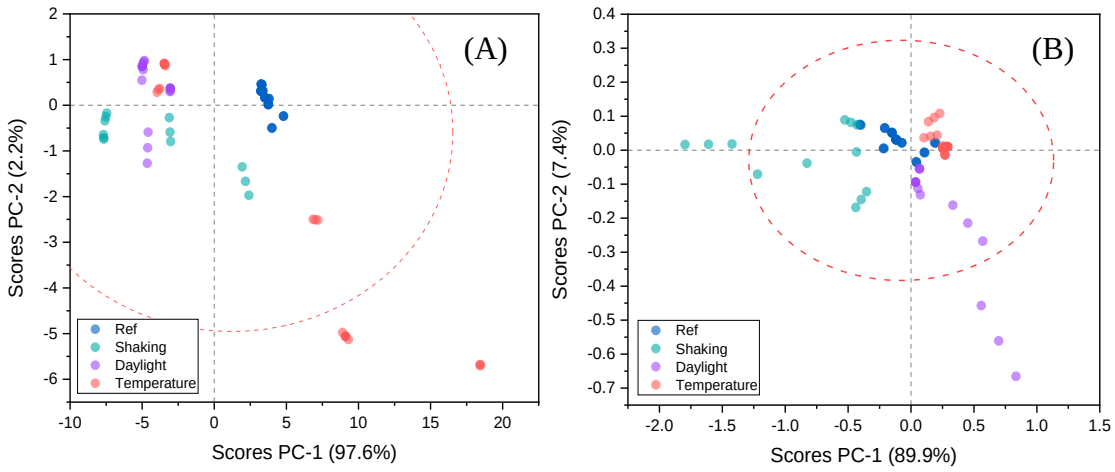


Figure 5.6 a & b showing PCA scores for a model based on the instrumental analysis (left [a]) and the fluorescence spectra (right[b]) captured for a 25 mg/ml BSA solution under simulated stress conditions. Clustering of the reference samples (blue) can be seen in both cases. While the stressed samples show less delineation between clusters owing to the overlap of stressor effects between different stress conditions.

5.5.2.2. Partial Least Squares Regression

In this study, the stability markers for concentration, aggregation, and conformation were estimated from the fluorescence data to evaluate the “reference batch” detection method. For this, several PLS models were constructed and subsequently, the prediction accuracy of the models was examined by cross-validation. The models were developed for all the response variables including HMW species, monomer, LMW species, UV-measured concentration, alpha helix, beta-sheet, and random coil contents. For each response, individual PLS models were built and considered as "independent" models. The results for predictability, fit and cross-validation for these models can be seen in Table 5.2.

Table 5.2 Summary of Independent PLS Models using the fluorescence anisotropy as predictors for the variables listed below. The method of normalisation was standard normal variate.

	R^2	RMSEcv	Root mean Press	Latent variables	Variance explained	Method of normalisation
HMW	0.95	0.994	0.312	10	95.50%	SNV
Monomer	0.96	0.849	0.279	10	96.08%	
LMW	0.95	0.251	0.307	9	95.03%	
Concentration (UV)	0.53	0.158	0.954	4	52.50%	
Alpha Helix	0.48	0.065	0.917	4	48.18%	
Beta Sheet	0.49	0.024	0.912	4	48.70%	
Random Coil	0.50	0.026	0.900	4	50.35%	

Additionally, semi-independent models were developed by grouping multiple responses for aggregation markers (HMW, monomer, LMW, concentration) and developing a separate model for conformation markers (alpha helix, beta content, random coil). The models were tested twice for different pre-processing normalisation methods for the spectra. The PLS results for spectroscopy data using a standard normal variate normalisation can be seen in Table 5.3. The alternative models using the 365 nm peak

maxima to normalise the polarisations before anisotropy calculations are shown in Table 5.4. The latter approach results in a pair of models that favour the aggregation markers over the conformational ones (85% vs. 36% variance explained). The models built using standard normal variate normalisation are more balanced (70% vs. 66% variance explained) between the aggregation markers and the conformational stability.

Table 5.3 PLS models for 2 combinations of variables the first includes the aggregation markers and UV absorbance, while the second includes the conformational stability markers only. Fluorescence Anisotropy data was used as a predictor, and the method of normalisation was standard normal variate (SNV).

	R^2	RMSEcv	Root mean Press	Latent variables	Variance explained	Method of normalisation
HMW	0.82	1.947				
Monomer	0.81	1.977				
LMW	0.74	0.601	0.7	6	70.01%	
Concentration (UV)	0.53	0.062				SNV
Alpha Helix	0.65	0.052				
Beta Sheet	0.65	0.020	0.84	7	66.26%	
Random Coil	0.68	0.021				

The table shows a high predictive ability for aggregation markers, as expected due to higher variance in aggregation data demonstrated by the PCA model. However, the relatively higher RMSE values indicate the possibility of overfitting of the data. This is especially true for models built to model the aggregation data alone. The results also demonstrate the advantage of using SNV normalisation as opposed to standard lambda max normalisation. The latter results in little explanation of the variance in conformational components most likely owing to the known efficiency of SNV normalisation when dealing with spectral scattering (Witteveen et al., 2022).

Table 5.4 PLS models for 2 combinations of variables the first includes the aggregation markers and UV absorbance, while the second includes the conformational stability markers only. Fluorescence Anisotropy data was used as a predictor, and the method of normalisation was according to the 365 nm peak maximum.

	R^2	RMSEcv	Root mean Press	Latent variables	Variance explained	Method of normalisation
HMW	0.96	1.034				
Monomer	0.96	1.035				
LMW	0.94	0.290	0.56	12	85.33%	
Concentration (UV)	0.58	0.157				To lambda max
Alpha Helix	0.40	0.078				
Beta Sheet	0.40	0.030	0.95	5	36.12%	
Random Coil	0.42	0.032				

Consequently, a single conformation response model was developed to eliminate the effects of collinearity observed in the previous models and used the aggregation markers and the beta content only as responses. The model attributes can be seen in Table 5.5. The model utilised standard normal variate normalisation and was able to explain 87.92% of the variance in the data set using 9 components. The weakest fit of the modelled variables belongs to the conformational data with an R^2 value of 0.63. However, the results are in line when considering the large variability of conformational data seen on CD (Figure 5.4). The model fits (R^2 values) are generally in line with the variances in the data set indicated by the PCA model for the secondary techniques.

Table 5.5 PLS model of the aggregation variables and a single conformational variable (namely the beta content). Fluorescence Anisotropy data was used as a predictor, and the method of normalisation was standard normal variate (SNV).

Parameter	R^2	RMSEcv	Root mean Press	Latent variables	Variance explained	Method of normalisation
HMW	0.96	0.676				
Monomer	0.96	0.663				
LMW	0.94	0.247				
Concentration	0.88	0.018	0.487	9	87.92%	SNV
Beta Sheet	0.63	0.049				

The estimated profile of the stability responses was investigated using the model with the highest explained variance for all the components to show the predictive power of the PLS model. The optimum model resulting in the most variance explained while maintaining a good fit for all components is the single conformation co-model. The fit for the model is displayed in Figure 5.7 below. In this figure, one can see that the aggregation and denaturation profiles were reasonably well-estimated from the fluorescence spectra for both anisotropy and aggregation-related scattering. The corresponding R^2 and RMSE values for these models can be seen in Table 5.5, indicating that the prediction accuracy was the most reliable. It is worth noting that the prediction errors were in the same order of magnitude as the relative standard deviations of values where the conformational data collected via CD showed a higher variability (24% at the highest) than the aggregation (4.1% at the highest). This demonstrates that the constructed PLS models could provide an accurate estimation of the responses.

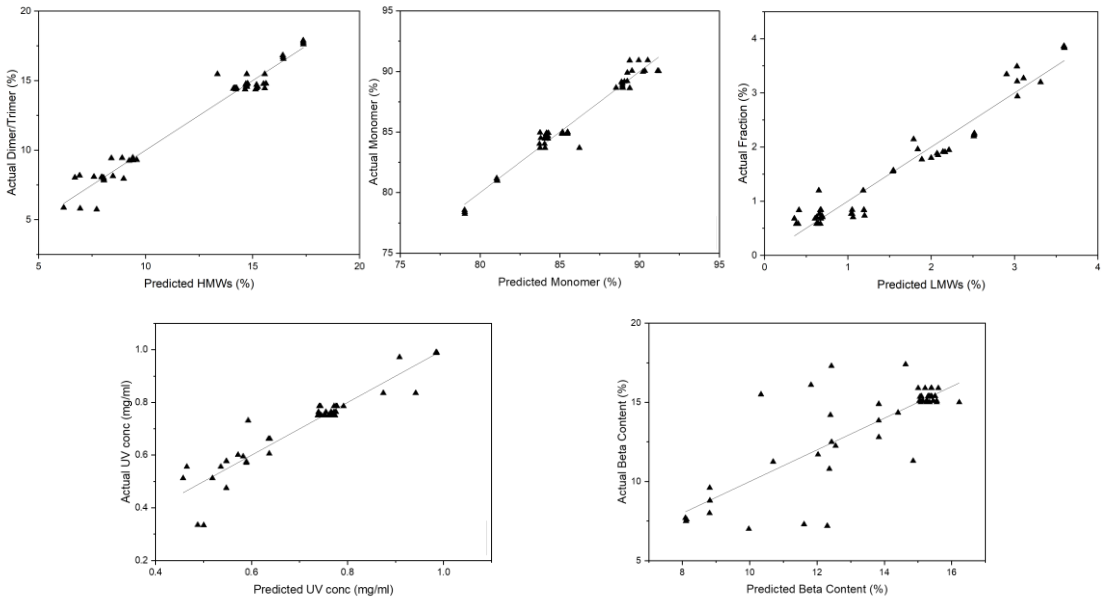


Figure 5.7 Actual vs. predicted results for the single conformation indicator model. Proving the usability of the fluorescence spectra as a predictor for aggregation and secondary structure of a protein solution

The samples also used in the “reference batch” testing are shown in red and show good distribution throughout the testing space. The secondary structure variables other than β-sheet content were excluded from calculations due to their collinearity. The figure shows the true results as measured by SEC HPLC, UV, and CD in comparison to the predicted values obtained during the calibration and validation of the PLS model.

5.6. Conclusion

The application of fluorescence anisotropy and multivariate analysis presented in this research was able to successfully differentiate denatured samples and native protein solutions. The PLS model shows that the fluorescence spectra can be used as a comprehensive source of information on the protein conformation and stability. The model was able to identify the main sources of variance within the data set, indicating the reproducibility of aggregation results over the measured time range. The model also helps in differentiating samples by the major form of degradation within the samples, which gives rise to possible applications in analytical tracing of sources of degradation. The analytical technique outlined in this chapter can be utilized to create handheld devices that can monitor the stability of biopharmaceutical products from the point of manufacture to administration. The customizability of this technology allows for its application in various industrial settings, including the use of in-line systems for pharma and biopharma industries to conduct rapid, non-destructive quality control inspections of product release. The technology can also provide a comprehensive overview of product quality during stability screening.

Chapter 6
General Discussion

6. Chapter 6: General Discussion

Proteins are valuable biomolecules in pharmaceutical development, with functions as active pharmaceutical ingredients (Johnston, 2007), carriers (Hong et al., 2020), and excipients (Johnston et al., 2010; Mercade-Prieto and Gunasekaran, 2009). The characterisation of proteins is therefore an important part of quality control in pharmaceutical manufacturing (Hinz, 2006; Johnston, 2007; Liu et al., 2022). To date attempts have been made to characterise proteins through glass surfaces or in their original containers to reduce the effects of manipulation or preserve sample sterility (Affleck et al., 2021; André et al., 2015; Berry et al., 2016; Cook and Ward, 2011; Kim and Morr, 1996; Wen et al., 2010; Werk et al., 2015). The characterisation of protein solutions and formulations using a novel intrinsic fluorescence approach, while retaining vial closure, was a major focus of this thesis. The aim of this thesis was to design novel spectrophotometric analytical methods to characterise protein formulations and apply these techniques using a combination of spectroscopy and multivariate spectral analysis.

6.1. Thesis findings

The development of novel analytical techniques for biopharmaceuticals is of interest to both pharmaceutical manufacturers and healthcare providers. Techniques simplifying analytical workflows have the potential to save on training, research, and manufacturing costs. The current state of the protein analytics field as discussed in Chapter 1, trends towards a reliance on well-established but destructive techniques to measure the stability and quality characteristics of a protein formulation. Therefore, the development of novel, accurate and efficient analytical techniques is of vital importance. This thesis provides an overview of the development of one such technique for various applications in protein

characterisation. Non-destructive, in-vial analytical applications were carried out in this study including measuring the reconstitution time of a lyophilized protein formulation using intrinsic fluorescence (Chapter 2), measuring protein stability markers using fluorescence anisotropy (Chapter 3), automating stability measurement and data analysis (Chapter 5), and exploring the properties of intrinsic fluorescence of proteins at excitation and emission wavelengths measurable in-vial (Chapter 4).

In this PhD project multivariate analysis was applied in multiple ways to aid with the elucidation of internal structures within large datasets (Chapters 2 and 5), as well as to verify the conclusions reached as a result of complex data sets with large numbers of spectroscopic variables (Chapter 5). Multivariate deconvolution techniques including algorithms of multivariate curve resolution were also applied successfully to increase the understanding of fluorescence sources of intrinsic fluorescence emissions explored in earlier chapters (Chapter 4).

The value of this research stems from the demonstration of method development, and the application of multivariate computational analysis to aid in the characterisation and enhancement of the technique itself. The novel application of the near-UV excitation wavelength to obtain fluorescence emission in a range of the spectrum transmissible through borosilicate glass is unprecedented in the literature. The thesis is primarily concerned with, and explores, the areas of fluorimetry, protein analytics, and multivariate data handling.

6.1.1. Borosilicate glass transmissible intrinsic protein fluorescence

Chapter 4 of this thesis discussed the causes and details of the intrinsic fluorescence of proteins at 365 nm excitation/450 nm emission. The intrinsic fluorescence of proteins is

caused by the three aromatic amino acids (fluorophores) tryptophan, tyrosine, and phenylalanine abundant in proteins (Edelhoch, 1967; Ichikawa and Terada, 1977). Using 3D emission excitation matrices, we were able to describe the overall protein fluorescence at 365 nm excitation/ 450 nm emission in terms of the fluorophore emissions. While these emissions have not been extensively studied in the literature(Moriyama et al., 1996), they are advantageous as they can be transmitted through borosilicate glass surfaces (Chapter 3). The most recent publication dealing with the excitation 365 nm/emission 450 nm fluorescence of proteins is the publication by Wang et al. (Wang et al., 2019) where the general photoluminescence of proteins in the visible regions was analysed for root causes and concentration-dependent effects. The novelty of the application of this emission at 450 nm presented here emanates from its applicability to in-vial monitoring and in-line analysis without disruption to the protein formulation.

Advances in manufacturing have led to high throughput protein drug manufacturing pipelines (André et al., 2015; Hinz, 2006; Liu et al., 2022; Razinkov et al., 2013). In Chapter 2 the transmission of the 365 nm excitation/450 nm emission allowed us to probe the intrinsic fluorescence during the dissolution/reconstitution of a model protein through the walls of a borosilicate glass vial. Fluorescence is ideal when attempting to capture the minutiae of a kinetic process such as dissolution. The dissolution process of a protein is influenced by multiple variables involving the rehydration of the protein, domain changes, and interactions with formulation excipients which lead to electromagnetic changes in the microenvironment surrounding the fluorophores within the protein molecule (Beech et al., 2015; Kulkarni et al., 2018; Sane et al., 2020; Zhou et al., 2016; Zimmermann et al., 2000). The ability to capture the changes in the protein fluorophore microenvironments

could hypothetically be leveraged for any of the kinetic events that cause changes in protein including the unfolding (Monsellier and Bedouelle, 2005) and aggregation (Clarke and Oprysa, 2004) as discussed in Chapters 3 and 5.

Fluorescence anisotropy is a modification of fluorescent emission detection that utilizes polarised light and has been demonstrated to detect changes in protein conformation during unfolding (Gijsbers et al., 2016). Chapter 3 showed the development of a bespoke analytical instrument designed to measure protein stability using fluorescence anisotropy at 365 nm excitation/450 nm emission in-vial. In Chapter 5 we were able to expand on the results of Chapter 3 with more complex degradation phenomena including UV-degradation, impulse-induced damage through shaking, and temperature-induced unfolding and aggregation. Chapter 5 included several iterations of a predictive model using the fluorescence anisotropy spectra as predictors of protein stability markers. The total amount of variance captured by the fluorescence spectra through the models varied depending on the pre-processing conditions as well as the inclusion of differently degraded protein samples. For example, the high-temperature degraded samples which are known to have the most orderly degradation kinetics (Alford et al., 2008; Borzova et al., 2016) produced models with the highest total explained variance (model parameters shown in Chapter 5). As well as, showing the smallest degrees of variability throughout other methods of protein stability measurements (Chapter 5).

Despite the relatively low intensities of fluorescent emissions after excitation at 365 nm when compared to 280 nm (Chapter 2) for proteins. They have been used successfully to identify the characteristic changes of proteins during or as a result the kinetic processes of dissolution, unfolding, and aggregation. Aided by the specialised post-processing

approaches utilised in each application, the in-vial fluorescence has much potential for future explorations.

6.1.2. Multivariate analysis techniques

As this thesis is mainly concerned with the interpretation of various fluorescent emissions producing large data matrices, specialised processing tools were required to handle the statistical calculations (Agresti, 2015). The growing complexity of the data sets required more specialised or complementary applications of multivariate analysis techniques (Afifi et al., 2019).

In Chapter 2 we produced fluorescence data matrices for a reconstituting protein formulation. As it was assumed the protein will reach an equilibrium state upon complete dissolution (Kramer et al., 2012), the study aimed to simply relate the fluorescence spectra to the equilibrium of the dissolving protein. It was determined that a linear regression at any one wavelength of emission over time would be insufficient in quantifying reconstitution, due to the nonlinearity of the fluorescence growth in cases of quenching due to the protein concentration. PCA has been used to quantify the kinetics of complex chemical reactions previously (Gokulakrishnan et al., 2006). Therefore, principal component analysis (PCA) was used to break down the changes in the fluorescence spectra over time. The PCA was able to reduce the variance in the fluorescence emissions captured in time series down to a single principal component with an average explained variance of 92%. The use of PCA in this chapter exemplifies the dimensionality reduction typical of multivariate analysis in general and PCA specifically(Bhattacharya and Burman, 2016).

While Chapter 2 primarily leverages the dimensionality reduction property of multivariate analysis, Chapter 5 utilizes the ability of PCA to explore the relationships of multiple variables at once (Wang, 2009). The breakdown of variable relationships using PCA is a common way of dealing with spectroscopic results (Bhatia et al., 2018; Kamga et al., 2013; Lee et al., 2012). While spectroscopic data can capture the nuanced aspects of complex material interactions due to the bilinear nature of spectroscopic matrices, it is often not feasible to experimentally isolate variables to determine causal relationships (Rencher and William, 2012). The study conducted in Chapter 5 used PCA to separate degraded samples of protein from their non-degraded reference samples. By analysing fluorescence anisotropy data the loadings of the models indicated the areas of interest in the emission spectrum, and aligned with the findings of Chapter 3. The scores of the first and second principal components were successfully used to differentiate non-degraded samples from degraded ones. The PCA analysis was followed up by a partial least squares regression (PLSR) to determine the predictive ability of the fluorescence spectra with regard to the protein stability markers of aggregation, secondary structure and concentration. PLSR displays another aspect of multivariate analysis through the modelling of complex multivariable relationships among the latent variables in the dataset (Höskuldsson, 1988). After PLSR model optimisation the final model reached showed a correlation between the fluorescence anisotropy spectra and the protein stability markers. Both PCA and PLSR modelling provided the statistical backing to the fluorescence anisotropy analytical method discussed in Chapters 3 and 5 as well as identifying the main sources of variability in the degraded protein samples.

Chapter 4 features the novel use of multivariate curve resolution (MCR) to correlate protein fluorescence spectra with their compositional individual fluorophore spectra. MCR is a multivariate technique that has been used for mixture separation, primarily via the iterative variant of the algorithm MCR- alternating least squares (MCR-ALS) (Bayat et al., 2020; Lawton and Sylvestre, 1971). The presence of fluorophores within the protein structures alters their fluorescence emissions, therefore the pure spectra of the fluorophore emissions could not be obtained via direct measurement. However, the MCR-ALS algorithm has been shown to predict spectroscopic information without defined reference spectra in the literature (Queiroz et al., 2021; Smith et al., 2019). The results of both the loadings and scores of the MCR-ALS application in Chapter 4 showed a correlation between the estimated components and the individual fluorophore spectra in an isolated form. The scores also correlated with the fluorophore concentrations in the proteins tested. While the results do not indicate a concrete relationship between the MCR-ALS results and measured fluorescent emissions, they demonstrated the potential use of the algorithm as an aid in the deconvolution of protein fluorescence.

6.1.3. Non-destructive protein analysis

The bespoke setup development for the evaluation of in-vial fluorescence is detailed in Chapter 2 with modifications for the measurement of fluorescence anisotropy in Chapter 3. The device was built using readily available components in a compact form factor requiring only a power source and a spectrophotometer for complete operation. The capture of fluorescence from the bottom of the vial rather than from the side was important to prevent a loss of emissions through refraction as a result of the curvature of the vial (Chapter 3).

The use of low-intensity fluorescent emissions at 450 nm required high excitation power to achieve good resolution. To ensure the LED power was consistent throughout all measurements the power of the transmitted light was checked with and without the vial and found to be linearly related to the diode current, indicating no optical non-linear absorption effects from the vial (Chapter 3). As UV light is also known to denature proteins, the transmission was recorded in the presence of a protein sample at the start and end of measurements, as well as, after a period of recovery. The results indicated that the protein absorbance decreased non-significantly and were able to completely recover after the rest period.

All in all, the research presented in this thesis provides an all-in-one platform for the qualification of a range of protein solutions. Firstly, the bespoke apparatus provides the potential for upgradability and portability. Secondly, the application of fluorescence and fluorescence anisotropy at 365 nm excitation /450 nm emission allows measurements to be performed completely isolated from the environment and without any damage to the samples. Finally, the application of multivariate post-processing techniques presents opportunities for understanding the complex processes involved in protein kinetics as well as the quantification of the processes of dissolution and degradation.

6.2. Summary of key findings

The overall aim of this thesis was to develop a novel analytical technique to investigate the stability properties and critical quality attributes of protein solutions in-vial using spectroscopic analysis. The selection of an appropriate excitation-emission wavelength range was essential to achieving in-vial characterisation. Multivariate analysis was used to simplify the interpretation of protein spectra.

The main novel aspect of the project was the ability to measure protein concentrations, conformation, and degree of aggregation for samples housed in borosilicate containers. As an outcome of this study, the key findings to consider in characterising proteins using fluorescence in-vial are summarised as follows:

I. The use of an excitation wavelength at 365 nm and detection of the subsequent fluorescence emissions at 450 nm enables the observation of protein molecules within borosilicate glass vials. Standard fluorescence techniques rely on excitations under 280 nm and are therefore unusable through borosilicate glass, where the transmission cutoff lies around 300 nm. While the details of 365 nm excitation had been available in the literature, it was not used due to its comparatively lower efficiency. However, the ability to bypass glass barriers presented in this thesis is of particular importance to manufacturing and lab-testing applications.

II. The application of multivariate analysis to a spectral time series to quantify the reconstitution time for a lyophilised formulation. Simplifying the process of tracking changes in spectra and utilising all the information presented by each spectrum to accurately determine the time point at which an equilibrium is reached.

III. Identification of protein stability parameters within fluorescence anisotropy data captured in-vial using 365 nm excitation. The technique used for identifying aggregation and conformational stability markers simultaneously was compared against traditional methods such as SEC-HPLC, CD, and DLS, confirming the usability of fluorescence anisotropy as a viable alternative with the added benefit

of being able to perform analysis in a sealed vial. The PLS model generated enables the qualification of protein stability in-vial, using an efficient, easy-to-use, and non-destructive technique.

IV. The application of multivariate curve resolution to decompose protein fluorescence spectra to the spectra of the individual fluorophores in their makeup. MCR has proven to be effective in handling the complex information contained in these spectra and has the potential to be applied to various types of protein fluorescence decompositions.

6.3. Limitations

The limitations of this study centre on the quantum yields of the protein intrinsic fluorescence at the excitations used to bypass the transmissibility of borosilicate glass. In Chapters 3 and 4, fluorescence anisotropy was used to measure the stability of a model protein successfully. However, the technique was not attempted to measure the structural changes for edge case proteins that contain very few or none of the fluorophore amino acids. Characterising such proteins with weak signals would require additional resources, equipment, and time for optimisation.

Another primary limitation of the measurement instrument is the actual volume under the effect of the excitation beam of the LED source. The Chapter 2 and 3 supplementary data include detailed calculations on the geometries of the volume affected by the beam which was calculated at 12.6 mm^3 (see appendix sections S2.1 and S3.1). The total volume of the solution is formulation-dependent, for a 2 ml sample the total volume will be close to 2000 mm^3 (depending on the specific density of the total dissolved solids) thus the measured volume is around 0.63 % of the total volume. While the small measurement

point is not uncommon for spectroscopic instruments, the lack of homogeneity in highly degraded protein samples can cause high degrees of variability in measurements. The measurement variability has been observed for UV and circular dichroism measurements in Chapter 5 as well.

The availability of model proteins to test the lower limits of detection of the instrument was also limited. While smaller macromolecules such as BSA and lysozyme are plentiful and easy to obtain, larger proteins such as immunoglobulins are not readily available. The use of smaller proteins limits the number of available fluorophores, resulting in a weak and noisier signal. The upper limit to concentrations used was imposed by the fluorescence inner-filter effects (quenching), thus future efforts to quantify higher concentrations would require accounting for and bypassing the quenching.

6.4. Future work

The presented work investigates the potential of in-vial fluorescence in characterising protein formulations. The framework proposed allows for further research into the use of specialized spectroscopic techniques to enhance the current state of large-molecule analytical techniques.

Future work is recommended in the following areas:

I. The use of different spectroscopic techniques such as near-infrared (NIR) and Raman spectroscopies to characterise the stability of proteins in solution. NIR spectroscopy is a well-established and well-characterised technique, while Raman technology has rapidly developed to rival other analytical methods.

II. Exploration of the fluorescence anisotropy technique limitations including the lower limits of real-time detection for folding changes in a denaturing protein.

III. Use of MCR for the characterisation and deconvolution of a sample mixture of proteins.

IV. Application of the MCR multivariate deconvolution to more energetic emissions for proteins at 280 nm excitations and below.

V. Explore the use of spectroscopic techniques to study the thawing process of frozen protein solutions in real-time.

VI. Investigate the usage of fluorescence anisotropy to detect, quantify and study the structures of nucleic acids.

VII. Application of in-vial fluorescence measurements to viral capsid proteins as a practical application for biotechnological development.

6.5. Conclusions

The overall aim of this thesis was to develop and study the use of in-vial fluorescence for the characterisation of protein solutions. The primary objective of this thesis was achieved by designing a bespoke apparatus capable of measuring fluorescence within a standard 10ml borosilicate lyophilisation glass vial. The apparatus utilised an excitation wavelength of 365 nm capable of transmission through the glass vial. The measurements were enabled by the intrinsic fluorescence of proteins at 450 nm, while the emission was previously known, it has seen little exploitation in the literature due to its low efficiency when compared to emissions at 350 nm. The apparatus enabled the measurement of an accurate reconstitution time for a lyophilised protein formulation. The reconstitution time was chosen as the studied critical quality attribute due to its wide use, as well as, a lack of instrumentational methods for its measurement ~~in industrial settings~~. The measured

reconstitution times were more precise than traditionally measured ones and provided a more accurate time by accounting for subvisible undissolved particulates.

Protein aggregation and unfolding markers were studied using a modified apparatus utilising polarised filters to measure fluorescence anisotropy. The stability measured via anisotropy correlated with the results of established protein characterisation techniques including SEC-HPLC, CD, and UV absorption. The fluorescence anisotropy apparatus developed was able to measure a model protein's stability in-vial using a scalable, accurate and non-destructive method. The automation of the process using multivariate analysis and R-programming was used to build a holistic, and easy-to-use protein stability characterisation protocol for use by laboratory scientists and healthcare providers. Finally, the application of MCR successfully related the decomposed fluorophore spectra to the characteristic fluorescence of 5 proteins. The technique provides a platform for the further characterisation and understanding of protein fluorescence spectra.

This thesis makes a noteworthy contribution to the field and advances the existing knowledge on protein fluorescence spectroscopy. Additionally, it suggests new areas of inquiry and provides recommendations that could be further explored. For instance, the potential of using a fluorescence platform to characterize individual proteins within a mixture or a matrix that is rich in fluorophores.

Bibliography

Adamian, L., Liang, J., 2002. Interhelical hydrogen bonds and spatial motifs in membrane proteins: Polar clamps and serine zippers. Proteins: Structure, Function and Genetics 47, 209–218.

Administration, F. and D., 2005. Food and Drug Administration Compliance Program Guidance Manual. Food and Drug Administration: Rockville, MD, USA.

Affleck, R.P., Khamar, D., Lowerre, K.M., Adler, N., Cullen, S., Yang, M., McCoy, T.R., 2021. Near Infrared and Frequency Modulated Spectroscopy as Non-Invasive Methods for Moisture Assessment of Freeze-Dried Biologics. J Pharm Sci 110, 3395–3402.

Afifi, A., May, S., Donatello, R., Clark, V., 2019. Practical multivariate analysis.

Agresti, A., 2015. Multivariate Analysis: Discrete Variables (Overview). International Encyclopedia of the Social & Behavioral Sciences: Second Edition 131–136.

Agro', A.F., Albergoni, V., Cassini, A., 1974. Intrinsic fluorescence of a protein devoid of tyrosine and tryptophan: Horse hepatocuprein. FEBS Lett 39, 164–166.

Ahern, T.J., Manning, M.C., 1992. Stability of protein pharmaceuticals.

Akyurt, M., Zaki, G., Habeebullah, B., 2002. Freezing phenomena in ice–water systems. Energy Convers Manag 43, 1773–1789.

Alford, J.R., Kendrick, B.S., Carpenter, J.F., Randolph, T.W., 2008. High concentration formulations of recombinant human interleukin-1 receptor antagonist: II. Aggregation kinetics. J Pharm Sci 97, 3005–3021.

Al-hussein, A., Gieseler, H., 2013. Investigation of histidine stabilizing effects on LDH during freeze-drying. J Pharm Sci 102, 813–826.

Allison, S.D., Randolph, T.W., Manning, M.C., Middleton, K., Davis, A., Carpenter, J.F., 1998. Effects of Drying Methods and Additives on Structure and Function of Actin: Mechanisms of Dehydration-Induced Damage and Its Inhibition. Arch Biochem Biophys 358, 171–181.

Alostad, H., Davulcu, H., 2015. Directional prediction of stock prices using breaking news on twitter. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT).

Anand, U., Jash, C., Mukherjee, S., 2011. Protein unfolding and subsequent refolding: a spectroscopic investigation. Physical Chemistry Chemical Physics 13, 20418–20426.

Andrade, G.C., Medeiros Coelho, C.M., Uarrota, V.G., 2020. Modelling the vigour of maize seeds submitted to artificial accelerated ageing based on ATR-FTIR data and chemometric tools (PCA, HCA and PLS-DA). Heliyon 6, e03477.

Andrade, J., 1985. Surface and Interfacial Aspects of Biomedical Polymers. Springer US.

André, S., Cristau, L. Saint, Gaillard, S., Devos, O., Calvosa, É., Duponchel, L., 2015. In-line and real-time prediction of recombinant antibody titer by in situ Raman spectroscopy. Anal Chim Acta 892, 148–152.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat Surv 4, 40–79.

Atkinson, W.L., Pickering, L.K., Schwartz, B., Weniger, B.G., Iskander, J.K., Watson, J.C., 2002. General recommendations on immunization. Recommendations of the Advisory Committee on Immunization Practices (ACIP) and the American Academy of Family Physicians (AAFP). MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control 51, 1–35.

Avkiran, N.K., Avkiran, N.K., 2018. Measuring the Systemic Risk of Regional Banks in Japan with PLS-SEM. Theoretical Economics Letters 8, 2024–2037.
Awotwe-Otoo, D., Agarabi, C., Read, E.K., Lute, S., Brorson, K.A., Khan, M.A., Shah, R.B., 2013. Impact of controlled ice nucleation on process performance and quality attributes of a lyophilized monoclonal antibody. Int J Pharm 450, 70–78.
Awotwe-Otoo, D., Agarabi, C., Wu, G.K., Casey, E., Read, E., Lute, S., Brorson, K.A., Khan, M.A., Shah, R.B., 2012. Quality by design: impact of formulation variables and their interactions on quality attributes of a lyophilized monoclonal antibody. Int J Pharm 438, 167–175.
Balestrieri, C., Colonna, G., Giovane, A., Irace, G., Servillo, L., 1978. Second-Derivative Spectroscopy of Proteins: A Method for the Quantitative Determination of Aromatic Amino Acids in Proteins. search.ebscohost.comC Balestrieri, G Colonna, A Giovane, G Irace, L ServilloEuropean journal of biochemistry, 1978•search.ebscohost.com 90, 433–440.
Bayat, M., Marín-García, M., Ghasemi, J.B., Tauler, R., 2020. Application of the area correlation constraint in the MCR-ALS quantitative analysis of complex mixture samples. Anal Chim Acta 1113, 52–65.
Beech, K.E., Biddlecombe, J.G., van der Walle, C.F., Stevens, L.A., Rigby, S.P., Burley, J.C., Allen, S., 2015. Insights into the influence of the cooling profile on the reconstitution times of amorphous lyophilized protein formulations. Eur J Pharm Biopharm 96, 247–254.
Beechem, J.M., Brand, L., 1985. Time-resolved fluorescence of proteins. Annu Rev Biochem VOL. 54, 43–71.
Berns, R., 2019. Billmeyer and Saltzman's principles of color technology.
Berry, B.N., Dobrowsky, T.M., Timson, R.C., Kshirsagar, R., Ryll, T., Wiltberger, K., 2016. Quick generation of Raman spectroscopy based in-process glucose control to influence biopharmaceutical protein product quality during mammalian cell culture. Biotechnol Prog 32, 224–234.
Berry, M.W., 1992. Large-Scale Sparse Singular Value Computations. http://dx.doi.org/10.1177/109434209200600103 6, 13–49, Accessed on 13/02/2024.
Bhambhani, A., Blue, J.T., 2010. Lyophilization strategies for development of a high-concentration monoclonal antibody formulation: benefits and pitfalls. Am Pharm Rev 13, 31–38.
Bhartia, R., Hug, W.F., Salas, E.C., Reid, R.D., Suapati, K.K., Tsapin, A., Abbey, W., Nealson, K.H., Lane, A.L., Conrad, P.G., 2008. Classification of organic and biological materials with deep ultraviolet excitation. Appl Spectrosc 62, 1070–1077.
Bhatia, H., Mehdizadeh, H., Drapeau, D., Yoon, S., 2018. In-line monitoring of amino acids in mammalian cell cultures using raman spectroscopy and multivariate chemometrics models. Eng Life Sci 18, 55–61.
Bhattacharya, P.K., Burman, P., 2016. Multivariate Analysis. Theory and Methods of Statistics 383–429.
Bischl, B., Lang, M., Bossek, J., Horn, D., Richter, J., Surmann, D., 2017. BBmisc: Miscellaneous Helper Functions for B. Bischl.
Blue, J., Yoder, H., 2009. LYOPHILIZATION-Successful Lyophilization Development of Protein Therapeutics. Am Pharm Rev 12, 90.

Borzova, V.A., Markossian, K.A., Chebotareva, N.A., Kleymenov, S.Y., Poliansky, N.B., Muranov, K.O., Stein-Margolina, V.A., Shubin, V. V., Markov, D.I., Kurganov, B.I., 2016. Kinetics of Thermal Denaturation and Aggregation of Bovine Serum Albumin. PLoS One 11.

Brenner, V., 2015. Causes of supply chain disruptions: An empirical analysis in cold chains for food and pharmaceuticals, Causes of Supply Chain Disruptions: An Empirical Analysis in Cold Chains for Food and Pharmaceuticals.

Burke, C.J., Steadman, B.L., Volkin, D.B., Tsai, P.K., Bruner, M.W., Middaugh, C.R., 1992. The adsorption of proteins to pharmaceutical container surfaces. Int J Pharm 86, 89–93.

Burstein, E.A., Vedenkina, N.S., Ivkova, M.N., 1973. Fluorescence and the location of tryptophan residues in protein molecules. Photochem Photobiol 18, 263–279.

Butreddy, A., Janga, K.Y., Ajjarapu, S., Sarabu, S., Dudhipala, N., 2021. Instability of therapeutic proteins — An overview of stresses, stabilization mechanisms and analytical techniques involved in lyophilized proteins. Int J Biol Macromol 167, 309–325.

Butterfield, D., Stadtman, E., 1997. Protein oxidation processes in aging brain. Elsevier.

Cabra, V., Vázquez-Contreras, E., Moreno, A., Arreguin-Espinosa, R., 2008. The effect of sulfhydryl groups and disulphide linkage in the thermal aggregation of Z19 α-zein. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1784, 1028–1036.

Calloni, G., Lendel, C., Campioni, S., Giannini, S., Gliozzi, A., Relini, A., Vendruscolo, M., Dobson, C.M., Salvatella, X., Chiti, F., 2008. Structure and dynamics of a partially folded protein are decoupled from its mechanism of aggregation. J Am Chem Soc 130, 13040–13050.

Camilloni, C., Guerini Rocco, A., Eberini, I., Gianazza, E., Broglia, R.A., Tiana, G., 2008. Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. Biophys J 94, 4654–4661.

Campbell, F.W., Green, D.G., 1965. Optical and retinal factors affecting visual resolution. J Physiol 181, 576–593.

Canet, D., Doering, K., Dobson, C.M., Dupont, Y., 2001. High-Sensitivity Fluorescence Anisotropy Detection of Protein-Folding Events: Application to α-Lactalbumin. Biophys J 80, 1996–2003.

Cao, W., Krishnan, S., Ricci, M.S., Shih, L.-Y., Liu, D., Gu, J.H., Jameel, F., 2013. Rational design of lyophilized high concentration protein formulations-mitigating the challenge of slow reconstitution with multidisciplinary strategies. European Journal of Pharmaceutics and Biopharmaceutics 85, 287–293.

Cappius, H.J., Börner, J., Südmeyer, H., Schmitt, F.J., 2015. Fluorescence-optical handheld non-contact sensor for rapid cleaning validation of surfaces. Photonics Lasers Med 4, 169–175.

Carpenter, J.F., Pikal, M.J., Chang, B.S., Randolph, T.W., 1997. Rational design of stable lyophilized protein formulations: some practical advice. Pharm Res 14, 969–975.

Chandran, S., Nashine, V., Morcos, M., Saluja, A., 2013. A small-scale, material-saving approach to rank-order lyophilized formulations based on reconstitution time. J Pharm Innov 8, 187–194.

Chansanroj, K., Thanawattanawanich, P., 2015. Lyophilization of pharmaceutical products: From concept to reality.

Chen, H., Rhoades, E., 2008. Fluorescence characterization of denatured proteins. Curr Opin Struct Biol.

Chen, Y., Parrini, C., Taddei, N., Lapidus, L.J., 2009. Conformational properties of unfolded HypF-N. J Phys Chem B 113, 16209–16213.
Chi, E.Y., Krishnan, S., Randolph, T.W., Carpenter, J.F., 2003. Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. Pharm Res 20, 1325–1336.
Chodankar, S., Aswal, V.K., Kohlbrecher, J., Vavrin, R., Wagh, A.G., 2009. Small-angle neutron scattering study of structure and kinetics of temperature-induced protein gelation. Phys Rev E Stat Nonlin Soft Matter Phys 79.
Chrysochoos, J., 1974. Concentration Fluorescence Quenching of L(-)-Tyrosine in Aqueous and Nonaqueous Solutions. Spectroscopy Letters 7, 235–246.
Chullipalliyalil, K., Elkassas, K., McAuliffe, M.A.P., Vucen, S., Crean, A., 2023. In-Vial Detection of Protein Denaturation Using Intrinsic Fluorescence Anisotropy. Anal Chem 95, 2774–2782.
Chullipalliyalil, K., Lewis, L., McAuliffe, M., 2019. Deep UV laser-Induced fluorescence for pharmaceutical cleaning validation. ACS Publications - Analytical Chemistry 92, 1447–1454.
Clarke, R.J., Oprysa, A., 2004. Fluorescence and light scattering. J Chem Educ 81, 705–707.
Colandene, J.D., Maldonado, L.M., Creagh, A.T., Vrettos, J.S., Goad, K.G., Spitznagel, T.M., 2007. Lyophilization Cycle Development for a High-Concentration Monoclonal Antibody Formulation Lacking a Crystalline Bulking Agent. J Pharm Sci 96, 1598–1608.
Cook, I.A., Ward, K.R., 2011. Headspace moisture mapping and the information that can be gained about freeze-dried materials and processes. PDA J Pharm Sci Technol 65, 457–467.
Cui, Y., Cui, P., Chen, B., Li, S., Guan, H., 2017. Monoclonal antibodies: formulations of marketed products and recent advances in novel delivery system. Drug Dev Ind Pharm 43, 519–530.
Davies, K.J.A., Delsignores, M.E., Lin, S.W., 1987. Protein Damage and Degradation by Oxygen Radicals 11. Modification of amino acids*. Journal of Biological Chemistry 262, 9902–9907.
Davies, M.J., 2005. The oxidative environment and protein damage. Biochim Biophys Acta 1703, 93–109.
Davis, K., 2008. Biopharmaceutical manufacturing: The challenge of global regulatory compliance. Pharmaceutical Technology.
De Juan, A., Jaumot, J., Tauler, R., 2014. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. Analytical Methods 6, 4964–4976.
Deprettere, E.F., 1989. SVD and signal processing: algorithms, applications, and architectures, Volumen 1 43–91.
Deutscher, M.P., 1990. [8] Maintaining protein stability. In: Methods in Enzymology. Elsevier, pp. 83–89.
Dias, C., Chan, H., 2014. Pressure-dependent properties of elementary hydrophobic interactions: ramifications for activation properties of protein folding. The Journal of Physical Chemistry ACS Publications 118, 7488–7509.
Diederich, P., Hansen, S.K., Oelmeier, S.A., Stolzenberger, B., Hubbuch, J., 2011. A sub-two minutes method for monoclonal antibody-aggregate quantification using parallel

interlaced size exclusion high performance liquid chromatography. J Chromatogr A 1218, 9010–9018.
Dill, K.A., 1990. Dominant Forces in Protein Folding. Biochemistry 29, 7133–7155.
Dobson, C.M., 2003. Protein folding and misfolding. Nature 426, 884–890.
dos Santos Rodrigues, F.H., Delgado, G.G., Santana da Costa, T., Tasic, L., 2023. Applications of fluorescence spectroscopy in protein conformational changes and intermolecular contacts. BBA Advances 3, 100091.
Draper, N.R., Smith, H., 1998. R-square statistic. Applied regression analysis 33.
Dwevedi, A., Dubey, V.K., Jagannadham, M. V., Kayastha, A.M., 2010. Insights into pH-induced conformational transition of β-galactosidase from pisum sativum leading to its multimerization. Appl Biochem Biotechnol 162, 2294–2312.
Ecker, D.M., Jones, S.D., Levine, H.L., 2015. The therapeutic monoclonal antibody market. MAbs 7, 9–14.
Edelhoch, H., 1967. Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. Biochemistry 6, 1948–1954.
Edsall, J.T., McKenzie, H.A., 1983. Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. Adv Biophys 16, 53–183.
Eftink, M.R., 1991. Fluorescence techniques for studying protein structure. Methods Biochem Anal 35, 127–205.
Eftink, M.R., 1994. The Use of Fluorescence Methods to Monitor Unfolding Transitions in Proteins. Biophys J 66, 482–501.
ElKassas, K., Chullipalliyalil, K., McAuliffe, M., Vucen, S., Crean, A., 2021. Fluorescence spectroscopy for the determination of reconstitution time of an in-vial lyophilised product. Int J Pharm 120368.
ElKassas, K.&, Chullipalliyalil, K., McAuliffe, M., Vucen, S., Crean, A., 2021. Fluorescence spectroscopy for the determination of reconstitution time of an in-vial lyophilised product. Int J Pharm 597, 120368.
Elkhabaz, A., Sarkar, S., Simpson, G., Taylor, L., 2019. Characterization of Phase Transformations for Amorphous Solid Dispersions of a Weakly Basic Drug upon Dissolution in Biorelevant Media. Pharm Res 36.
Elliott, D.F., 1987. Handbook of digital signal processing : engineering applications. Academic Press.
EMA, 2013. Committee for Medicinal Products for Human Use (CHMP) Paediatric Committee (PDCO) Guideline on pharmaceutical development of medicines for paediatric use Guideline on pharmaceutical development of medicines for paediatric use.
Esbensen, K.H., Geladi, P., 2010. Principles of Proper Validation: use and abuse of re-sampling for validation. J Chemom 24, 168–187.
Fainerman, V.B., Miller, R., 2005. Equilibrium and dynamic characteristics of protein adsorption layers at gas-liquid interfaces: Theoretical and experimental data. Colloid Journal 67, 393–404.
Federici, M.M., 1994. The Quality Control of Biotechnology Products. Biologicals 22, 151–159.
Feitelson, J., 1964. On the mechanism of fluorescence quenching. Tyrosine and similar compounds. Journal of Physical Chemistry 68, 391–397.

Fink, J.L., Belle, A.L.; P., Marcus, J., Gjerde, R.A., Schmuchler, D.T., Fritz, G., 1990. Mechanism of Acid-Induced Folding of Proteins1". Proc. Natl. Acad. Sci. U.S.A 29, 35.
Formoso, C., Forster, L.S., 1975. Tryptophan Fluorescence Lifetimes in Lysozyme*. Journal of Biological Chemistry 250, 3738–3745.
Freifelder, D., 1982. Physical biochemistry : applications to biochemistry and molecular biology 761.
Geidobler, R., Konrad, I., Winter, G., 2013. Can controlled ice nucleation improve freeze-drying of highly-concentrated protein formulations? J Pharm Sci 102, 3915–3919.
Geisser, S., 1975. The Predictive Sample Reuse Method with Applications. J Am Stat Assoc 70, 320.
Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal Chim Acta 185, 1–17.
Gervasi, V, Dall Agnol, R., Cullen, S., McCoy, T., Vucen, S., Crean, A., 2018. Parenteral protein formulations: An overview of approved products within the European Union. Eur J Pharm Biopharm 131, 8–24.
Gervasi, V., Dall Agnol, R., Cullen, S., McCoy, T., Vucen, S., Crean, A., 2018. Parenteral protein formulations: An overview of approved products within the European Union. European Journal of Pharmaceutics and Biopharmaceutics.
Gething, M.J., Sambrook, J., 1992. Protein folding in the cell. Nature 355, 33–45.
Giese, A.T., French, C.S., 1955. The Analysis of Overlapping Spectral Absorption Bands by Derivative Spectrophotometry. Applied Spectroscopy, Vol. 9, Issue 2, pp. 78-96 9, 78–96.
Gijsbers, A., Nishigaki, T., Sánchez-Puig, N., 2016. Fluorescence anisotropy as a tool to study protein-protein interactions. JoVE.
Gill, S., Von Hippel, P., 1989. Calculation of protein extinction coefficients from amino acid sequence data. Analytical biochemistry, .
Giurleo, J.T., He, X., Talaga, D.S., 2008. β-Lactoglobulin Assembles into Amyloid through Sequential Aggregated Intermediates. J Mol Biol 381, 1332–1348.
Gokulakrishnan, P., Lawrence, A.D., McLellan, P.J., Grandmaison, E.W., 2006. A functional-PCA approach for analyzing and reducing complex chemical mechanisms. Comput Chem Eng 30, 1093–1101.
Goto, Y., Calciano, L.J., Fink, A.L., 1990. Acid-induced folding of proteins. Proceedings of the National Academy of Sciences 87, 573–577.
Grigolato, F., Arosio, P., 2019. Sensitivity analysis of the variability of amyloid aggregation profiles. Phys. Chem. Chem. Phys. 21, 1435–1442.
Grimm, L., Yarnold, P., 1995. Reading and understanding multivariate statistics.
Grossman, S.H., 1984. Fluorescence analysis of denaturation and reassembly of dansylated creatine kinase. Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular.
Grum, F., Paine, D., Zoeller, L., 1972. Derivative Absorption and Emission Spectrophotometry. Applied Optics, Vol. 11, Issue 1, pp. 93-98 11, 93–98.
Gryczynski, I., Wiczk, W., Johnson, M.L., Lakowicz, J.R., 1988. Lifetime distributions and anisotropy decays of indole fluorescence in cyclohexane/ethanol mixtures by frequency-domain fluorometry. Biophys Chem 32, 173–185.
Guckeisen, T., Hosseinpour, S., Peukert, W., 2021. Effect of pH and urea on the proteins secondary structure at the water/air interface and in solution. J Colloid Interface Sci 590, 38–49.

Guo, W., Shea, J.E., Berry, R.S., 2005. The physics of the interactions governing folding and association of proteins. Ann N Y Acad Sci 1066, 34–53.
Gupta, A., Mahalakshmi, R., 2019. Helix-strand interaction regulates stability and aggregation of the human mitochondrial membrane protein channel VDAC3. Journal of General Physiology 151, 489–504.
Gupta, A., Mahalakshmi, R., 2020. Single-residue physicochemical characteristics kinetically partition membrane protein self-assembly and aggregation. Journal of Biological Chemistry 295, 1181–1194.
Hageman, M.J., 1988. The role of moisture in protein stability. Drug Dev Ind Pharm 14, 2047–2070.
Hales, J.E., Aoudjane, S., Aeppli, G., Dalby, P.A., 2021. Proof-of-concept analytical instrument for label-free optical deconvolution of protein species in a mixture. J Chromatogr A 1641, 461968.
Hammes, G.G., Hammes, G.G., 2005. Spectroscopy for the biological sciences 172.
Hanrahan, P., Krueger, W., 1993. Reflection from layered surface due to subsurface scattering. Proc ACM SIGGRAPH 93 Conf Comput Graphics 165–174.
Harkes, R., Kukk, O., Mukherjee, S., Klarenbeek, J., van den Broek, B., Jalink, K., 2021. Dynamic FRET-FLIM based screening of signal transduction pathways. Sci Rep 11.
Harris, S.A., Patel, B.A., Gospodarek, A., Desai, J., de Janon Gutiérrez, A., Botonjic-Sehic, E., Brower, M., Pinto, N.D.S., 2021. Determination of protein concentration in downstream biomanufacturing processes by in-line index of refraction. Biotechnol Prog 37, e3187.
Harvey, D.T., Hanson, B.A., 2022. Understanding Scores and Loadings. CRAN.
Hawkins, B.K., Honigs, D.E., 1987. A comparison of spectroscopic techniques for protein quantification in aqueous-solutions. Am Biotechnol Lab 5, 26.
Heinisch, O., 1962. Steel, R. G. D., and J. H. Torrie: Principles and Procedures of Statistics. (With special Reference to the Biological Sciences.) McGraw-Hill Book Company, New York, Toronto, London 1960, 481 S., 15 Abb.; 81 s 6 d. Biom Z 4, 207–208.
Hellmann, N., Schneider, D., 2019. Hands On: Using Tryptophan Fluorescence Spectroscopy to Study Protein Structure. Methods in Molecular Biology 1958, 379–401.
Herrera, M.G., Veuthey, T. V., Dodero, V.I., 2016. Self-organization of gliadin in aqueous media under physiological digestive pHs. Colloids Surf B Biointerfaces 141, 565–575.
Hill, R.E., Bogdan, G.M., Dart, R.C., 2001. Time to reconstitution: purified Fab antivenom vs. unpurified IgG antivenom. Toxicon 39, 729–731.
Hinz, D.C., 2006. Process analytical technologies in the pharmaceutical industry: The FDA's PAT initiative. Anal Bioanal Chem 384, 1036–1042.
Hofmeister, F., 1888. Zur Lehre von der Wirkung der Salze - Zweite Mittheilung. Archiv für Experimentelle Pathologie und Pharmakologie 24, 247–260.
Hong, S., Choi, D.W., Kim, H.N., Park, C.G., Lee, W., Park, H.H., 2020. Protein-Based Nanoparticles as Drug Delivery Systems. Pharmaceutics 12, 1–28.
Höskuldsson, A., 1988. PLS regression methods. J Chemom 2, 211–228.
Hug, W.F., Bhartia, R., Sijapati, K., Beegle, L.W., Reid, R.D., 2014. Improved sensing using simultaneous deep-UV Raman and fluorescence detection-II. In: Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XV.

Hui, C.H., Triandis, H.C., 1985. Measurement in Cross-Cultural Psychology. http://dx.doi.org/10.1177/0022002185016002001 16, 131–152, Accessed on 13/02/2024.
Huynh, K., Partch, C.L., 2015. Current Protocols in Protein Science: Analysis of protein stability and ligand interactions by thermal shift assay. Current protocols in protein science / editorial board, John E. Coligan ... [et al.] 79, 28.9.1.
Ichikawa, T., Terada, H., 1977. Second derivative spectrophotometry as an effective tool for examining phenylalanine residues in proteins. Biochimica et Biophysica Acta.
Ichikawa, T., Terada, H., 1979. Estimation of state and amount of phenylalanine residues in proteins by second derivative spectrophotometry. Biochimica et Biophysica Acta - (BBA)-Protein, .
Jaenicke, R., 1993. What does protein refolding in vitro tell us about protein folding in the cell? Philos Trans R Soc Lond B Biol Sci 339.
Jain, S., Heiser, A., Venter, A.R., 2011. Spray desorption collection: An alternative to swabbing for pharmaceutical cleaning validation. Analyst.
James, N.A., Kejariwal, A., Matteson, D.S., 2014. BreakoutDetection: Breakout Detection via Robust E-Statistics.
Jansson, P.A., 1984. Deconvolution : with applications in spectroscopy 342.
Jaumot, J., Gargallo, R., De Juan, A., Tauler, R., 2005. Software description A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. Chemometrics and Intelligent Laboratory Systems 76, 101–110.
Johnson, R.E., Oldroyd, M.E., Ahmed, S.S., Gieseler, H., Lewis, L.M., 2010. Use of Manometric Temperature Measurements (MTM) to Characterize the Freeze-Drying Behavior of Amorphous Protein Formulations. J Pharm Sci 99, 2863–2873.
Johnson, W.C., 1999. Analyzing Protein Circular Dichroism Spectra for Accurate Secondary Structures. Proteins 35, 307–312.
Johnston, M.J.W., Nemr, K., Hefford, M.A., 2010. Influence of bovine serum albumin on the secondary structure of interferon alpha 2b as determined by far UV circular dichroism spectropolarimetry. Biologicals 38, 314–320.
Johnston, S.L., 2007. Biologic therapies: what and when? J Clin Pathol 60, 8.
Jollife, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci 374.
Jones, L.S., Bam, N.B., Randolph, T.W., 1997. Surfactant-Stabilized Protein Formulations: A Review of Protein-Surfactant Interactions and Novel Analytical Methodologies. In: Therapeutic Protein and Peptide Formulation and Delivery. American Chemical Society, pp. 206–222.
Kamga, M.H., Woo Lee, H., Liu, J., Yoon, S., 2013. Quantification of protein mixture in chromatographic separation using multi-wavelength UV spectra. Biotechnol Prog 29, 664–671.
Kanai, S., Liu, J., Patapoff, T.W., Shire, S.J., 2008. Reversible self-association of a concentrated monoclonal antibody solution mediated by Fab-Fab interaction that impacts solution viscosity. J Pharm Sci 97, 4219–4227.
Kato, K., Nakayoshi, T., Kurimoto, E., Oda, A., 2020. Mechanisms of Deamidation of Asparagine Residues and Effects of Main-Chain Conformation on Activation Energy. Int J Mol Sci 21, 1–14.

Kaylor, J., Bodner, N., Edridge, S., Yamin, G., Hong, D.P., Fink, A.L., 2005. Characterization of Oligomeric Intermediates in α-Synuclein Fibrillation: FRET Studies of Y125W/Y133F/Y136F α-Synuclein. J Mol Biol 353, 357–372.
Kehm, R., Baldensperger, T., Raupbach, J., Höhn, A., 2021. Protein oxidation - Formation mechanisms, detection and relevance as biomarkers in human diseases. Redox Biol 42.
Kejariwal, A., Lee, W., Vallis, O., Hochenbaum, J., Yan, B., 2013. Visual analytics framework for cloud infrastructure data. IEEE 16th International Conference on Computational Science.
Kelly, M.A., Vestling, M.M., Fenselau, C.C., Smith, P.B., 1992. Electrospray analysis of proteins: A comparison of positive-ion and negative-ion mass spectra at high and low pH. Organic Mass Spectrometry 27, 1143–1147.
Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1751, 119–139.
Khan, S.A.R., Razzaq, A., Yu, Z., Shah, A., Sharif, A., Janjua, L., 2022. Disruption in food supply chain and undernourishment challenges: An empirical study in the context of Asian countries. Socioecon Plann Sci.
Kim, Y.D., Morr, C. V., 1996. Dynamic headspace analysis of light activated flavor in milk. Int Dairy J 6, 185–193.
Kirch, W., 2008. Pearson's Correlation Coefficient. Encyclopedia of Public Health 1090–1091.
Kishore, D., Kayastha, A., 2012. A β-galactosidase from chick pea (Cicer arietinum) seeds: Its purification, biochemical properties and industrial applications. Food Chem.
Kishore, D., Kundu, S., Kayastha, A.M., 2012. Thermal, Chemical and pH Induced Denaturation of a Multimeric β-Galactosidase Reveals Multiple Unfolding Pathways. PLoS One 7.
Klijn, M.E., Hubbuch, J., 2021. Application of ultraviolet, visible, and infrared light imaging in protein-based biopharmaceutical formulation characterization and development studies. European Journal of Pharmaceutics and Biopharmaceutics 165, 319–336.
Kopac, T., Bozgeyik, K., Yener, J., 2008. Effect of pH and temperature on the adsorption of bovine serum albumin onto titanium dioxide. Colloids Surf A Physicochem Eng Asp.
Kramer, R.M., Shende, V.R., Motl, N., Pace, C.N., Scholtz, J.M., 2012. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. Biophys J 102, 1907.
Krebs, M.R.H., Devlin, G.L., Donald, A.M., 2007. Protein Particulates: Another Generic Form of Protein Aggregation? Biophys J 92, 1336.
Krebs, M.R.H., Domike, K.R., Donald, A.M., 2009. Protein aggregation: more than just fibrils. Biochem Soc Trans 37, 682–686.
Kroger, A., Bahta, L., Long, S., Sanchez, P., 2022. Advisory Committee on Immunization Practices (ACIP) General Best Guidance for Immunization. Centre for Disease Control (CDC).
Kucheryavskiy, S., 2020. mdatools – R package for chemometrics. Chemometrics and Intelligent Laboratory Systems 198, 103937.
Kueltzo, L.A., Ersoy, B., Ralston, J.P., Middaugh, C.R., 2003. Derivative absorbance spectroscopy and protein phase diagrams as tools for comprehensive protein characterization: A bGCSF case study. J Pharm Sci 92, 1805–1820.

Kulkarni, S.S., Patel, S.M., Bogner, R.H., 2020. Reconstitution Time for Highly Concentrated Lyophilized Proteins: Role of Formulation and Protein. J Pharm Sci 109, 2975–2985.
Kulkarni, S.S., Suryanarayanan, R., Rinella Jr., J. V, Bogner, R.H., 2018. Mechanisms by which crystalline mannitol improves the reconstitution time of high concentration lyophilized protein formulations. Eur J Pharm Biopharm 131, 70–81.
Lai, Z., McCulloch, J., Lashuel, H.A., Kelly, J.W., 1997. Guanidine hydrochloride-induced denaturation and refolding of transthyretin exhibits a marked hysteresis: Equilibria with high kinetic barriers. Biochemistry 36, 10230–10239.
Lakowicz, J.R., 1999. Principles of Fluorescence Spectroscopy. Principles of Fluorescence Spectroscopy.
Lapidus, L.J., 2017. Protein unfolding mechanisms and their effects on folding experiments. F1000Res 6, 1723.
Lawton, W.H., Sylvestre, E.A., 1971. Self Modeling Curve Resolution. Technometrics 13, 617.
Leach, S.J., Scheraga, H.A., 1960. Ultraviolet Difference Spectra and the Internal Structure of Proteins. OF BIOLOGICAL CHEMISTRY 235.
Lee, H.W., Christie, A., Liu, J.J., Yoon, S., 2012. Estimation of raw material performance in mammalian cell culture using near infrared spectra combined with chemometrics approaches. Biotechnol Prog 28, 824–832.
Levine, R.L., Marcia Federici, M., 1982. Quantitation of Aromatic Residues in Proteins: Model Compounds for Second-Derivative Spectroscopy. Biochemistry 21, 2600–2606.
Lewis, I., Edwards, H., 2001. Handbook of Raman spectroscopy: from the research laboratory to the process line.
Liland, K.H., Stefansson, P., Indahl, U.G., 2020. Much faster cross-validation in PLSR-modelling by avoiding redundant calculations. J Chemom 34, e3201.
Liu, Y., Zhang, C., Chen, J., Fernandez, J., Vellala, P., Kulkarni, T.A., Aguilar, I., Ritz, D., Lan, K., Patel, P., Liu, A., 2022. A Fully Integrated Online Platform For Real Time Monitoring Of Multiple Product Quality Attributes In Biopharmaceutical Processes For Monoclonal Antibody Therapeutics. J Pharm Sci 111, 358–367.
Lobley, A., Whitmore, L., Wallace, B.A., 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. Bioinformatics 18, 211–212.
Lu, J.R., Su, T.J., Thomas, R.K., 1999. Structural Conformation of Bovine Serum Albumin Layers at the Air–Water Interface Studied by Neutron Reflection. J Colloid Interface Sci 213, 426–437.
Mach, H., Thomson, J., Middaugh, C., 1991. Examination of phenylalanine microenvironments in proteins by second-derivative absorption spectroscopy. Archives of biochemistry - Elsevier.
Mahler, H.-C., Friess, W., Grauschopf, U., Kiese, S., 2009. Protein aggregation: Pathways, induction factors and analysis. J Pharm Sci 98, 2909–2934.
Mahler, H.C., Jiskoot, W., 2012. Analysis of Aggregates and Particles in Protein Pharmaceuticals. Analysis of Aggregates and Particles in Protein Pharmaceuticals.
Malik, A., De Juan, A., Tauler, R., 2015. Multivariate curve resolution: A different way to examine chemical data. ACS Symposium Series 1199, 95–128.

Manavalan, P., Johnson, W.C., 1987. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Anal Biochem 167, 76–85.
Mann, T.L., Krull, U.J., 2003. Fluorescence polarization spectroscopy in protein analysis. Analyst.
Manning, M.C., Chou, D.K., Murphy, B.M., Payne, R.W., Katayama, D.S., 2010. Stability of Protein Pharmaceuticals: An Update. Pharm Res 27, 544–575.
Margineanu, A., Chan, J.J., Kelly, D.J., Warren, S.C., Flatters, D., Kumar, S., Katan, M., Dunsby, C.W., French, P.M.W., 2016. Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM). Scientific Reports 2016 6:1 6, 1–17.
Matthias, D.M., Robertson, J., Garrison, M.M., Newland, S., Nelson, C., 2007. Freezing temperatures in the vaccine cold chain: A systematic literature review. Vaccine.
May, J.C., 2004. Regulatory control of freeze-dried products: Importance and evaluation of residual moisture. pp. 349–384.
McCreery, R., 2005. Raman spectroscopy for chemical analysis.
McGuire, R., Feldman, I., 1973. The quenching of tyrosine and tryptophan fluorescence by H2O and D2O. Photochem Photobiol 18, 119–124.
Meech, S., 2009. Excited state reactions in fluorescent proteins. Chemical Society Reviews.
Mei, G., Rosato, N., Silva, N., Rusch, R., Gratton, E., Savini, I., Finazzi-Agrd, A., Roma, di, Vergata, T., 1992. 7224-7230 proteins (Kuwajima, 1977; Dol-gikh etal. Biochemistry 31.
Mercade-Prieto, R., Gunasekaran, S., 2009. Alkali cold gelation of whey proteins. Part II: Protein concentration. Langmuir 25, 5793–5801.
Messinis, G., Hatziargyriou, N., 2018. Unsupervised classification for non-technical loss detection. Power Systems Computation Conference (PSCC).
Michalet, X., Weiss, S., Jäger, M., 2006. Single-molecule fluorescence studies of protein folding and conformational dynamics. Chem Rev 106, 1785–1813.
Michnik, A., Michalik, K., Drzazga, Z., 2008. Effect of UVC radiation on conformational restructuring of human serum albumin. J Photochem Photobiol B 90, 170–178.
Miles, A.J., Ramalli, S.G., Wallace, B.A., 2022. DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. Protein Science 31, 37–46.
Millar, D.P., 1996. Fluorescence studies of DNA and RNA structure and dynamics. Curr Opin Struct Biol 6, 322–326.
Mittag, J.J., Rädler, J.O., McManus, J.J., 2018. Peptide self-assembly measured using fluorescence correlation spectroscopy. Methods in Molecular Biology 1777, 159–171.
Modler, A.J., Gast, K., Lutsch, G., Damaschun, G., 2003. Assembly of amyloid protofibrils via critical oligomers - A novel pathway of amyloid formation. J Mol Biol 325, 135–148.
Monsellier, E., Bedouelle, H., 2005. Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. Protein Engineering, Design and Selection 18, 445–456.
Moran, F., 2001. SOMCD: method for evaluating protein secondary structure from UV circular dichroism spectra. … Structure, Function, and ….
Moriyama, Y., Ohta, D., Hachiya, K., Mitsui, Y., Takeda, K., 1996. Fluorescence behavior of tryptophan residues of bovine and human serum albumins in ionic surfactant solutions:

A comparative study of the two and one tryptophan(s) of bovine and human albumins. J Protein Chem 15, 265–272.
Motegi, H., Tsuboi, Y., Saga, A., Kagami, T., Inoue, M., Toki, H., Minowa, O., Noda, T., Kikuchi, J., 2015. Identification of Reliable Components in Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS): a Data-Driven Approach across Metabolic Processes. Scientific Reports 2015 5:1 5, 1–12.
Nafie, L., 2001. Theory of Raman scattering. Practical Spectroscopy Series.
Nafie, L., 2008. Theory of Raman scattering and Raman optical activity: near resonance theory and levels of approximation. Theoretical Chemistry Accounts - Springer 119, 39–55.
Natalello, A., Prokorov, V. V., Tagliavini, F., Morbin, M., Forloni, G., Beeg, M., Manzoni, C., Colombo, L., Gobbi, M., Salmona, M., Doglia, S.M., 2008. Conformational plasticity of the Gerstmann-Sträussler-Scheinker disease peptide as indicated by its multiple aggregation pathways. J Mol Biol 381, 1349–1361.
Neagu, A., Neagu, M., Dé, A.S., 2001. Fluctuations and the Hofmeister Effect.
Nielsen, M.M., Andersen, K.K., Westh, P., Otzen, D.E., 2007. Unfolding of β-Sheet Proteins in SDS. Biophys J 92, 3674.
Noble, J.E., Bailey, M.J.A., 2009. Chapter 8 - Quantitation of Protein. Methods Enzymol 463, 73–95.
Olivieri, A.C., Escandar, G.M., 2014. Application Example: PARAFAC. Practical Three-Way Calibration 233–250.
Otzen, D., 2011. Protein–surfactant interactions: A tale of many states. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1814, 562–591.
Otzen, D.E., 2002. Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. Biophys J 83, 2219–2230.
Otzen, D.E., Christiansen, L., Schülein, M., 1999. A comparative study of the unfolding of the endoglucanase Cel45 from Humicola insolens in denaturant and surfactant. Protein Sci 8, 1878.
Overcashier, D.E., Patapoff, T.W., Hsu, C.C., 1999. Lyophilization of protein formulations in vials: Investigation of the relationship between resistance to vapor flow during primary drying and small-scale product collapse. J Pharm Sci 88, 688–695.
Parastar, H., Tauler, R., 2014. Multivariate curve resolution of hyphenated and multidimensional chromatographic measurements: A new insight to address current chromatographic challenges. Anal Chem 86, 286–297.
Park, J.H., Jackman, J.A., Ferhan, A.R., Ma, G.J., Yoon, B.K., Cho, N.J., 2018. Temperature-Induced Denaturation of BSA Protein Molecules for Improved Surface Passivation Coatings. ACS Appl Mater Interfaces.
Parui, S., Manna, R.N., Jana, B., 2016. Destabilization of Hydrophobic Core of Chicken Villin Headpiece in Guanidinium Chloride Induced Denaturation: Hint of π-Cation Interaction. J Phys Chem B 120, 9599–9607.
Patel, S.M., Nail, S.L., Pikal, M.J., Geidobler, R., Winter, G., Hawe, A., Davagnino, J., Rambhatla Gupta, S., 2017. Lyophilized Drug Product Cake Appearance: What Is Acceptable? J Pharm Sci 106, 1706–1721.
Patonay, G., Beckford, G., Hänninen, P., 2014. UV-Vis and NIR Fluorescence Spectroscopy. Handbook of Spectroscopy: Second, Enlarged Edition 3–4, 999–1036.

Perera, K.D.C., Weragoda, G.K., Haputhanthri, R., Rodrigo, S.K., 2021. Study of concentration dependent curcumin interaction with serum biomolecules using ATR-FTIR spectroscopy combined with Principal Component Analysis (PCA) and Partial Least Square Regression (PLS-R). Vib Spectrosc 116, 103288.
Perrot, P., 1998. A to Z of Thermodynamics. A to Z of Thermodynamics.
Perucho, J., Gonzalo-Gobernado, · R, Bazan, · E, Casarejos, · M J, Jiménez-Escrig, · A, Asensio, M.J., Herranz, · A S, 2015. Optimal excitation and emission wavelengths to analyze amino acids and optimize neurotransmitters quantification using precolumn OPA-derivatization by HPLC. SpringerJ Perucho, R Gonzalo-Gobernado, E Bazan, MJ Casarejos, A Jiménez-Escrig, MJ AsensioAmino Acids, 2015•Springer 47, 963–973.
Picard, R.R., Cook, R.D., 1984. Cross-Validation of Regression Models. J Am Stat Assoc 79, 575.
Pignataro, M.F., Herrera, M.G., Dodero, V.I., 2020. Evaluation of Peptide/Protein Self-Assembly and Aggregation by Spectroscopic Methods. Molecules 2020, Vol. 25, Page 4854 25, 4854.
Poklar, N., Vesnaver, G., Lapanje, S., 1994. Denaturation behavior of α-chymotrypsinogen A in urea and alkylurea solutions: Fluorescence studies. J Protein Chem.
Polen, T., 2006. Enhanced packaging and reformulations give new life to biologics. Biopharm Int.
Prajapati, S., Bhakuni, V., Babu, K.R., Jain, S.K., 1998. Alkaline unfolding and salt-induced folding of bovine liver catalase at high pH. Eur J Biochem 255, 178–184.
Preston, K.B., Randolph, T.W., 2021. Stability of lyophilized and spray dried vaccine formulations. Adv Drug Deliv Rev 171, 50–61.
Privalov, P.L., 1997. Thermodynamics of protein folding 29, 447–474.
Privalov, P.L., Gill, S.J., 1988. Stability of Protein Structure and Hydrophobic Interaction. In: Anfinsen, C.B., Edsall, J.T., Richards, F.M., Eisenberg, D.S. (Eds.), Advances in Protein Chemistry. Academic Press, pp. 191–234.
Queiroz, A.L.P., Rodrigues, M., Zeglinski, J., Crean, A.M., Sarraguça, M.C., Vucen, S., 2021. Determination of co-crystal phase purity by mid infrared spectroscopy and multiple curve resolution. Int J Pharm 595, 120246.
R Core Team, 2019. R: A Language and Environment for Statistical Computing.
Radner, W., Benesch, T., 2019. Age-related course of visual acuity obtained with ETDRS 2000 charts in persons with healthy eyes. Graefe's Archive for Clinical and Experimental Ophthalmology 257, 1295–1301.
Rambhatla, S., Obert, J.P., Luthra, S., Bhugra, C., Pikal, M.J., 2005. Cake Shrinkage During Freeze Drying: A Combined Experimental and Theoretical Study. Pharm Dev Technol 10, 33–40.
Randolph, T.W., Jones, L.S., 2002. Surfactant-Protein Interactions. In: Carpenter, J.F., Manning, M.C. (Eds.), Rational Design of Stable Protein Formulations: Theory and Practice. Springer US, Boston, MA, pp. 159–175.
Razinkov, V., Treuheit, M., Becker, G., 2013. Methods of High Throughput Biophysical Characterization in Biopharmaceutical Development. Curr Drug Discov Technol 10, 59–70.
Rencher, A.C., 1998. Multivariate statistical inference and applications 559.
Rencher, A.C., William, F.C., 2012. Methods of multivariate analysis: Third edition. Methods of Multivariate Analysis: Third Edition 1–758.

Roche, J., Caro, J.A., Norberto, D.R., Barthe, P., Roumestand, C., Schlessman, J.L., Garcia, A.E., García-Moreno, B.E., Royer, C.A., 2012. Cavities determine the pressure unfolding of proteins. Proc Natl Acad Sci U S A 109, 6945–6950.
Rodrigues, J., Prosinecki, V., Marrucho, I., 2011. Protein stability in an ionic liquid milieu: on the use of differential scanning fluorimetry. Physical Chemistry Chemical Physics.
Roostaee, A., Côté, S., Roucou, B., 2009. Aggregation and amyloid fibril formation induced by chemical dimerization of recombinant prion protein in physiological-like conditions. ASBMB - Journal of Biological Chemistry.
Rosenfeld, M., Leonova, V., Konstantinova, M., Razumovskii, S., 2009. Self-assembly of fibrin monomers and fibrinogen aggregation during ozone oxidation. Springer Biochemistry (Moscow) 74, 54461.
Rossi, A., Taylor, C., 2011. Analysis of protein-ligand interactions by fluorescence polarization. Nat Protoc.
Rupley, J.A., Gratton, E., Careri, G., 1983. Water and globular proteins. Trends Biochem Sci 8, 18–22.
Saleh, B.E.A., Teich, M.C., 1991. Fundamentals of Photonics.
Sane, P., Bogner, R.H., Bhatnagar, B., Tchessalov, S., 2020. Reconstitution of Highly Concentrated Lyophilized Proteins: Part 1 Amorphous Formulations. J Pharm Sci 109, 1681–1691.
Scopes, R.K., 1974. Measurement of protein by spectrophotometry at 205 nm. Anal Biochem 59, 277–282.
Seal, H.L., 1967. Studies in the History of Probability and Statistics. XV The historical development of the Gauss linear model. Biometrika 54, 1–24.
Sharma, V.K., Kalonia, D.S., 2003. Steady-State Tryptophan Fluorescence Spectroscopy Study to Probe Tertiary Structure of Proteins in Solid Powders. J Pharm Sci 92, 890–899.
Shaw, K.L., Grimsley, G.R., Yakovlev, G.I., Makarov, A.A., Pace, C.N., 2001. The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. Protein Sci 10, 1206.
Shire, S.J., 2015. 1 - Introduction. In: Shire, S.J. (Ed.), Monoclonal Antibodies. Woodhead Publishing, pp. 1–15.
Shire, S.J., Shahrokh, Z., Liu, J., 2004. Challenges in the development of high protein concentration formulations. J Pharm Sci 93, 1390–1402.
Shlens, J., 2014. A Tutorial on Principal Component Analysis. Machine Learning - arXiv.
Sijbesma, E., Somsen, B.A., Miley, G.P., Leijten-Van De Gevel, I.A., Brunsveld, L., Arkin, M.R., Ottmann, C., 2020. Fluorescence anisotropy-based tethering for discovery of protein–protein interaction stabilizers. ACS Publications 15, 3143–3148.
Simonian, M.H., Smith, J.A., 2006. Spectrophotometric and Colorimetric Determination of Protein Concentration. Curr Protoc Mol Biol 76, 10.1.1-10.1A.9.
Smith, J.P., Holahan, E.C., Smith, F.C., Marrero, V., Booksh, K.S., 2019. A novel multivariate curve resolution-alternating least squares (MCR-ALS) methodology for application in hyperspectral Raman imaging analysis. Analyst 144, 5425–5438.
Snatzke, G., 2000. Circular Dichroism: An Introduction. Circular Dichroism: Principles and Applications 1–35.
Snowman, J.W., 1997. Lyophilization. In: Goldberg, E. (Ed.), Handbook of Downstream Processing. Springer Netherlands, Dordrecht, pp. 203–234.

Spalding, K., Bonnier, F., Bruno, C., Blasco, H., Board, R., Benz-de Bretagne, I., Byrne, H.J., Butler, H.J., Chourpa, I., Radhakrishnan, P., Baker, M.J., 2018. Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy. Vib Spectrosc 99, 50–58.
Sreerama, N., Woody, R.W., 1993. A self-consistent method for the analysis of protein secondary structure from circular dichroism. Anal Biochem 209, 32–44.
Stefani, M., 2008. Protein folding and misfolding on surfaces. Int J Mol Sci 9, 2515–2542.
Step, 2017. Committee for Human Medicinal Products ICH guideline Q8 (R2) on pharmaceutical development.
Stevens, A., Ramirez-Lopez, L., 2013. An introduction to the prospectr package.
Strang, G., 2016. Introduction in linear algebra. Introduction to Linear Algebra 585.
Swinehart, D.F., 1962. The Beer-Lambert Law. J Chem Educ 39, 333–335.
Takata, T., Oxford, J., Demeler, B., science, K.L.-P., 2008, undefined, 2008. Deamidation destabilizes and triggers aggregation of a lens protein, βA3-crystallin. Wiley Online LibraryT Takata, JT Oxford, B Demeler, KJ LampiProtein science, 2008•Wiley Online Library 17, 1565–1575.
Tanford, C., 1968. Protein Denaturation. Part A. Characterization of the Denatured State. Adv Protein Chem 23, 121–217.
Tchessalov, S., Kantor, A., Ll, L., Warne, N.W., 2010. Slow dissolution method for reconstitution of lyophilised material.
Thakral, S., Sonje, J., Munjal, B., Suryanarayanan, R., 2021. Stabilizers and their interaction with formulation components in frozen and freeze-dried protein formulations. Adv Drug Deliv Rev 173, 1–19.
The UniProt Consortium, 2023. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res 51, D523–D531.
The United States pharmacopeia, 2014. <1> Injections. USP 37-NF32.
Thermo Scientific, 2012. T124-TECHNICAL BULLETIN Influence of Buffer on Protein Quantification at 280 nm.
Tirivangani, T., Alpo, B., Kibuule, D., Gaeseb, J., Adenuga, B.A., 2021. Impact of COVID-19 pandemic on pharmaceutical systems and supply chain – a phenomenological study. Exploratory Research in Clinical and Social Pharmacy.
Tooke, P.B., 1988. Fourier self-deconvolution in IR spectroscopy. TrAC Trends in Analytical Chemistry 7, 130–136.
Tooke, P.B., James, D.I., 1985. Resolution Of Overlapped Bands By A Fourier Method. https://doi.org/10.1117/12.970816 0553, 291–292, Accessed on 13/02/2024.
Tsai, A.C., 2013. Achieving Consensus on Terminology Describing Multivariable Analyses. Am J Public Health 103, e1.
Uarrota, V.G., Moresco, R., Coelho, B., Nunes, E.D.C., Peruch, L.A.M., Neubert, E.D.O., Rocha, M., Maraschin, M., 2014. Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (Manihot esculenta Crantz) roots during postharvest physiological deterioration. Food Chem 161, 67–78.
Van Benthem, M.H., Keenan, M.R., 2004. Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. J Chemom 18, 441–450.
Van Buren, N., Rehder, D., Gadgil, H., Matsumura, M., Jacob, J., 2008. Elucidation of two major aggregation pathways in an IgG2 antibody. Elsevier 98, 3013–3030.

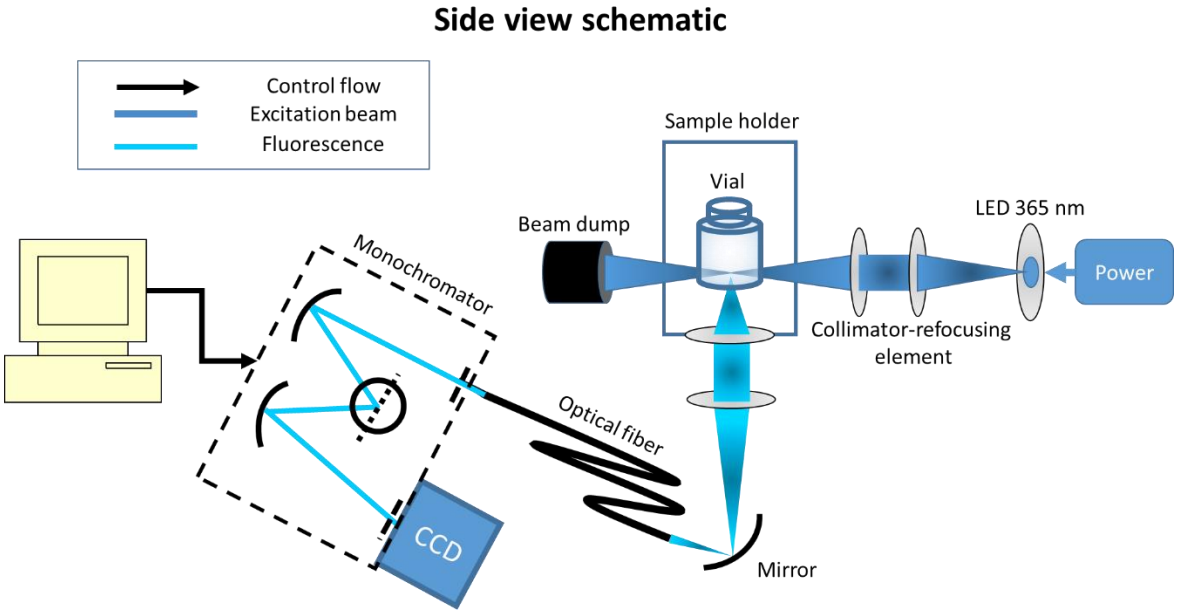
Varadhan, S.R.S., 1990. Prediction of Protein Structure and the Principles of Protein Conformation. Gerald D. Fasman, Ed. Plenum, New York, 1989. xiv, 798 pp., illus. $95. Science 247, 1351–1351.
Vlasova, I.M., Saletsky, A.M., 2009. Study of the denaturation of human serum albumin by sodium dodecyl sulfate using the intrinsic fluorescence of albumin. J Appl Spectrosc.
Wahl, V., Khinast, J., Paudel, A., 2016. Lyophilized protein powders: A review of analytical tools for root cause analysis of lot-to-lot variability. Trends in Analytical Chemistry 82, 468–491.
Wang, F., 2009. Factor Analysis and Principal-Components Analysis. International Encyclopedia of Human Geography 1–7.
Wang, Q., Dou, X., Chen, X., Zhao, Z., Wang, S., Wang, Y., Sui, K., Tan, Y., Gong, Y., Zhang, Y., Yuan, W.Z., 2019. Reevaluating Protein Photoluminescence: Remarkable Visible Luminescence upon Concentration and Insight into the Emission Mechanism. Angewandte Chemie 131, 12797–12803.
Wang, W., 1999. Instability, stabilization, and formulation of liquid protein pharmaceuticals. Int J Pharm 185, 129–188.
Wang, W., 2000. Lyophilization and development of solid protein pharmaceuticals. Int J Pharm 203, 1–60.
Wang, W., Nema, S., Teagarden, D., 2010. Protein aggregation—Pathways and influencing factors. Int J Pharm 390, 89–99.
Warburg, O., Christian, W., Warburg, O., Christian, W., 1941. Isolierung und Kristallisation des Gärungsferments Enolase. Naturwissenschaften 29, 589–590.
Wei, Y., Chen, L., Chen, J., Ge, L., He, R.Q., 2009. Rapid glycation with D-ribose induces globular amyloid-like aggregations of BSA with high cytotoxicity to SH-SY5Y cells. BMC Cell Biol 10.
Weiss, W.F., Young, T.M., Roberts, C.J., 2009. Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. J Pharm Sci 98, 1246–1277.
Wen, Z.Q., Cao, X., Phillips, J., 2010. Application of Raman Spectroscopy in Biopharmaceutical Manufacturing. Am Pharm Rev 13.
Werk, T., Huwyler, J., Hafner, M., Luemkemann, J., Mahler, H.-C., 2015. An Impedance-Based Method to Determine Reconstitution Time for Freeze-Dried Pharmaceuticals. J Pharm Sci 104, 2948–2955.
Whitmore, L., Wallace, B.A., 2008. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers 89, 392–400.
Wiemken, T., Furmanek, S., Mattingly, W., 2018. Methods for computational disease surveillance in infection prevention and control: statistical process control versus Twitter's anomaly and breakout detection. American Journal of Infection Control - Elsevier 46, 124–132.
Winogradoff, D., John, S., Aksimentiev, A., 2020. Protein unfolding by SDS: the microscopic mechanisms and the properties of the SDS-protein assembly. Nanoscale 12, 5422–5434.
Witteveen, M., Sterenborg, H.J.C.M., Leeuwen, T.G. van, Aalders, M.C.G., Ruers, T.J.M., Post, A.L., 2022. Comparison of preprocessing techniques to reduce nontissue-related variations in hyperspectral reflectance imaging. J Biomed Opt 27.

Wokaun, A., 1996. B. Schrader: Infrared and Raman Spectroscopy - Methods and Applications. VCH, Weinheim, 1995, DM 298,-, ISBN 3-527-26446-9. Berichte der Bunsengesellschaft für physikalische Chemie 100, 1268–1268.
Wold, H., 1966. Estimation of Principal Components and Related Models by Iterative Least Squares. [WWW Document]. Multivariate Analysis, Academic Press,. URL (accessed 11.30.23).
Wold, S., 1978. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. Technometrics 20, 397.
Wold, S., Ruhe, A., Wold, H., W. J. Dunn, I., 2006. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. https://doi.org/10.1137/0905052 5, 735–743, Accessed on 13/02/2024.
Wyatt, P.J., 1993. Light scattering and the absolute characterization of macromolecules. Anal Chim Acta 272, 1–40.
Xie, Z., Feng, X., Chen, X., 2022. Partial least trimmed squares regression. Chemometrics and Intelligent Laboratory Systems 221, 104486.
Xu, Q., Keiderling, T.A., 2004. Effect of sodium dodecyl sulfate on folding and thermal stability of acid-denatured cytochrome c: a spectroscopic approach. Protein Sci 13, 2949–2959.
Yang, A.-S., Honig, B., 1993. On the pH Dependence of Protein Stability. J Mol Biol 231, 459–474.
Yazdanparast, R., Esmaeili, M.A., Khodagholi, F., 2007. Control of aggregation in protein refolding: Cooperative effects of artificial chaperone and cold temperature. Int J Biol Macromol 40, 126–133.
Yohannes, G., Wiedmer, S.K., Elomaa, M., Jussila, M., Aseyev, V., Riekkola, M.L., 2010. Thermal aggregation of bovine serum albumin studied by asymmetrical flow field-flow fractionation. Anal Chim Acta.
Zhang, L., Sinha, S., Topp, E., 2010. Protein Conformation and Reactivity in Amorphous Solids. Formulation and Process Development Strategies for Manufacturing Biopharmaceuticals.
Zhao, R., So, M., Maat, H., Ray, N.J., Arisaka, F., Goto, Y., Carver, J.A., Hall, D., 2016. Measurement of amyloid formation by turbidity assay—seeing through the cloud. Biophysical Reviews 2016 8:4 8, 445–471.
Zhou, C., Cleland, D., Snell, J., Qi, W., Randolph, T.W., Carpenter, J.F., 2016. Formation of Stable Nanobubbles on Reconstituting Lyophilized Formulations Containing Trehalose. J Pharm Sci 105, 2249–2253.
Zimmermann, E., Muller, R.H., Mader, K., 2000. Influence of different parameters on reconstitution of lyophilized SLN. Int J Pharm 196, 211–213.

Appendix

S2. Supplementary Information Chapter 2

S2.1. Detailed set-up diagram and representative volume info



Supplementary Figure 2.1 Another schematic representation of the setup used to measure the UV fluorescence emission spectra of the constitution of a lyophilised formulation in-vial.

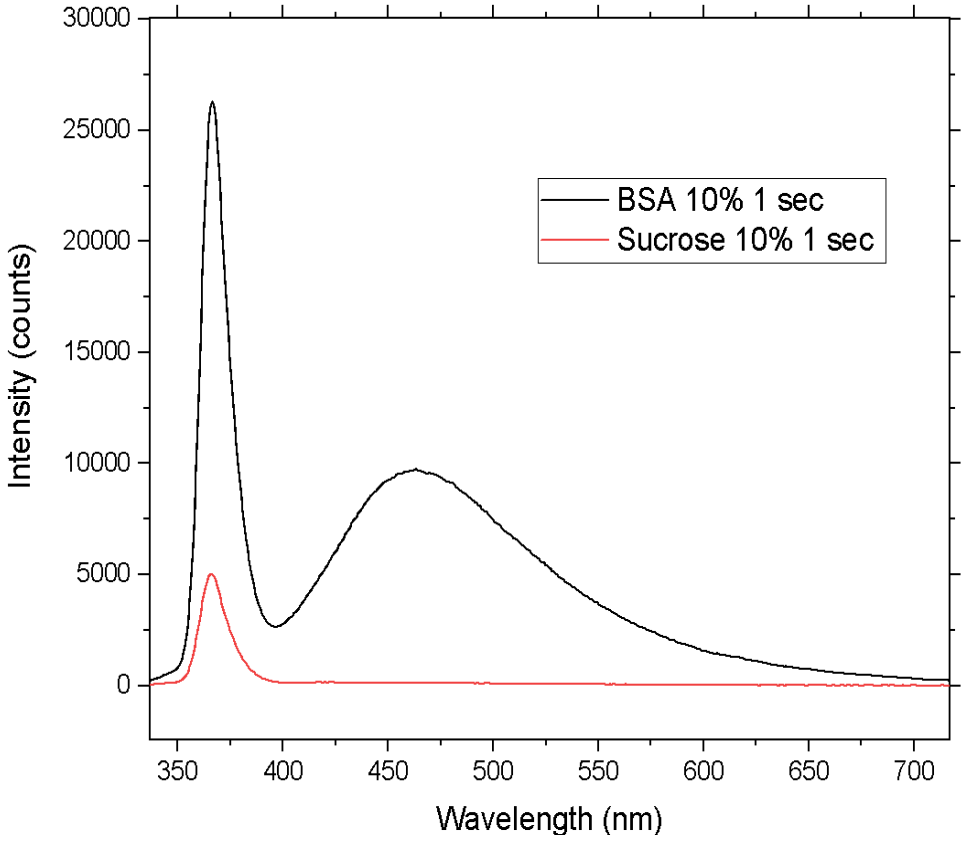
Representative volume of the vial

The beam is located 4 mm above the bottom of the vial. The beam radius was 1 mm at the beam waist, for the diameter of the vial 22 mm will give a representative volume of 12.6 mm^3. This is not a representation of the full volume in the vial and hence stirring would be a good option to increase the uniformity of the formulation during dissolution.

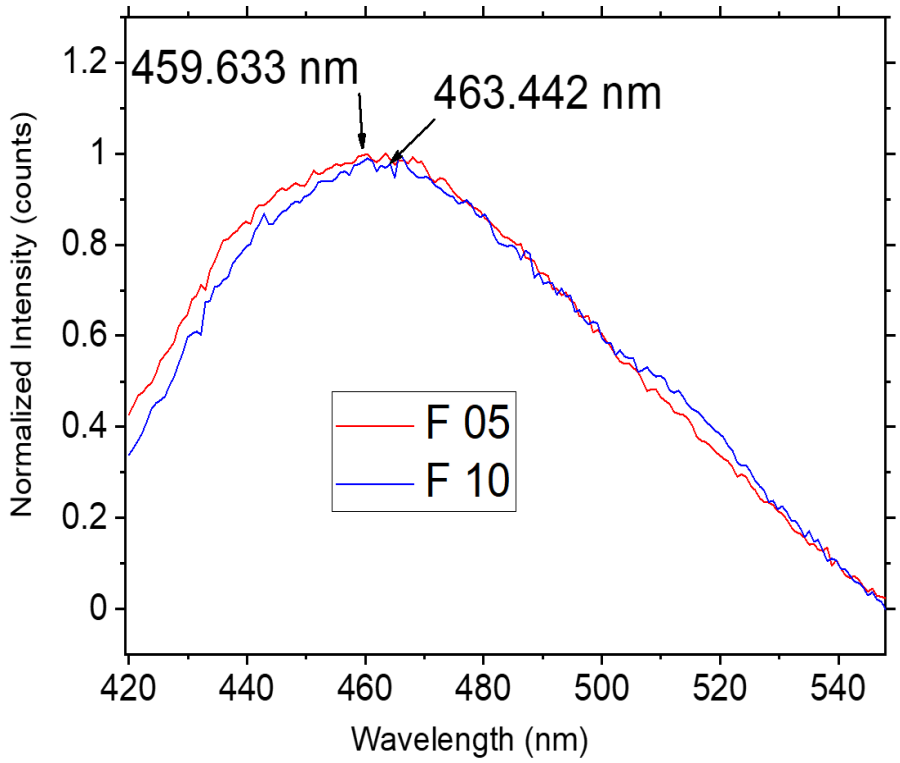
However, the inclusion of stirring needs to be also considered in relation to the specific application. The main objective of the work was to measure the reconstitution time of lyophilised protein formulations in-vial. In this study reconstitution without stirring was selected, which is a "worst-case scenario" and would result in the longest dissolution times. However, in benchmarking the instrumental set-up established against the manual visual

technique, variability between the reconstitution process between both measurements was minimised. The manual visual method, routinely performed for lyophilised protein formulations, involves either gentle agitation/swirling, or leaving the vial on a steady surface for a certain period of time, and therefore the less variable non-agitated approach was taken, and the potentially variable manual agitation or swirling was avoided.

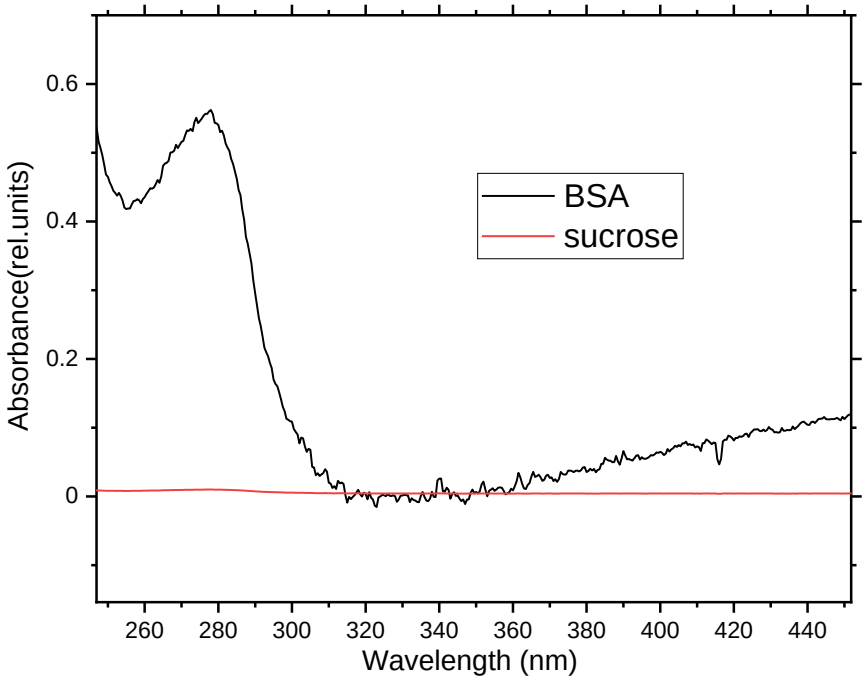
S2.2. Additional fluorescence and absorbance plots



Supplementary Figure 2.2 BSA and sucrose fluorescence response. Both are measured as solutions at 10% concentration. As expected, sucrose doesn't have any emission in the same region as BSA. Also, the difference in light intensity detected from the source indicates less elastic scattering.



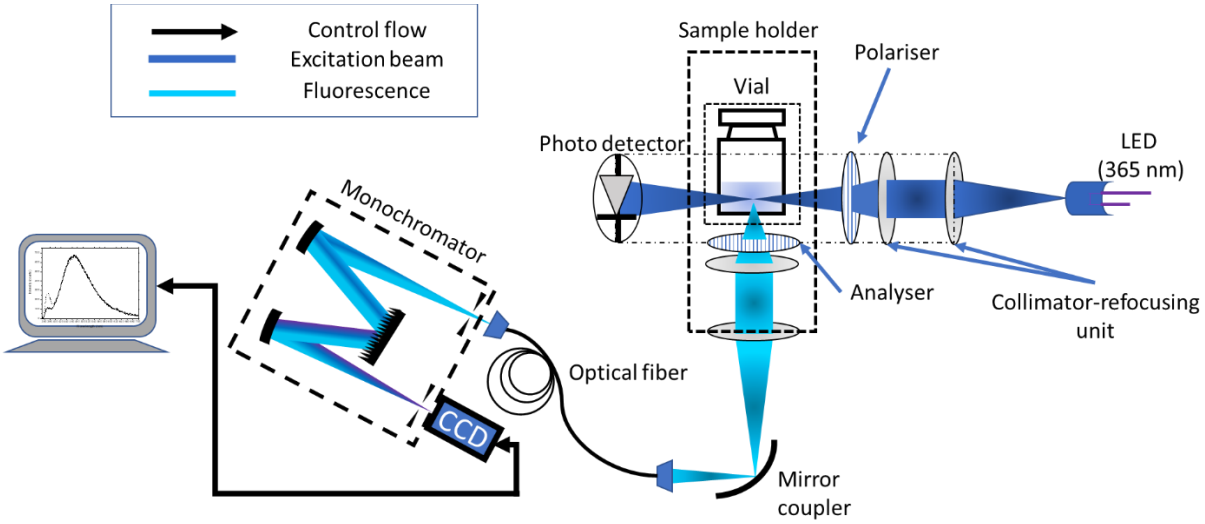
Supplementary Figure 2.3 Normalised Fluorescence spectra of F10 and F5 provided for comparison.



Supplementary Figure 2.4 Baseline corrected Absorption spectra of BSA and sucrose at the same concentration (0.1 ppm solutions)

S3. Supplementary information Chapter 3

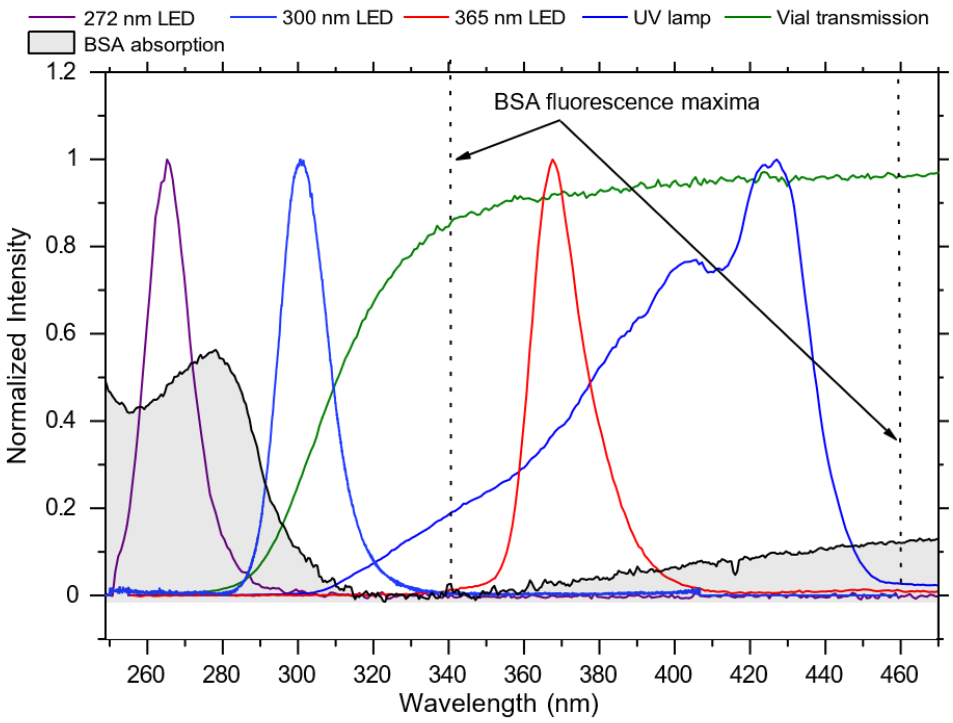
S3.1. Instrumentation



Supplementary Figure 3.1 Another schematic representation of the setup used to measure the UV fluorescence emission spectra of the stability of a protein formulation in-vial.

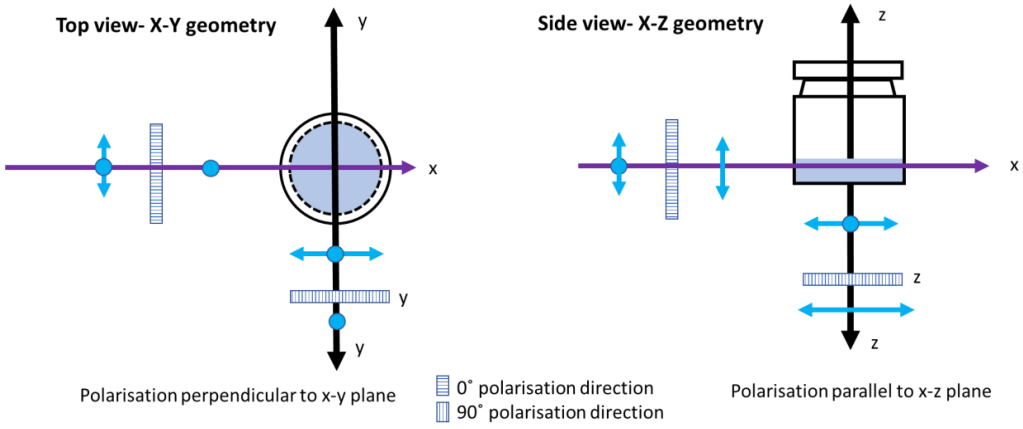
Vial transmission characterisation

For in-vial measurements, the broadband absorption spectrum (250 nm – 700 nm) of the lyophilization vial had less than 60% transmission for wavelengths shorter than 320 nm (Supplementary Figure 3.1). Hence, excitation wavelengths shorter than 320 nm could not produce sufficient intrinsic fluorescence to detect BSA in-vial at lower concentrations. A broadband source would not have enough intensity to generate fluorescence and also interferes with the emission at 340 nm and 460 nm, as shown in Supplementary Figure 3.1. Therefore, a 365 nm LED was used for the in-vial measurement set up, where the emission at 460 nm had maximum efficiency. The fluorescence collection was made from the bottom of the vial to avoid any loss of light due to the curvature of the vial. The excited region using 365 nm is a weak absorption region of BSA (the shaded region in Supplementary Figure 3.2 is the absorption spectrum of liquid BSA solution) and is enough to generate the required fluorescence.



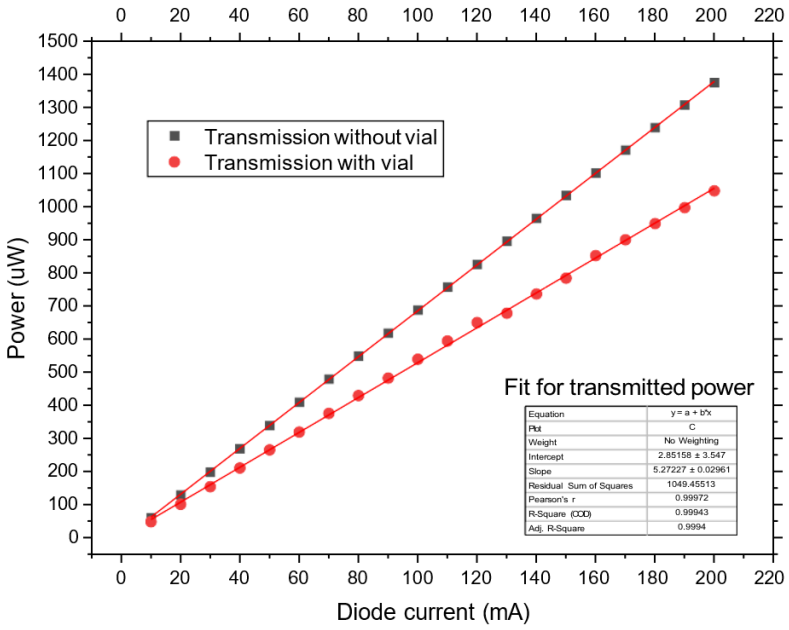
Supplementary Figure 3.2 Normalised transmission spectra of an empty lyophilization glass vial (in green) and emission spectra of different light sources in comparison. Spectra of LEDs at 272 nm (violet), 300 nm(blue), 365 nm (red) and a standard broadband UV lamp (blue) are shown. The re-scaled absorption spectrum of BSA is also included (shaded region) to show the influence of choosing the right excitation wavelength. Wavelengths below 300 nm ha ve less than 50% transmission from the vial. Dotted lines also show the position of two fluorescence maxima of BSA.

S3.2. Optical geometry of the set-up



Supplementary Figure 3.3 Polarisation directions on x-y and x-z directions of the vial. Z- direction is considered to be the length of the vial and x-y is the base plane. In both cases, the polarization direction is along the Z axis. The fluorescence will have to pass through the curvature of the vial when the excitation-emission direction is in the X-Y geometry, while it has a rather plane surface to pass through if the geometry is set up in the X-Z direction. In both cases, the polarisation depends on the fluorophore's rotation, but the output will be higher for heavier molecules in the emission polarisation direction.

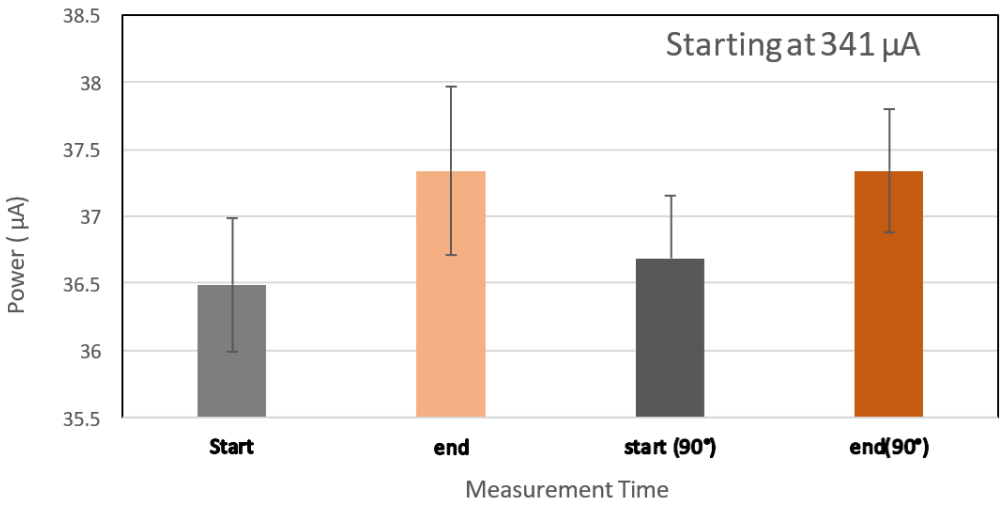
S3.3. UV denaturation characterization



Supplementary Figure 3.4 Characterization of transmitted power through the vial at 365 nm.

Supplementary Figure 3.4 shows that the used diode currents doesn't induce any optical non-linear absorption effects from the vial. Ultra-violet radiation is a well-known denaturant of proteins. The denaturant effect of the UV- LED excitation source used in the instrument was probed to ensure the validity of the results, the results are shown in Supplementary Figure 3.5.

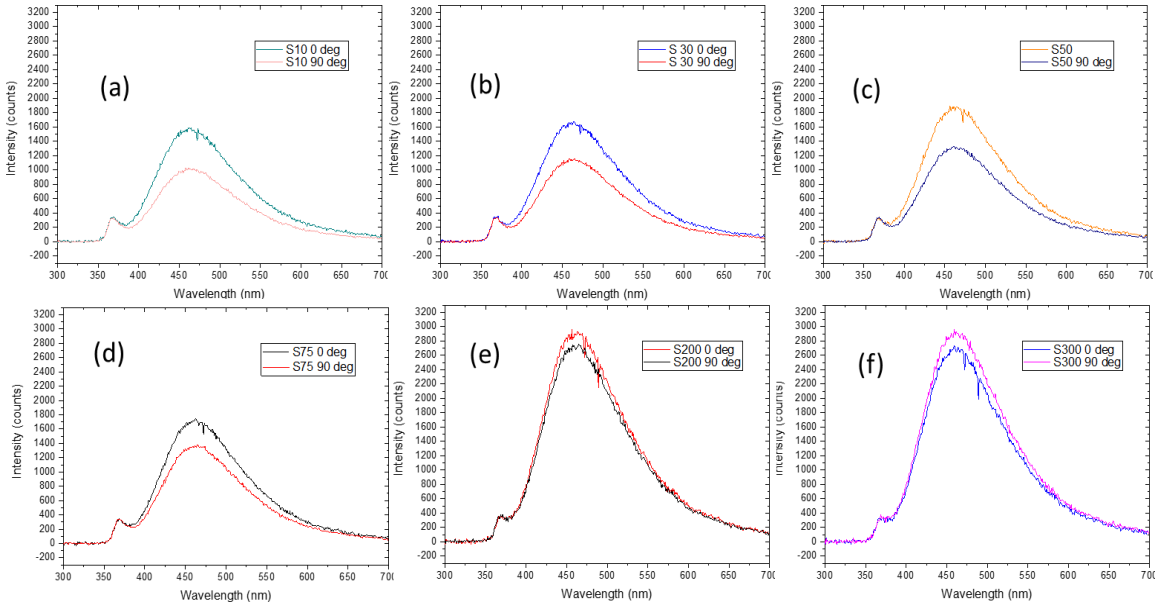
A photodetector was placed after the sample vial to detect the transmitted intensity. As the sample is exposed to the UV excitation source, increasing denaturation upon exposure can be identified by analysing the transmittance through the sample. An increase in transmittance (decrease in absorption) will indicate a higher denaturation level and vice versa. The figure below shows a minimal increase in power for transmitted light. Furthermore, upon a short rest and remeasurement, the samples returned to their original values, indicating the Brownian motion, and the volume of sample denatured is minimal.



Supplementary Figure 3.5 LED power transmitted through sample detected using a photodetector. Starting power was always 341 microamperes, reduced to 279 microamperes by an empty vial and further reduced to the values presented in the figure by the sample. The sample is slightly denatured by LED throughout the analysis, causing an increase in transmittance (reduction in absorbance). Some Samples were re-measured after a short rest and were shown to have fully recovered (returned to original values or similar).

S3.4. Fluorescence measurements details

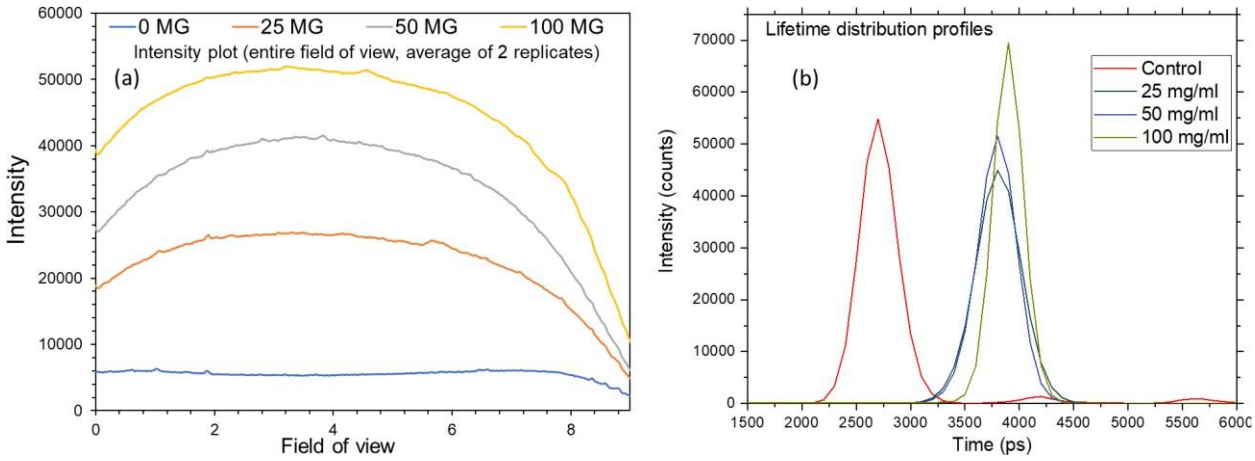
Fluorescence spectra at orthogonal polarisation for different SDS concentrations



Supplementary Figure 3.6 Fluorescence spectra in orthogonal polarisation directions obtained for different concentrations of SDS in BSA at 100 mg/ml. S 10 means 10 mM of SDS and so on. The spectra are normalized to the scattering source peak at 365 nm to demonstrate the difference in total intensities.

Fluorescence life time measurements

The fluorescence signal intensity (Supplementary Figure 3.7) increases with an increase in protein concentration; lifetime distribution profiles suggest a minor increase in lifetime with an increase in the concentration of BSA. A lifetime of ~3.8 ns was found for BSA samples, while control wells produced some lifetime signal at ~2.7 ns (Supplementary Figure 3.7(b)). No phosphorescence was observed for the same emission.



Supplementary Figure 3.7 Lifetime distribution profiles for BSA at concentrations 25 mg/ml, 50 mg/ml and 100 mg/ml (a) Intensity across field of view (b) lifetime of BSA at different concentrations. A slight change in life time can be observed for 100 mg/ml, possibly due to concentration quenching.

S4. Supplementary information Chapter 4

S4.1. MCR-ALS Spectral Loadings

Correlation parameters for spectral loadings

Supplementary Table 4.1 BSA Correlation coefficients for the matched fluorophore spectra and respective MCR components, BSA 1, BSA 2 and BSA 3 related to MCR component 1, 2 and 3 resp.

		trp avg	tyr avg	phe avg	BSA 1	BSA 2	BSA 3
trp avg	Pearson Corr.	1	0.62162	0.89596	0.95914	-0.4892	0.27264
	p-value	--	1.49E-25	6.57E-81	8.37E-125	5.32E-15	3.25E-05
tyr avg	Pearson Corr.	0.62162	1	0.60925	0.6988	-0.58719	0.07971
	p-value	1.49E-25	--	2.36E-24	1.94E-34	2.45E-22	0.23265
phe avg	Pearson Corr.	0.89596	0.60925	1	0.91632	-0.20555	0.55012
	p-value	6.57E-81	2.36E-24	--	5.41E-91	0.00189	2.81E-19
BSA 1	Pearson Corr.	0.95914	0.6988	0.91632	1	-0.52283	0.24327
	p-value	8.37E-125	1.94E-34	5.41E-91	--	2.96E-17	2.22E-04
BSA 2	Pearson Corr.	-0.4892	-0.58719	-0.20555	-0.52283	1	0.45889
	p-value	5.32E-15	2.45E-22	0.00189	2.96E-17	--	3.60E-13
BSA 3	Pearson Corr.	0.27264	0.07971	0.55012	0.24327	0.45889	1
	p-value	3.25E-05	0.23265	2.81E-19	2.22E-04	3.60E-13	--

Supplementary Table 4.2 Pepsin Correlation coefficients for the matched fluorophore spectra and respective MCR components. Pepsin 1, Pepsin 2 and Pepsin 3 related to MCR component 1, 2, and 3 respectively.

		Pepsin 1	Pepsin 2	Pepsin 3	trp avg	tyr avg	phe avg
Pepsin 1	Pearson Corr.	1	-0.7506	0.48999	0.8942	0.6441	0.98723
	p-value	--	3.46E-42	4.74E-15	3.89E-80	7.04E-28	1.11E-180
Pepsin 2	Pearson Corr.	-0.7506	1	-0.08258	-0.82188	-0.79141	-0.70682
	p-value	3.46E-42	--	2.16E-01	1.17E-56	8.83E-50	1.58E-35
Pepsin 3	Pearson Corr.	0.48999	-0.08258	1	0.18183	0.0306	0.496
	p-value	4.74E-15	2.16E-01	--	6.12E-03	0.64727	1.95E-15
trp avg	Pearson Corr.	0.8942	-0.82188	0.18183	1	0.62162	0.89596
	p-value	3.89E-80	1.17E-56	6.12E-03	--	1.49E-25	6.57E-81
tyr avg	Pearson Corr.	0.6441	-0.79141	0.0306	0.62162	1	0.60925
	p-value	7.04E-28	8.83E-50	0.64727	1.49E-25	--	2.36E-24
phe avg	Pearson Corr.	0.98723	-0.70682	0.496	0.89596	0.60925	1
	p-value	1.11E-180	1.58E-35	1.95E-15	6.57E-81	2.36E-24	--

Supplementary Table 4.3 Lysozyme Correlation coefficients for the matched fluorophore spectra and respective MCR components. Lysz 1, Lysz 2 and Lysz 3 related to MCR component 1, 2 and 3 respectively.

		trp avg	tyr avg	phe avg	Lysz 1	Lysz 2	Lysz 3
trp avg	Pearson Corr.	1	0.62162	0.89596	0.70517	0.53307	0.60993
	p-value	--	1.49E-25	6.57E-81	2.67E-35	5.42E-18	2.04E-24
tyr avg	Pearson Corr.	0.62162	1	0.60925	0.51536	0.46011	0.1322
	p-value	1.49E-25	--	2.36E-24	9.86E-17	3.06E-13	0.04714
phe avg	Pearson Corr.	0.89596	0.60925	1	0.65656	0.76889	0.53859
	p-value	6.57E-81	2.36E-24	--	2.97E-29	2.14E-45	2.12E-18
Lysz 1	Pearson Corr.	0.70517	0.51536	0.65656	1	0.32983	0.06284
	p-value	2.67E-35	9.86E-17	2.97E-29	--	3.90E-07	3.47E-01
Lysz 2	Pearson Corr.	0.53307	0.46011	0.76889	0.32983	1	0.19077
	p-value	5.42E-18	3.06E-13	2.14E-45	3.90E-07	--	4.00E-03
Lysz 3	Pearson Corr.	0.60993	0.1322	0.53859	0.06284	0.19077	1
	p-value	2.04E-24	0.04714	2.12E-18	3.47E-01	4.00E-03	--

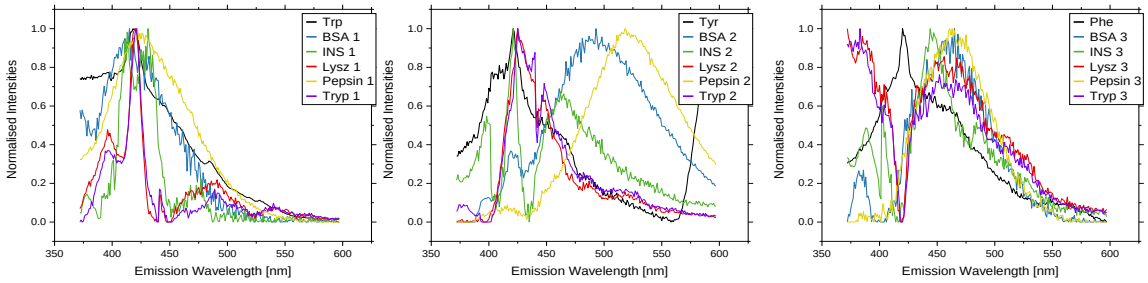
Supplementary Table 4.4 Trypsin Correlation coefficients for the fluorophore spectra and respective MCR components. Trypsin 1, Trypsin 2 and Trypsin 3 related to MCR component 1, 2 and 3 respectively.

		trp avg	tyr avg	phe avg	Trypsin 1	Trypsin 2	Trypsin 3
trp avg	Pearson Corr.	1	0.62162	0.89596	0.6692	0.51215	0.61947
	p-value	--	1.49E-25	6.57E-81	1.03E-30	1.64E-16	2.42E-25
tyr avg	Pearson Corr.	0.62162	1	0.60925	0.55107	0.40627	0.12344
	p-value	1.49E-25	--	2.36E-24	2.38E-19	2.16E-10	0.06394
phe avg	Pearson Corr.	0.89596	0.60925	1	0.6626	0.75044	0.50653
	p-value	6.57E-81	2.36E-24	--	6.07E-30	3.68E-42	3.94E-16
Trypsin 1	Pearson Corr.	0.6692	0.55107	0.6626	1	0.35587	-0.06994
	p-value	1.03E-30	2.38E-19	6.07E-30	--	3.77E-08	2.95E-01
Trypsin 2	Pearson Corr.	0.51215	0.40627	0.75044	0.35587	1	0.2237
	p-value	1.64E-16	2.16E-10	3.68E-42	3.77E-08	--	7.06E-04
Trypsin 3	Pearson Corr.	0.61947	0.12344	0.50653	-0.06994	0.2237	1
	p-value	2.42E-25	0.06394	3.94E-16	2.95E-01	7.06E-04	--

Supplementary Table 4.5 Insulin Correlation coefficients for the matched fluorophore spectra and respective MCR components. INS 2 and INS 3 related to MCR components 2 and 3 respectively. The first component for insulin while correlated to the Trp fluorescence, has no physical meaning since insulin does not contain Trp residues. Please see supplementary spectral loadings for further discussion on insulin MCR calculations.

		INS 1	INS 2	INS 3	trp avg	tyr avg	phe avg
INS 1	Pearson Corr.	1	0.15072	0.14235	0.71054	0.62794	0.7856
	p-value	--	2.34E-02	3.24E-02	4.80E-36	3.45E-26	1.34E-48
INS 2	Pearson Corr.	0.15072	1	0.35426	0.44944	0.20202	0.59649
	p-value	2.34E-02	--	4.38E-08	1.23E-12	2.28E-03	3.61E-23
INS 3	Pearson Corr.	0.14235	0.35426	1	0.4303	0.14008	0.5649
	p-value	3.24E-02	4.38E-08	--	1.33E-11	0.03533	1.88E-20
trp avg	Pearson Corr.	0.71054	0.44944	0.4303	1	0.62162	0.89596
	p-value	4.80E-36	1.23E-12	1.33E-11	--	1.49E-25	6.57E-81
tyr avg	Pearson Corr.	0.62794	0.20202	0.14008	0.62162	1	0.60925
	p-value	3.45E-26	2.28E-03	0.03533	1.49E-25	--	2.36E-24
phe avg	Pearson Corr.	0.7856	0.59649	0.5649	0.89596	0.60925	1
	p-value	1.34E-48	3.61E-23	1.88E-20	6.57E-81	2.36E-24	--

Insulin Spectral loadings



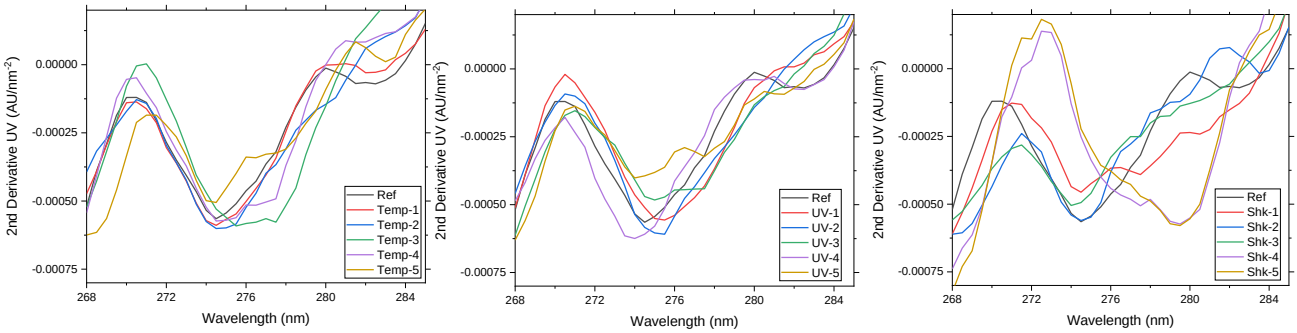
Supplementary Figure 4.1 Normalised MCR components for BSA, Insulin, Lysozyme, Pepsin, and Trypsin, as compared to their matched reference spectrum. The black line represents reference peaks for pure aqueous solutions of the aromatic amino acids a) Tryptophan b) Tyrosine and c) Phenylalanine at 1mg/ml.

Insulin MCR analysis shows a third (or in this case 1st matched component) representing tryptophan, despite the physical lack of tryptophan in the structure when analysed at three MCR components. It was decided that for MCR results to be comparable across the studied proteins a two-component model for insulin alone would be insufficient. As seen in the figure above, the three-component model could not be optimised for insulin leading to a noisy 1st component

S5. Supplementary information Chapter 5:

S5.1. UV derivatives

2nd derivative UV-analysis of the samples at 276 nm was used to confirm the unfolding of the protein (ref). Fresh samples as well as days 1 and 2 of the stability study peaks were within 0.5 nm of each other. Day 7 again showed the most significant change with a red-shift of 1.8 nm, while days 14 and 21 were less red-shifted (compared to a fresh sample) showing a peak within 1 nm of the reference.



Supplementary Figure 5.1 UV 2nd derivative spectra of 25 mg/ml BSA samples incubated at 40°c for (1,2,7, 14, and 21 days). No major changes are seen on days 1 and 2 (< 0.5 nm redshift), the greatest shift is seen on day 7(green) with a 2 nm difference. A re-shifting to the opposite direction (blue shift) is seen on days 14, and 21.

S5.2. Extra modelling for fold under temp incubation

An extra model was constructed using reference and 40°C incubated samples only. As these samples show the least variability in unfolding the models are shown to be more robust with higher R^2, lower RMPRESS values, and higher cumulative variance explained per model.

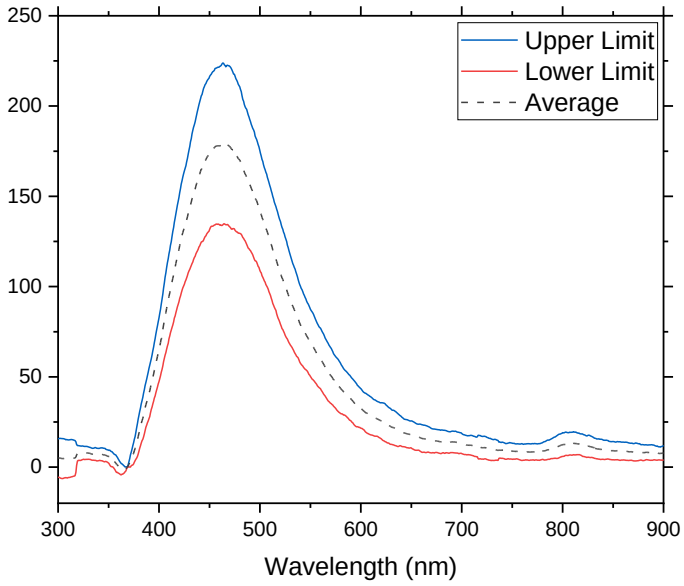
Supplementary Table 5.1 Summary of thermally degraded protein PLS Models using fluorescence anisotropy as predictors for the variables listed below.

	TEMPERATURE-DENATURED SAMPLES ONLY MODEL					
	R^2	RMSE	Root mean Press	Latent variables	Variance explained	Method of normalisation
HMW	**0.807**	1.622				
Monomer	0.747	1.847				
LMW	0.779	0.407	0.64	7	78.89%	
Concentration (UV)	0.681	0.094				To lambda max
Alpha Helix	0.830	0.028				
Beta Sheet	0.828	0.011	0.66	5	77.60%	
Random Coil	0.847	0.011				
	R^2	RMSE	Press	lv	Variance explained	Method of normalisation
HMW	0.970	0.626				
Monomer	0.976	0.626				
LMW	0.967	0.171	0.46	7	90.33%	
Concentration (UV)	0.701	0.054				SNV
Alpha Helix	0.923	0.019				
Beta Sheet	0.931	0.007	0.414	5	92.79%	
Random Coil	0.930	0.008				

S5.3. Alternative reference batch analysis methodology

While considering the challenges associated with implementing a PLS model, such as that developed in Chapter 3, an alternative screening methodology based on comparison with a reference batch was developed.

The samples subjected to light, temperature and shaking stress were tested against the generated “reference batch” after being processed similarly (normalisation & polarisation subtraction). Multiple samples can be analysed via a moving window of a set number of nanometres across the processed spectral x-axis. The window compares the values of the sample anisotropy against the reference batch limits; if the sample anisotropy is within the limits, the analysis moves on to the following window. Otherwise, the analysis is halted, a deemed to exceed the limits of acceptability set and the next sample is analysed. The standard deviation (and multiples of which from 1 to 50 standard deviations) was used to set the “limits of acceptability” for further stability testing. Supplementary Figure 5.2 below outlines a general example of this process.



Supplementary Figure 5.2 Fluorescence anisotropy for 9 reference samples as an average, upper and lower limit were set as a multiple of the standard deviation of the reference samples.

For this study, a simple analytical validation experiment was conducted to obtain the ideal value for each of the parameters. Analytical data of the protein solution was obtained using circular dichroism (secondary structure), UV absorbance (chemical stability), size-exclusion chromatography (aggregation), and dynamic light scattering (aggregation and hydrodynamic size).

This method can be adjusted by altering a number of controllable parameters associated with the approach; type of normalisation used in pre-processing, smoothing window and polynomial order, the acceptable range set in multiples of standard deviations of the reference batch, comparison start point, and comparison window (in nm). The choice of controllable parameters would be dependent on the type and granularity of the required analysis. The choice to test at multiple levels of acceptability was to demonstrate flexibility

in the choice of protein and the applicability of different regulatory standards (EMA, 2013; Step, 2017).

To verify the feasibility of the automated detection as well as the fluorescence as a method for capturing the variability in the protein state the PCA results were compared to the following result matrix generated by the "reference batch" method. Results of the automated random testing at different levels of acceptance limits can be seen in Supplementary Figure 5.2, and show similar results to the PCA clustering. Reference samples are within the acceptability limits, while denatured samples fail to pass.

Of the samples randomly selected, 3rd timepoint shaking samples (7 days) show the highest level of deviation from the mean reference spectrum which is corroborated by nanodrop UV absorbance seen in Figure 5.2. In addition, the CDNN results in Figure 5.4 show a deviation of ~5% beta content on average, and a significant increase in variability for this timepoint. The PCA results in Figure 5.6a and b show that the shaking timepoint 3 (7 days) sample is located furthest from the reference sample cluster of the samples included in the random test, while remaining inside the confidence interval.

Supplementary Table 5.2 Acceptance (within limits = Y, outside limits = N) for samples tested using the reference batch script. All reference samples are within acceptance criteria between 1.5~2 SD while all stressed samples are outside the limits for the same range.

Type	Sample Label/ timepoint	Time (days)	Standard deviations multiplier for acceptance limits								
			1	1.5	2	2.5	3	3.5	4	4.5	5
Reference	3	0	N	Y	Y	Y	Y	Y	Y	Y	Y
Reference	5	0	Y	Y	Y	Y	Y	Y	Y	Y	Y
Reference	6	0	Y	Y	Y	Y	Y	Y	Y	Y	Y
Reference	8	0	Y	Y	Y	Y	Y	Y	Y	Y	Y
Shake	3	3	N	N	N	N	N	N	N	Y	Y
UV	1	1	N	N	Y	Y	Y	Y	Y	Y	Y
UV	2	2	N	N	N	Y	Y	Y	Y	Y	Y
Temperature	2	2	N	N	N	N	Y	Y	Y	Y	Y
Temperature	4	7	N	N	N	N	Y	Y	Y	Y	Y

The analysis was run in triplicate for each of the samples, the accepted condition was only applied if all 3 replicates were considered a pass. All samples tested were removed from the training dataset before the analysis.

The first reference sample failed to pass within acceptance criteria for 1 standard deviation, most likely owing to its lying on the boundary between native and high-temperature incubated as can be seen from the clustering in Figure 5.6. The remaining reference samples tested passed all limit levels tested. The totality of the stability study samples was retested using this method and a table of the outcomes can be found in Supplementary Table 5.3 below.

All stability study results

Supplementary Table 5.3 the results of all automated AUC testing conducted on the stability study samples under conditions of Temperature denaturation (temp), shaking (Shk), and daylight exposure (UV). Timepoints included 1, 2, 7, 21, and 30 days and are labelled (Tp1, Tp2, Tp3, Tp4, Tp5) respectively. Row numbers indicate the number of standard deviations used to set the limits of acceptability. The main body of the table refers to passing samples (1 – green) and failing samples (0 – clear).

	Shk_Tp1	Shk_Tp2	Shk_Tp3	Shk_Tp4	Shk_Tp5	UV_Tp1	UV_Tp2	UV_Tp3	UV_Tp4	UV_Tp5	Tmp_Tp1	Tmp_Tp2	Tmp_Tp3	Tmp_Tp4	Tmp_Tp5
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2.5	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
3	0	0	0	0	0	1	1	0	0	0	1	1	1	1	0
3.5	0	0	0	0	0	1	1	0	0	0	1	1	1	1	1
4	0	0	0	0	0	1	1	0	0	0	1	1	1	1	1
4.5	0	0	1	0	0	1	1	0	0	0	1	1	1	1	1
5	0	0	1	0	0	1	1	0	0	0	1	1	1	1	1
5.5	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1
6	0	0	1	0	1	1	1	0	0	0	1	1	1	1	1
6.5	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1
7	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1
7.5	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1
8	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1
8.5	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1
9	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1
9.5	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1
10	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1
10.5	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1
11	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1
11.5	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
12	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
12.5	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
13	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
13.5	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
14	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
14.5	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
15	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1