National University of Ireland – Galway

College of Sciences, School of Chemistry

Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES) for the analysis of Immunoglobulin G (IgG) type proteins

Thesis presented for the degree of PhD

of the National University of Ireland − Galway

by **Marina Steiner-Browne**

Student ID – 13261116

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

− August 2020 −

Acknowledgements

I would like to thank my supervisor Alan Ryder for giving the opportunity to develop my research skills as an international visiting student back in 2014. What was supposed to be a simple summer internship was extended to an undergrad final project and to this PhD.

I also want to thank my colleagues from the Nanoscale Biophotonics Laboratory (NBL), it was a pleasure to work with such a great group. Special thanks to Yannick Casamayou-Boucau for helping me on this journey since 2014, Saioa Elcoroaristizabal for sharing her knowledge in chemometrics, and Ana Luiza de Faria e Silva for being in this crazy journey with me.

This work has emanated from research supported in part by a research grant from Science Foundation Ireland and is co-funded under the European Regional Development Fund under Grant number (14/IA/2282, Advanced Analytics for Biological Therapeutic Manufacture, to AGR).

Contents

Commonly used abbreviations

ARMES: Anisotropy resolved multidimensional emission spectroscopy

ASDR: Absent spectral data region

CONCORDIA: Core consistency diagnostic test

EEM: Excitation emission matrix

FRET: Förster resonance energy transfer

HPLC-SEC: High-performance liquid chromatography-size exclusion

HPW: High purity water

IFE: Inner filter effect

IgG: Immunoglobulin G

I_F: Fluorescence intensity

LFH: Laminar flow hood

LOR: Limit of reporting

MCR-ALS: Multivariate curve resolution alternating least square

MDF: multidimensional fluorescence

NS: Native state

PARAFAC: Parallel factor analysis

PBS: Phosphate saline buffer

pEEM, pMDF, pTSFS: Polarized EEM, MDF, TSFS

PFC: PARAFAC component

Phe: Phenylalanine

***r*:** Anisotropy

rIgG: Rabbit Immunoglobulin G

RSD: Relative standard deviation

SimI: Similarity index

SNR: Signal-to-noise ratio

t-EEM: TSFS spectra converted into an EEM like layout

Trp: Tryptophan

TSFS: Total synchronous fluorescence spectroscopy

Tyr: Tyrosine

UV-Vis: Ultraviolet-visible spectroscopy

Abstract

Immunoglobulin G (IgG) is the main antibody used in the biopharmaceutical industry for therapeutic purposes. Monoclonal antibodies (mAb) are growing in the marked due to its high specificity and safety. Physical and chemical stresses can lead to unfolding of the tertiary structure, which could then refold into a different structure, and/or aggregate. Changes in protein structure are dangerous and can cause adverse immunogenicity issues. One key factor to ensure efficacy and safety is to understand and measure the stability of mAbs in solution. A combination of different analytical techniques, some of them are expensive and time consuming, is necessary to monitor mAbs stability, and to assess protein aggregation. The aim of this project was to apply a newly developed fluorescence-based method, Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES), for the rapid characterisation of IgG type proteins in solution, which could be used to monitor IgG structure. This method combined polarized multidimensional fluorescence spectroscopy (pMDF), anisotropy, and chemometric analysis, such as Parallel factor analysis (PARAFAC), to try to resolve emission from different fluorophores, which were used to follow protein structure.

ARMES measurements of IgG solutions in its native state were carried out with polarized Excitation Emission Matrix (pEEM) and polarized Total Synchronous Fluorescence Spectroscopy (pTSFS) measurements. It was the first time that PARAFAC analysis was attempted on IgG, as there was no consensus about the optimal pre-processing method for this type of data, we evaluated several methods. Although there was insufficient fluorescence fluctuation in the native state for PARAFAC analysis to resolve different fluorophore populations, we were able to photophysically characterise the IgG, which served as a baseline for monitoring protein stability using ARMES. Thus, for the IgG protein, even if ARMES could not recover individual fluorophore emission, it was showed to be a suitable technique for monitoring protein stability.

1. Introduction

The main objective of this thesis is to develop a fluorescence-based method for the rapid qualitative characterization of Immunoglobulin G (IgG) type proteins in the solution phase. This is to be accomplished using the newly developed Anisotropy resolved multidimensional emission spectroscopy (ARMES),[1-3] and careful experimental design to acquire the spectral data, which is then analysed with multivariate (chemometrics) methods. First, we must explain the reason behind our interest in IgG. The second part of this chapter introduces general fluorescence concepts, which help to understand the photophysical process that affect ARMES data from protein. The last part shows how ARMES can be used to develop an analytical method for characterizing and monitoring changes in IgG structure.

1.1 Immunoglobulin G (IgG)

IgG is a common class of antibodies (protein) used for therapeutic purposes.[4, 5] Monoclonal antibodies (mAb) are in the spotlight of the pharmaceutical industry, with IgGs accounting for ~99% of the mAbs marketed.[6] These therapeutic proteins are highly specific, with lower safety risks when compared with other therapeutic classes. The quick advance in knowledge of the diseases at a molecular level, combined with reduced risks of adverse effects, make the mAbs the first products candidates to advance to human clinical trials.

IgG has a flexible 'Y'-shaped structure (Figure 1.1), composed by a constant (Fc) and a variable region (Fab), connected by di-sulphide bonds, and a hinge region, with a molecular weight of ~150 kDa. Changes in the amino acids (AAs) sequence of the Fab portion (hypervariable region) are responsible for the IgG diversity. Variations in the connection between Fab and Fc portion, generates different IgG isotypes, which are species dependent: humans have four isotypes,[7, 8] while rabbits have only one.[9-11] These differences in the Fab portion between the IgG isotypes, as well as changes in AAs sequences, and protein structure determine protein stability.[12, 13] Protein structure can change upon physical and chemical stresses. The most serious changes involve unfolding of the tertiary structure, and this can sometimes be followed by refolding and/or aggregation. Improper manipulation during manufacturing, storage, and use

can lead to protein unfolding, and aggregation, which can cause a loss in function, and adverse immunogenicity issues.[14] The stability of proteins in solution is, therefore, a key factor for their successful therapeutic use.

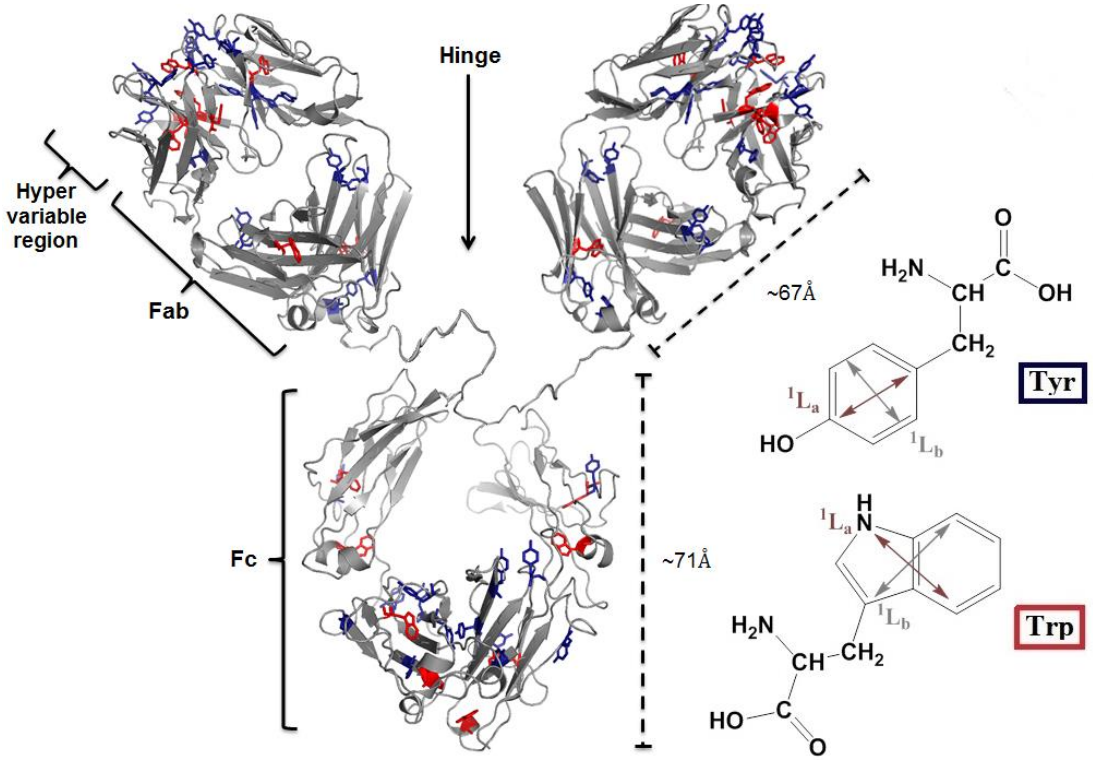

Figure 1.1: Structure of rabbit IgG (rIgG) showing the major structural elements and locations of the Trp (red) and Tyr (blue) fluorophores. Trp and Tyr structures show the different transition dipoles. Protein structure was created by the combinations of Fc and Fab images from the RCSB Protein data bank (**www.rcsb.org**) of PBD ID 2VUO,[15] and 4HBC,[16] respectively.

The various quality attributes of therapeutic proteins can be measured and monitored using a variety of techniques. For example, high-performance liquid chromatography-size exclusion (HPLC-SEC), and dynamic light scattering (DLS) are most commonly used to assess protein aggregation.[17] Spectroscopic methods, such as circular dichroism (CD), and Fourier transform infrared (FTIR) can also be used to characterize the native structure, and monitor changes in protein structure.[18, 19] While far-UV CD is sensitive to changes in secondary structure, near-UV CD detects changes in tertiary structure, however, protein concentration has to be carefully selected, as sedimentation, and light scattering can negatively impact CD. FTIR can be used with liquid and solid samples in high concentrations, and with large aggregates, but proteins can bind/adhere to the crystal, and the water signal has to be carefully removed from the spectra. Therefore, a combination of different analytical tools, some of which are

expensive and time consuming, are required to characterize protein structure, stability, and aggregation in solution.[20]

1.2 Intrinsic fluorescence spectroscopy (IFS)

The biopharmaceutical industry spends billions trying to develop new drugs, which includes the identification, and development of quality control tools. The development of a single analytical technique that could monitor protein stability, changes in structure, particle size, and aggregation is of paramount importance. This new analytical method should be robust, less complex, and more economic,[21] which could be achieved using intrinsic fluorescence spectroscopy (IFS). IFS can be used to study protein structure in solution, due to the presence of phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) fluorophores in proteins.[22-24] Some advantages of the use of fluorescence are that this method is non-destructive, non-invasive, and highly sensitive.

1.2.1 Basic principles of fluorescence

The emission of light from electronically excited states is known as luminescence. Depending on the nature of the excited state, luminescence is divided into two main categories: fluorescence and phosphorescence.[22] These are generally better illustrated by a Jablonski diagram, as shown in Figure 1.2.

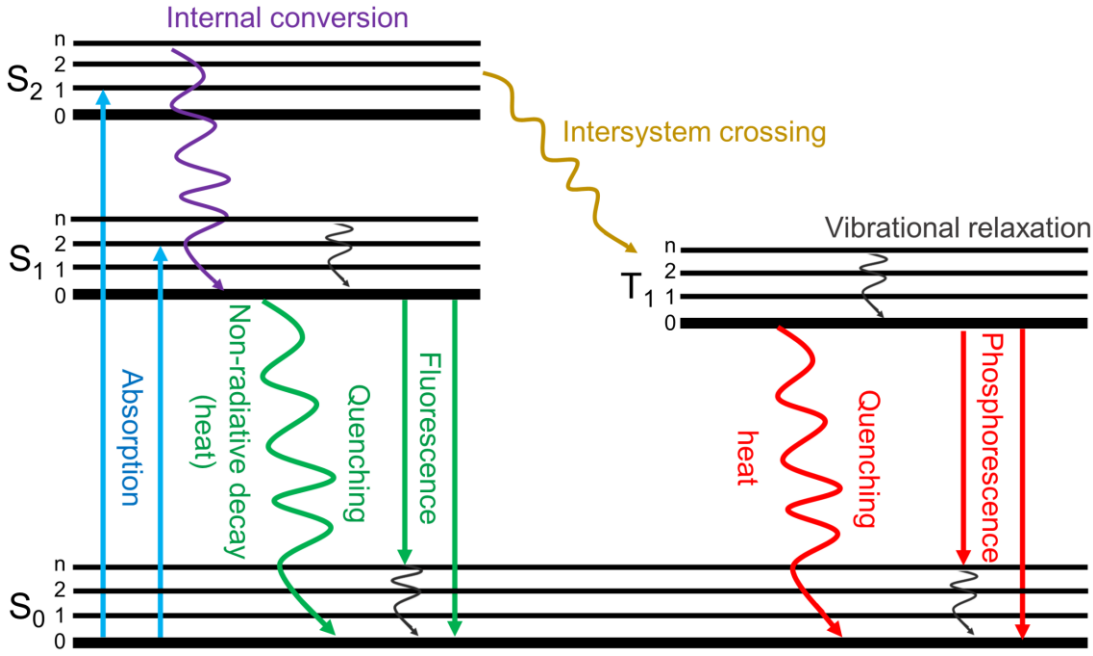

Figure 1.2: Illustration of one form of the Jablonski diagram.[22]

The phenomena of fluorescence typically occurs when an electron in a molecule is photo-excited from the ground state (S_0) to an excited singlet state (S_1, S_2, or S_n), and then returns to S_0, rapidly emitting a photon (emission of light).[22] This phenomenon is fast, with a fluorescence lifetime (τ) of 0.1−10 ns, which describes the average time that an electron spent in the excited state. Phosphorescence emission occurs from triplet excited state (T_1), in which the lowest vibrational level of S_1, and the upper vibrational level of T_1 have the same vibrational level. While in the singlet excited state the electron spin has an opposite direction from that in S_0, which allows the transition from S_1 to S_0, the electron spin present in T_1 and S_0 have the same spin orientation. Thus, transitions from T_1 to S_0 are forbidden in quantum theory, increasing the phosphorescence lifetime (milliseconds to seconds). As this process is very slow, intersystem crossing, and vibrational relaxation are favored over the phosphorescence phenomena, which can be observed at low temperature, or a rigid medium at room temperature.[24]

Fluorescence and phosphorescence are the phenomena in which the excited molecules return to S_0 emitting a photon to release energy. However, the emission of a photon is not the only way to release energy from the excited state, and this energy can also be dissipated as heat, or transferred to other molecules via collisions or dipole-dipole coupling. These types of non-radiative decays are known as quenching effects and are competing with the fluorescence, and phosphorescence.

1.2.2 Fluorophores and their characteristics

Fluorophores are molecules that can absorb and re-emit light. These are generally aromatic molecules and are sometimes, in the context of protein fluorescence, split into two groups: intrinsic and extrinsic.[22-24] Extrinsic fluorophores (*e.g.* 1-8,anilinonaphthalene sulphonate and rhodamine B)[25, 26] can be added to a sample, enabling or improving the use of fluorescence techniques to analyse the sample. However, the use of an extrinsic fluorophore can induce local perturbations in the sample, and this must be taken in consideration when interpreting the data.[24] Intrinsic fluorophores are those molecules that are naturally fluorescent, such as the aromatic amino acids (AAs) Phe, Tyr, and Trp. Both intrinsic and extrinsic fluorophores are characterized by their excitation/emission spectra (and Stokes shift), quantum yield, lifetime, and anisotropy (Section 1.4.1).

The excitation spectrum is measured by fixing an emission wavelength (λ_{em}) and scanning the fluorescence intensity over a range of excitation wavelengths (λ_{ex}). The emission spectrum is measured at a fixed λ_{ex} and measuring fluorescence intensity over a range of λ_{em} (Figure 1.3). The fluorescence emission spectrum is highly sensitive to chemical structure of the fluorophores and the surrounding environment. As observed in the Jablonski diagram (Figure 1.2), the energy of the emission is less than that of the absorption, which is known as Stokes shift ($\Delta\lambda$).[22] The Stokes shift can be influenced by solvent effects (change in fluorophore environment), and/or energy transfer, which can be very useful in the study of protein structure via IFS.[23, 24]

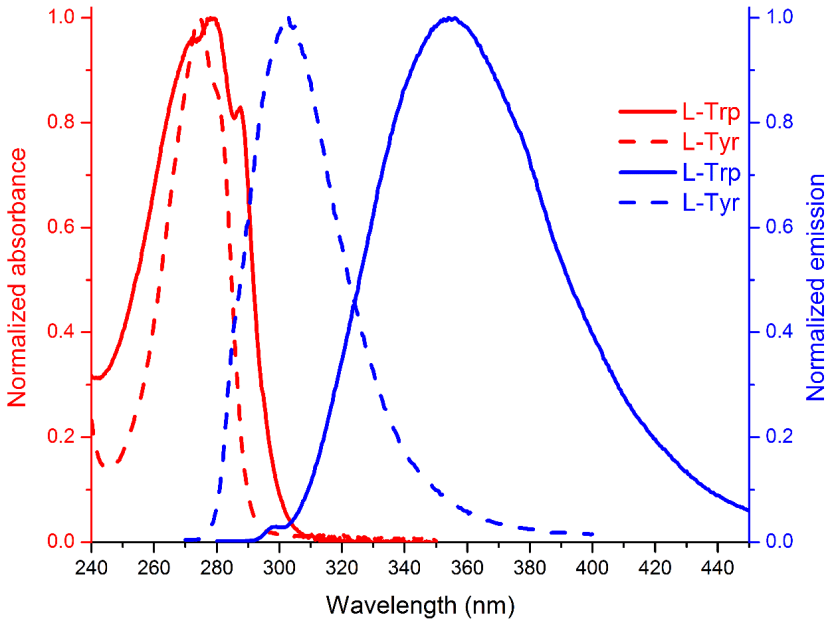

Figure 1.3: Absorption and emission spectra of L-Tyr, and L-Trp in PBS solution (0.1 M phosphate buffer solution, pH 7). Data from PhotoChemCad 2.1 software.[27-29]

The fluorescence quantum yield (Φ_F) is the ratio between the number of emitted photons and the number of absorbed photons. The closer the Φ_F is to one, the brighter, and more efficient is the fluorophore.[24] Phe has a small quantum yield and transmit its energy to Tyr (Phe-to-Tyr energy transfer). Thus, the Phe fluorescence cannot usually be discriminated in the presence of much stronger Tyr, and/or Trp emission (Table 1.1).[23] As fluorophore emission can be influenced by changes in its environment, the amount of time that a fluorophore spends in the excited state (τ) can influence the shape of the emission spectra.

Table 1.1: Fluorescence characteristics of Trp, Tyr, and Phe AAs (0.1 M phosphate buffer, pH 7, measured at room temperature).[22, 23]

Amino acids	λ_{ex} (nm)	λ_{em} (nm)	Quantum yield	Lifetime (ns)
Trp	280	350	0.13	2.8
Tyr	275	303	0.14	3.3
Phe	258	282	0.03	6.8

Protein IFS generally uses an excitation ≥280 nm, meaning that Phe is not excited (Table 1.1). Tyr and Trp can be excited at λ_{ex}=276 nm, and when the λ_{ex}≥295 nm, mainly the Trp is excited. Nevertheless, emission from Phe would not be observed unless the protein lacks Tyr and Trp. Tyr fluorophores are generally present in a larger quantity than the Trp fluorophores.[23] Thus, even though the Trp is a stronger emitter, Tyr emission can still be observed in protein fluorescence when very few Trp are present.[30] Although it is possible to detect Tyr emission in multifluorophore proteins, its utility is restricted as its emission is very insensitive to changes in solvent polarity. On the other hand, Trp emission is highly sensitive to changes in polarity and/or local environment, which can be very useful on the study of protein structural changes.[22, 23, 31-33]

As Trp emission is highly dependent on its environment (*e.g.* solvent polarity, temperature, pH), it is possible to classify different Trp classes. Burstein proposed that there are five main classes of Trp present in proteins,[34] three of them are used to follow changes in the Trp environment:[35] class I (λ_{em}=330−332 nm, $\Delta\lambda$=48−50 nm), with the Trp buried in a non-polar region of the protein; class II (λ_{em}=340−342 nm, $\Delta\lambda$=53−55 nm), when the Trp is limited exposed to the solvent; and class III (λ_{em}=350−353 nm, $\Delta\lambda$=59−61 nm), with the Trp completely exposed to the solvent. The other two classes are observed as structured spectra at λ_{em}=308 and 316 nm, known as class A and class S, respectively, and do not concern us here. Although these different classes of Trp can be easily identified for single-Trp proteins, the presence of multiple Trp (and Tyr) fluorophores make it harder to resolve individual classes of Trp. For multi-Trp proteins, the Trp emission spectrum represents an average of all Trp within its structure.[22] Thus, a lack of change in the average emission does not necessarily means that protein structure is unchanged.[a]

[a] The Trp spectrum can remain unchanged if the structural change is a result of some Trp being blue shifted (to shorter wavelengths), and some red shifted (to longer wavelengths), resulting in the same average class of Trp as observed before any changes in protein structure.

In addition, the presence of two Trp excited states, ^1L_a and ^1L_b, which are perpendicular to each other (Figure 1.1), increases the Trp spectral complexity (Figures 1.1/1.4).[22, 36-41] Trp emits from its ^1L_a excited state, unless its local environment is completely nonpolar. The ^1L_b state is insensitive to solvent environment and emits in nonpolar solvents. Valeur and Weber[36] used the excitation anisotropy spectra to resolve the excitation spectra of the ^1L_a and ^1L_b excited states of Trp (Figure 1.4). The ^1L_a transition state shows unstructured emission spectra, while the ^1L_b transition state has structured emission spectra (Figure 1.4). The maximum λ_{ex} for the ^1L_a and ^1L_b excited states of Trp appears at ~280, and ~290 nm, respectively, and at longer wavelengths (λ_{ex}=295−300 nm), only the ^1L_a is excited. Because only the ^1L_a Trp is excited at $\lambda_{ex}\geq$295 nm, several studies use λ_{ex}=295 nm to monitor tertiary structural changes in IgG during physical/chemical stress.[17, 42, 43]

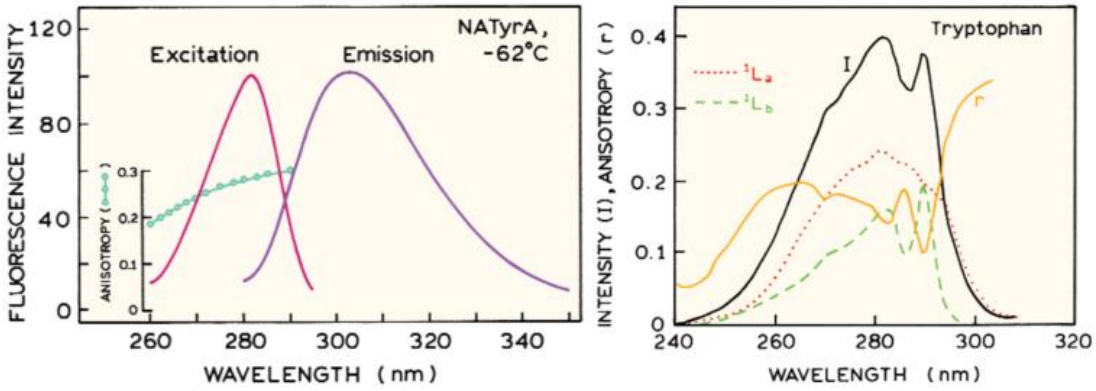

Figure 1.4: (Left) Emission spectrum and excitation anisotropy spectra of N-acetyl-L-tyrosinamide (NATyrA). Fluorescence anisotropies were measured in a mixture of 70% propylene glycol with 30% buffer at −62°C.[44, 45] (Right) Excitation anisotropy spectra of Trp in propylene glycol at −50°C. Also shown the anisotropy spectra (yellow line), and the anisotropy-resolved spectra of the ^1L_a (red dotted), and ^1L_b (green dashed) transitions.[38] Reproduced from Lakowicz,[22] with permission of Springer Nature.

1.3 Photophysical processes that affect the spectral properties of fluorophores

The fluorophores within a protein are sensitive to its environment, meaning that the emission spectrum of proteins is sensitive to its structure. Even though Tyr emission is very insensitive to environmental change, it is highly influenced by intermolecular, and intramolecular interactions. Any changes in the protein environment affect how the fluorophores interact with each other and with their environment.[22-24] Here, we describe the main photophysical processes that can affect the spectral properties of Tyr, and Trp fluorophores.

1.3.1 Quenching

Fluorescence quenching occurs when a fluorophore interacts with a quencher molecule, resulting in a decrease in fluorescence intensity. The quenching process is a result from various molecular interactions, which requires the fluorophore and the quencher to be in contact and is divided into two main groups: static and dynamic quenching. Static quenching occurs when a quencher associates to the fluorophore in the ground state, forming a non-fluorescent complex. Dynamic quenching is a result of non-radiative relaxation by internal conversion, and intersystem crossing (Figure 1.2), which occurs when a quencher collides with the fluorophore in the excited state, resulting in the return of the fluorophore to the ground state, without emitting a photon.[22, 24] In addition, the decrease in fluorescence intensity can be caused by an inner filter effect (IFE) phenomenon as a result of high optical density; or as a result of high turbidity, causing light loss due to scatter.

1.3.2 Förster resonance energy transfer (FRET)

Förster resonance energy transfer (FRET) is a non-radiative transfer of energy, which occurs when a donor (D) molecule transfers its energy to an acceptor (A) molecule, decreasing the fluorescence intensity of the donor, and increasing that of the acceptor. Donor and acceptor can be identical or chemically distinct, which are known as homo-, and hetero-transfer, respectively. This energy transfer is generated by a long-range dipole-dipole interaction (Förster's mechanism), occurring without the emission of a photon. The main requirements for FRET are donor and acceptor molecules are close to each other; a spectral overlap between the emission spectrum of the donor, and the absorption spectrum of the acceptor; and the alignment between donor and acceptor dipoles.

The Förster distance (R_0) is the distance in which FRET is 50% efficient, which is typically between 10 to 60 Å. FRET efficiency (E) depends on the distance (R) between donor and acceptor, and can be calculate as Equation 1 in the presence of an acceptor.[22] In the presence of a fluorescent acceptor, such as Trp, the relative FRET efficiency (E_{rel}) can be calculated using the ratio between donor (I_D) and acceptor (I_A) emission intensity (Equation 2).[39] FRET rate can be used to measure the distance between donor and acceptor, and changes in spectral features of a sample, such as the unfolding of a protein.[22, 24]

$$E = \frac{1}{1+\left(\frac{R}{R_0}\right)^6} \qquad \textbf{Equation 1}$$

$$E_{rel} = \frac{I_A}{I_D + I_A} \qquad \textbf{Equation 2}$$

The fluorescence spectra of a protein are an overlap of Tyr and Trp exposed to different environments.[22, 46, 47] FRET can occur due to the Tyr and Trp spectral overlap (Figure 1.3), and Tyr-to-Trp FRET is known to be the most common energy transfer process in protein. IgGs typically have ~50/60 Tyr and ~20/30 Trp fluorophores (Figure 1.1), many are in proximity (≤20 Å), and FRET is expected. We know that homo-, and hetero-FRET rates will be high, and have a large effect on the measured emission. For proteins in their native state, emission of Tyr is frequently quenched as a result of energy transfer to Trp. As protein unfolds, this energy transfer is reduced, and an increase in Tyr emission[b] is often observed in non-native protein.[22, 48, 49]

1.3.3 Inner filter effect (IFE)

The fluorescence emission is detected in the middle of the cuvette (Figure 1.5), meaning that the light can be attenuated before reaching the detector, decreasing fluorescence intensity in the region of spectral overlap, and modifying spectral shape. This is called inner filter effect (IFE). IFE depends on non-molecular properties of the sample, such as the path length of the sample container, optical density of the sample (Absorbance, A), and the geometry arrangement of the excitation and emission paths.[22, 24]

High optical density could attenuate the recorded fluorescence intensity and may result in a red shift (to longer emission wavelength) due to changes in spectral shape.[50-52] Most of the fluorescence spectrophotometers use a right-angle geometry (as used in this thesis), and a 10 mm pathlength quartz cuvette, meaning that IFE will be significant even for relatively low concentration solutions. The IFE in the right-angle geometry (Figure 1.5) can be referred as primary IFE, attenuating the excitation light; and, secondary IFE, attenuating the emission light.[50] The total attenuation of fluorescence caused by IFE at each excitation/emission wavelength pair is a function of the absorbance at each wavelength pair, and the cuvette pathlength (Figure 1.5). As

[b] A change in Tyr-to-Trp FRET process is observed as a simultaneous decrease in Trp and an increase in Tyr emission.

the absorbance spectrum decreases with increasing wavelength (Figure A-1), IFE is more severe at shorter wavelengths (blue edge), affecting the shape, maximum peak of the fluorescence spectra, and red-shifting the emission with the increase in sample concentration.[50]

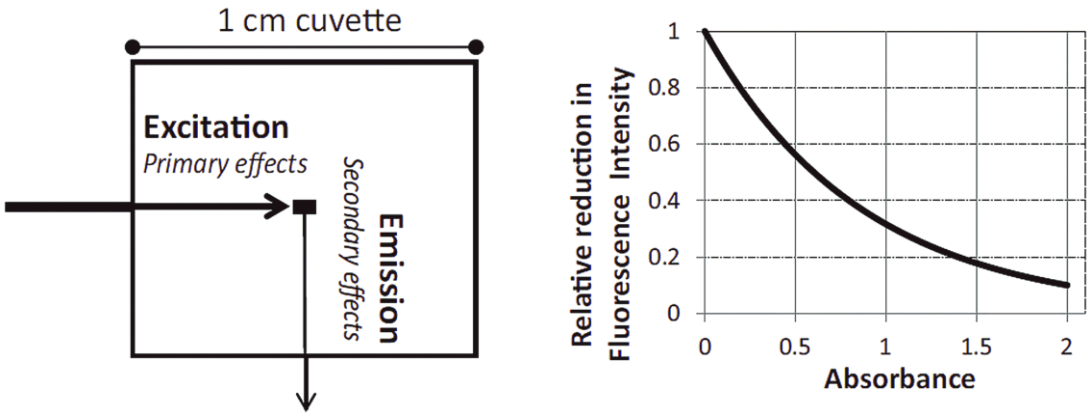

Figure 1.5: (Left) Schematic of the primary and secondary IFE in a 10×10 mm cuvette with a right-angle geometry. The total attenuation of fluorescence intensity caused by primary and secondary IFE can be estimated from the absorbance spectrum of the sample. (Right) Plot showing the relative loss in fluorescence intensity due to IFE as a function of increasing A_{Total} (absorbance at any reference excitation/emission wavelength pair) in a 10×10 mm cuvette. Reproduced from Kothawala et al. (2013),[50] with permission of John Wiley and Sons.

The four methods used for dealing with IFE are sample dilution, pathlength change, mathematical correction, or explicit inclusion of the IFE.[53] The simplest way to avoid IFE is to reduce sample concentration, but this is not always a viable option. Reducing sample concentration involves changing the sample and requires sample handling, introducing a source of error. In addition, diluting a sample would not simply reduce optical density, but would be accompanied by a decrease in fluorescence intensity, a very poor signal-to-noise ratio (SNR), and increased light scatter.[54] Thus, avoiding IFE reduces spectral resolution, and thus a mathematical correction is often preferred. The mathematical correction (Section 2.4.3) approach can be used when the sample optical density is <1.5 (in a 10×10 mm cuvette).[50] Reducing the cuvette's pathlength lowers the absorbance, enabling optimal IFE corrections.[55] When it is not possible to correct for IFE (*e.g.* A>1.5), the alternative is to accept that IFE is a characteristic of the sample,[1, 56] which can be useful in some cases.

1.4 Anisotropy resolved multidimensional emission spectroscopy (ARMES)

ARMES combines anisotropy, multidimensional fluorescence spectroscopy (MDF) measurements, and chemometrics[1] analysis to try to extract more information from multifluorophore proteins. ARMES uses anisotropy as an additional layer of photophysical information in combination with the excitation wavelength (λ_{ex}), emission wavelength (λ_{em}) or wavelength offset ($\Delta\lambda$), and intensity (I). This extra information can help to provide additional information with which one can characterize proteins and differentiate the emission of each fluorophore or a class of fluorophores. The combination of anisotropy and MDF has previously enabled the differentiation and quantification of fluorophores with similar emission properties in complex mixtures, based on their rotational speed and hydrodynamic volume, and thus the molecular size, or for macromolecules the mobility/flexibility of the constituent fluorophores.[1, 3, 57]

1.4.1 Fluorescence anisotropy

The base of fluorescence anisotropy lies on the principle of photoselective excitation of fluorophores by polarized light.[22, 24] According to this principle, photons with electric vectors aligned parallel to the transition moment of the fluorophores will be preferentially absorbed. The transition moment of a fluorophore has a particular orientation (Figure 1.1) with respect to its molecular axis, but the fluorophore will be randomly oriented in an isotropic solution. Exciting a sample with polarized light allows one to selectively excite those fluorophores whose absorption alignment is parallel to that of the excitation light (Figure 1.6). Upon excitation, molecules rotate during the excited-state lifetime, with the anisotropy being determined by the degree of rotational diffusion. Small molecules rotate relatively quickly, decreasing anisotropy values, bigger molecules rotate more slowly, increasing anisotropy.

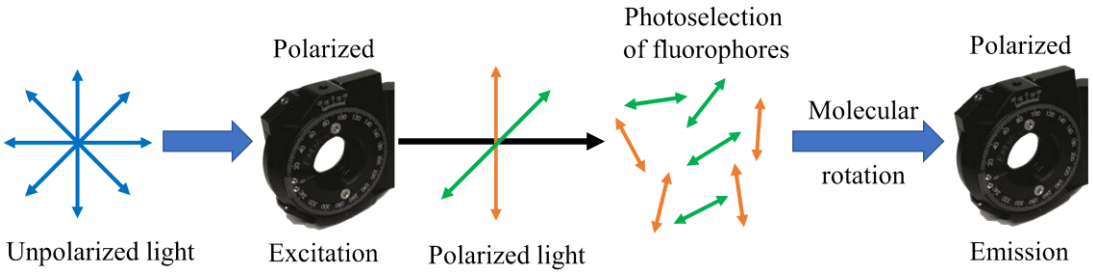

Figure 1.6: Polarization of incident light by a vertically (orange) and horizontally (green) oriented polarizer, resulting in the photoselection of randomly distributed fluorophores.

The degree of emission polarization can be expressed as anisotropy (r) or polarization (P), as defined by Equation 3, and Equation 4, respectively; where $I_{\parallel}$ and $I_{\perp}$ are the fluorescence intensities of the vertically ($\parallel$), and horizontally ($\perp$) polarized emission, obtained when the sample is excited with vertically polarized light.[24, 58] Anisotropy and polarization describe the same phenomenon, and these values are interchangeable (Equation 5), however, the use of anisotropy is preferred.[59] As observed in Equation 3, the difference in intensity between the polarizations ($I_{\parallel}-I_{\perp}$) is normalized by the total intensity ($I_T=I_{\parallel}+2\times I_{\perp}$), making the anisotropy values independent of the total emission intensity.[c]

$$r = \frac{I_{\parallel}-I_{\perp}}{I_{\parallel}+2\times I_{\perp}} \qquad \textbf{Equation 3}$$

$$P = \frac{I_{\parallel}-I_{\perp}}{I_{\parallel}+I_{\perp}} \qquad \textbf{Equation 4}$$

$$r = \frac{2\times P}{3-P} \leftrightarrow P = \frac{3\times r}{2+r} \qquad \textbf{Equation 5}$$

Four polarized intensity measurements, I_{VV}, I_{VH}, I_{HV}, I_{HH} (V=vertical, H=horizontal; the first letter corresponds to the orientation of the excitation, and the second the emission polarizers), are necessary to calculate emission anisotropy. The detector channel can have different sensitivity to the vertical and horizontal components of the polarized fluorescence emission, and this is also wavelength dependent.[60] This difference can be corrected for each wavelength by using the G-factor (Equation 6), which can be calculated by exciting the sample horizontally and detecting the emission at both vertical and horizontal polarizations (HV and HH).

[c] Anisotropy is dimensionless, meaning that absolute anisotropy values obtained from different machines, and/or laboratories can be compared.

Thus, the horizontally polarized excitation is necessary to calculate the true anisotropy (Equation 7).[22, 60]

$$G = \frac{I_{HV}}{I_{HH}} \quad \textbf{Equation 6}$$

$$r = \frac{I_{VV} - G \times I_{VH}}{I_{VV} + 2 \times G \times I_{VH}} \rightarrow r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2 \times I_{\perp}}; \text{ where } I_{\parallel} = I_{VV}, \text{ and } I_{\perp} = G \times I_{VH} \quad \textbf{Equation 7}$$

The G-factor is not the only correction necessary to recover the true anisotropy. Anisotropy depolarization can be caused by trivial reasons, such as light scattering (*e.g.* Rayleigh and Raman scattering), IFE, misalignment, and/or inefficiency of the polarizers.[58] Although Rayleigh scattering and IFE are generally associated with experimental problems related to the optical condition of the experiment,[50, 60, 61] Rayleigh scatter can provide useful information about the sample.[56, 57, 62, 63] Scattered light is highly polarized (r=1), and a small portion of scattered light can cause a significant change in anisotropy,[54, 64] generally increasing the measured anisotropy value. In the presence of IFE, the reabsorption of emitted photons means that anisotropy values will be reduced, even at low concentration.[22, 54, 58, 64] While it is possible to correct the spectra for light scattering and IFE, one must ensure the correct alignment of the polarizers, and to select the best polarizer for the wavelength range being used.[57]

In the absence of depolarizing processes, maximum measured anisotropy (r_0) range between −0.2 and 0.4,[36, 45] but these values are reduced in biomolecules[32, 58, 65] due to rotational diffusion, FRET, and protein flexibility. In non-viscous solutions molecules can freely rotate, and the orientation of the polarized emission is randomized, resulting in anisotropy values near to zero.[22] Rotational diffusion can be reduced or eliminated using low temperatures, and/or high viscosities. Changes in anisotropy due to depolarization can give information on the size and shape of molecules, and, mobility and microenvironment of fluorophores.[22, 24, 58, 59, 66] For multifluorophore proteins like albumins, insulin, and IgG, anisotropy is not constant across the emission space, and can vary significantly.[57, 67] This heterogeneity is due to the interlinked emission from all the fluorophores within protein structure, which

will change upon structural alteration. Thus, changes in anisotropy can be used to measure protein denaturation, and aggregation.[d]

1.4.2 Multidimensional fluorescence spectroscopy (MDF)

For the analysis of multifluorophore complex mixtures or proteins, MDF provides a convenient approach to extracting more information in comparison to conventional two-dimensional (2D) fluorescence spectroscopy. The commonest MDF measurement methods are excitation-emission matrix (EEM)[68] and total synchronous fluorescence spectroscopy (TSFS).[69] Three-dimensional (3D) EEM and TSFS spectra provide a spectral signature of the multiple fluorophores presents in a protein or a complex multifluorophore sample.[67, 70, 71]

1.4.2.1 Excitation-emission matrix (EEM)

An EEM spectra is composed of several excitation and emission spectra, collected by keeping a fixed λ_{ex}, and scanning the emission. EEM data generally includes Rayleigh first ($\lambda_{ex}=\lambda_{em}$) and second ($\lambda_{ex}=2\times\lambda_{em}$) order, which appear as multiple diagonal bands in the EEM spectra (Figure 1.7), and Raman light scattering. Rayleigh scattering can be used to monitor particle formation, such as aggregation caused by physical and chemical stressors.[72]

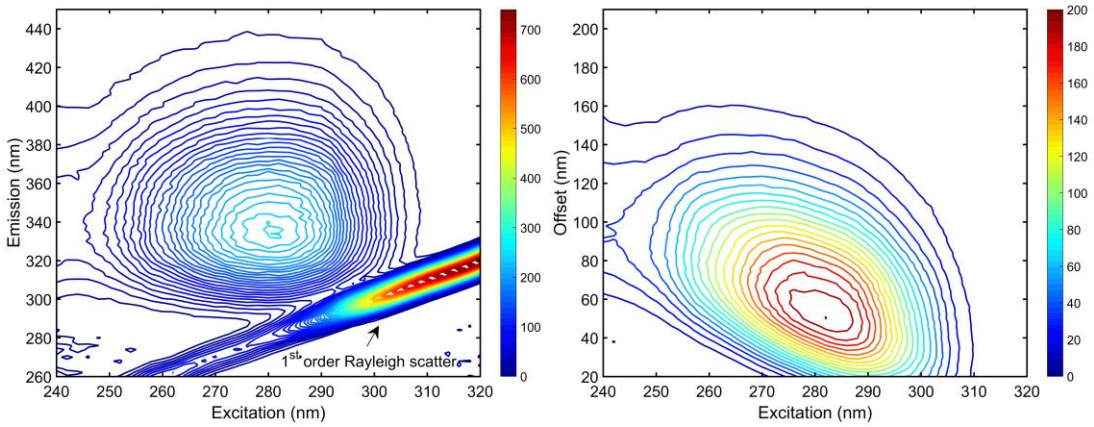

Figure 1.7: (Left) EEM and (right) TSFS contour plots of a rIgG measured at 20°C. The first order Rayleigh scatter appears as a diagonal band in the EEM plot. Visually, there is no Rayleigh scatter in the TSFS plot.

[d] Protein aggregates are a cluster of multiple molecules, meaning that there is a significant increase in size, decreasing rotational rate, and increasing anisotropy.

1.4.2.2 Total synchronous fluorescence spectroscopy (TSFS)

TSFS spectra are comprised of a series of synchronous fluorescence (SF) collected over a series of λ_{ex} in a wavelength offset ($\Delta\lambda$), collected when the excitation and emission are changed at the same time with a constant delta (Figure 1.7). This difference in data acquisition gives some advantages for TSFS over EEM: *(i)* the degree of Rayleigh scatter collected can be reduced at the collection point; *(ii)* TSFS has a more distinctive fluorescence signature.[71]

The fact that Rayleigh scatter can be reduced during data collection for TSFS measurements requires much less, or no data pre-processing for light scattering removal in comparison with EEM. Thus, TSFS measurements may preserve more spectral information from the emission blue edge, particularly the Tyr components. However, one must understand the TSFS data prior to any data analysis. TSFS structure is intrinsically different from EEM (Figures 1.7/2.4) and requires precise pre-processing steps before implementing curve resolution-based methods for their analysis.[71] The only way of applying PARAFAC analysis for the IgG measured with TSFS is restructuring the TSFS into an EEM like layout (t-EEM) to generate the trilinearity[71, 73-75] required. This conversion is possible because the equations for fluorescence intensity spectroscopy are correlated,[71, 73] where I_{EEM} and I_{TSFS} are the fluorescence intensity for the EEM (Equation 8), and TSFS (Equation 9) measurements, respectively, K is the instrumental factor constant, c is the fluorophore concentration, d is the cuvette pathlength, which is different for excitation and emission here, E_X and E_M are the excitation and emission profile, and λ_{ex} and λ_{em} are the excitation and emission wavelength. From the EEM and TSFS equations, it is possible to derivate Equation 10, which is used to rearrange the TSFS into t-EEM.

$$I_{EEM}(c, \lambda_{ex}, \lambda_{em}) = KcdE_X(\lambda_{ex})E_M(\lambda_{em}) \qquad \textbf{Equation 8}$$

$$I_{TSFS}(c, \lambda_{ex}, \Delta_\lambda) = KcdE_X(\lambda_{ex})E_M(\lambda_{ex} + \Delta_\lambda) \qquad \textbf{Equation 9}$$

$$I_{EEM}(c, \lambda_{ex}, \lambda_{em}) = I_{TSFS}(c, \lambda_{ex}, \Delta_\lambda) \rightarrow \lambda_{em} = \Delta_\lambda + \lambda_{ex} \qquad \textbf{Equation 10}$$

1.4.3 Chemometrics

The analysis of this multidimensional data requires the use of chemometric modelling techniques to accurately resolve contributions from different fluorophores in the protein. Chemometrics combines mathematical and statistical tools to retrieve

chemical information,[76-78] allowing one to extract relevant, and useful analytical information from complex MDF data.[79, 80] For ARMES, multi-way decomposition methods, like multivariate curve resolution (MCR),[81, 82] and parallel factor analysis (PARAFAC)[83-85] are required to identify the spectral contribution of individual constituents. These contributions can then be associated with specific fluorophore emission.

Previously, it was showed that ARMES of Human Serum Albumin (HSA) using MCR modelling yielded multiple components,[86] however these studies used thin film polarizers (TFP), with no transmission below 300 nm, which affected the TSFS data structure. Further developments of ARMES used dual wire grid polarizers (dWGP) to enable short (<300 nm) wavelength excitation.[87] Validation measurements using dWGP have shown that one can accurately recover individual fluorophore emission in the absence of IFE, and FRET.[88] These early studies also undertook chemometric (MCR) analysis on all four polarization measurements (VV, VH, HV, and HH), mainly in an effort to confirm that the chemometric modelling was robust.[87, 89] The use of MCR in the initial development of ARMES was justified by the use of TSFS measurements. As the TSFS data lacks the trilinearity required for PARAFAC analysis, MCR was the selected multi-way decomposition method.[1, 2] However, further development of ARMES using dWGP and TSFS measurements showed that the resolution of meaningful components by MCR was not possible,[57] and the trilinearity intrinsic to the EEM measurements[90] was required to enable the use of MCR. The best MCR models resolved for the EEM data were obtained using trilinear constraints,[88] which can only be applied when no IFE and FRET are present. While it is possible to correct for IFE, FRET is an intrinsic part of the protein signal, and it is not possible to correct for FRET. As FRET is present in complex proteins, it was clear that the use of PARAFAC should be preferred over MCR.[62] Here, the full understanding of the TSFS data allowed the use of PARAFAC for the IgG measured with both EEM, and TSFS techniques.[91, 92]

PARAFAC is one of the most popular curve resolution methods[79] that requires the three-way array data ($\underline{\mathbf{X}}$) to be trilinear in nature. PARAFAC (Equation 11)[83] is a trilinear model method with two independent sets of variables (excitation and emission profiles), and one variable dependent on both spectral profiles.[79, 83, 85] The PARAFAC model of a three-dimensional array is given by three loading matrices, **A** (samples), **B**

(λ_{em}), and **C** (λ_{ex}) with elements a_{if}, b_{jf}, and c_{kf}, where $i = 1, \ldots, I$; $j = 1, \ldots J$; $k = 1, \ldots, K$; and F is the number of components. The sum of squares of residuals in the model (**E**), e_{ijk}, is minimized by the trilinear model.

$$x_{ijk} = \sum_{f=1}^{F} a_{ij} b_{jf} c_{kf} + e_{ijk} \qquad \textbf{Equation 11}$$

EEM three-way array data generally fulfil this trilinearity requirement. Still, the presence of a diagonal Rayleigh and Raman scatter adversely affect PARAFAC analysis. Rayleigh and Raman scattering are non-trilinear data, and these diagonal signals must be eliminated (or at least reduced) by the data pre-processing.[71, 85] In contrast, TSFS three-way arrays are not trilinear because the offset ($\Delta\lambda$) and excitation modes are not independent. Consequently, it would not be possible to apply a trilinear decomposition, such as PARAFAC, to TSFS data, unless pre-processed to give the trilinearity required. Converting the TSFS into EEM three-way layout (Equation 10), allowed the use of PARAFAC decompositions of the multifluorophore protein solution.

Before any chemometric method can be applied, the data must be pre-processed to correct for systematic bias in the dataset (*e.g.* instrumental factor, IFE, and G-factor corrections), remove signals unrelated to protein fluorescence (Rayleigh and Raman scattering), and normalize the datasets to remove large differences in intensity between samples.[53, 60, 61, 85]

Every instrument is different, and various aspects (*e.g.* lamp, PMT voltage) can affect spectral shape, and fluorophore maximum peaks,[22, 53, 61] meaning that the excitation and emission fingerprint of a sample will vary between instruments. Thus, one cannot compare its results with literature absolute values, unless appropriate excitation and emission instrumental correction (Section 2.4.1) are applied to the spectra. One must correct for instrumental factor to guarantee that any changes observed in the spectrum are related to the sample, not to the instrument. The spectral shift caused by the instrument itself can affect the analysis of stressed proteins based on changes in Trp classes. The literature is clear about the fluorescence emission for different Trp classes,[35, 93] making the instrumental correction a critical factor here. Thus, if Trp emission is not correct for instrumental factor, a ~10 nm red shift could indicate that the Trp is fully exposed to the solvent (λ_{em}=350−353 nm), meaning that

the protein was completely unfold, while it was actually only partially exposed to the solvent (λ_{em}=340−342 nm).

IFE correction was critical here for accurate PARAFAC resolution and anisotropy calculations because the fluorescence intensity obtained with polarized measurements were much weaker than normal, unpolarized MDF measurements.[86] Rayleigh and Raman scattering[22] are non-trilinear, and should not be present in the data for PARAFAC analysis[71, 85] of multifluorophore mixtures like proteins. Thus, Rayleigh and Raman scattering must be removed, or at least attenuated, from the data prior to undertaking PARAFAC analysis on the emission component of the measurement.

1.5 Outline of the thesis

Chapter 2 describes the general methods used for sample preparation, data collection, and how the data had to be handled to enable the use of PARAFAC analysis for monitoring protein structure and stability. The pre-processing steps that were common for the data collected with the different pMDF measurements are explained in Chapter 2. The steps that were exclusively applied for either pEEM or pTSFS measurements are described in Chapters 3, 4, and 5.

The photophysical characterization of the rIgG in its native state (15−35°C) was measured with pEEM (Chapter 3) and pTSFS (Chapter 4). Chapter 3 shows the first attempt to use a combination of pEEM measurements and PARAFAC analysis for the photophysical characterization of an IgG type protein in solution. The main challenge in this chapter was to solve the issues related to the data to enable a trustworthy resolution. Even though several methods were applied, the main outcome was that we could characterize IgG structure in its native state, but we could not discriminate different fluorophore populations as previously described using ARMES. For the pEEM data, it was clear that the Rayleigh scattering was having a major impact on PARAFAC analysis when there were no significant structural changes.

To try to reduce Rayleigh scatter influence, the pMDF method was changed. Chapter 4 uses pTSFS for the photophysical characterization of IgG to see if this could improve PARAFAC resolution. However, TSFS has a different data structure, which had to be accounted for before PARAFAC analysis. Thus, the pTSFS data were handled to match the pEEM data structure, allowing a careful comparison of

chemometric resolution of both pEEM and pTSFS data. The reduction of Rayleigh scatter influence improved PARAFAC resolution, with two components (Trp and Tyr) being resolved for the pTSFS data. This was the first time that Tyr and Trp fluorophores were resolved from IgG data measured with MDF techniques, indicating that the use of pTSFS was more suitable for the analysis of small structural changes.

Chapter 5 shows that once the IgG native state was understood, it was possible to use pMDF/PARAFAC to monitor protein structure, stability, and aggregation for the first time. Even though the use of pTSFS was superior in comparison with pEEM for the analysis of subtle structural changes, the use of pEEM is still preferred over pTSFS when significant structural changes are present. Large structural changes can be followed or accompanied by changes in particle size and aggregation, which can be extracted from the Rayleigh scatter. Rayleigh scatter and protein spectral information can be collected simultaneously with pEEM measurements, enabling the analysis of protein structural stability, changes in particle size, and aggregation with a single measurement. PARAFAC analysis of thermally stressed (20−70°C) IgG solutions showed that it was possible to discriminate the native from the non-native protein structure.

2. Materials and Methods

This chapter describes the materials, sample preparations, instruments, and collection parameters used for the experiments presented in this thesis. The pre-processing steps and data analysis that were common to all the experiments are also described here.

2.1 Materials

Polyclonal IgG from rabbit serum (≥95% essentially salt-free, lyophilized powder) were purchased from Sigma-Aldrich (lot #SLBM2617V) and used without further purification. High-performance liquid chromatography-size exclusion (HPLC-SEC) analysis of the rabbit IgG (rIgG) used here showed a monomer content of ~80±1%.[e] Sodium phosphate monobasic, sodium phosphate dibasic heptahydrate, and sodium chloride were used to prepare a 0.01 M Phosphate 0.150 M saline buffer (PBS) at pH 6.5±0.1 in high purity water (HPW). HPW was purchased from Honeywell (HPLC grade), all the other materials were purchased from Sigma-Aldrich and used as received. Isopropanol and HPW were used to clean up the working space, glassware, and cuvettes. Cuvettes were cleaned up with a 2% Hellmanex™ III (Hellma Analytics) solution at the end of each experiment set.

2.2 Instrumentation and parameters for data collection

UV-visible absorbance spectra (200−800 nm) were collected using a Cary 60 UV-vis spectrophotometer (Agilent, product No. G6860A, serial #MY14090017) and a temperature controller single-cell holder Peltier accessory (Agilent, product No. SPV 1X0, serial #1411216) at a scan rate of 1200 nm min^{-1}. All spectroscopic measurements were carried out using 10×2 mm pathlength quartz cuvettes (Lightpath Optical, UK). Absorbance spectra were collected with a 2 mm pathlength; for the MDF measurements, samples were excited along the short axis (2 mm), and emission collected from the long axis (10 mm) to reduce IFE.[94]

[e] HPLC-SEC of the rIgG was collected (22/08/2018) and analysed by Ana Luiza de Faria e Silva. Data collected after one freeze-thaw cycle. No HPLC-SEC was collected for the freshly prepared rIgG solutions.

EEM, TSFS, and ARMES, were performed using a Cary Eclipse fluorimeter (Agilent, product No. G9800A, serial #MY15350008) fitted with bespoke dual wire grid polarizers (dWGP[f]),[57, 88] and a temperature-regulated multi-cell holder (Agilent, product No. G9844A, serial #MY15290007). A validation of the Cary Eclipse was performed at the beginning of each set of experiments.[g] The Cary Eclipse used here (serial #MY15350008) passed in all validation parameters every time it was tested.

Polarized EEM (pEEM) data were collected over an excitation and emission range of λ_{ex}=240−320 nm and λ_{em}=260−450 nm using 2 nm step increments. Polarized TSFS (pTSFS) spectra were collected over an excitation range of λ_{ex}=240−320 nm and $\Delta\lambda$ interval of 20−210 nm at 2 nm step increments in each case. Slit width of excitation and emission monochromators were 10 nm, scan rate 1200 nm min^{-1}, and the photomultiplier tube (PMT) detector voltage was set to 650 V. It took ~8−10 minutes to collect each polarized spectrum (pEEM and pTSFS, respectively).

All samples were measured with four different polarizer settings: VH (vertical-horizontal), VV (vertical-vertical), HH (horizontal-horizontal), and HV (horizontal-vertical). The anisotropy at each emission wavelength was calculated using the standard anisotropy formula,[22] which was then used to construct the corresponding *aniso*-MDF maps.

2.3 Sample preparation and data collection

PBS (0.01 M Phosphate 0.150 M saline pH 6.5±0.1) was prepared adding ~0.072 g of sodium phosphate monobasic, ~0.0804 g of sodium phosphate dibasic heptahydrate, and ~0.781 g of sodium chloride in a 100 mL Fisherbrand amber volumetric flask, which was filled to 100 mL with HPW. PBS solutions were shaken 20 times, sealed with parafilm, and stored at 4−8°C. The same PBS solution was used to prepare the replicate samples of each designed experiment, avoiding small variations in buffer composition. The HPW was not sterilized, PBS buffer was membrane filtered (0.2 µm) using polyethersulfone (PES) Captiva Premium Syringe

[f] Composed of two Ultra broadband WGP (ThorLabs, 25 mm, product No. WP25M-UB).

[g] A software called 'validate' was used for this purpose to check some parameters, such as: *(i)* the accuracy and reproducibility of excitation and emission wavelengths using a Xenon lamp; *(ii)* the excitation and emission spectral bandwidth accuracy (Xenon lamp); *(iii)* the accuracy of 0% transmittance for Rayleigh scatter and Raman band using a certified water standard (Starna, 3Q-10-Water); *(iv)* the lack of stray light (ground silica diffuser); and *(v)* Raman water sensitivity at excitation 350 and 500 nm, to make sure that the SNR was bigger than the tolerance for Raman water band.

filters (Agilent) fitted in 20 mL syringes (BD Plastik™) immediately before being used for rIgG sample preparation. An aliquot of PBS was used to collect blank pEEM and pTSFS spectra (10 replicate measurements) to calculate the Limit of reporting (LOR) used for IFE correction (Section 2.4.3).

All rIgG sample preparation was carried out in a laminar flow hood (LFH) using aseptic techniques, and membrane filtered (0.2 μm) using PES Captiva Premium Syringe filters (Agilent) with 20 mL syringes (BD Plastik™), to minimize contamination. The rIgG powder is not very dense, and is sensitive to static electricity, making it hard to recover it from the vial. The first rIgG solution was prepared in a volumetric flask (Fisherbrand amber, 10 mL), but foam formation was a real problem, requiring ~40 minutes for the foam to dissipate. Thus, rIgG solutions were prepared in its original vial, adding 10 mL of PBS to the vial to minimize sample manipulation and foam formation. The vial was closed with its own cap, shaken 30 times with figure "8" movements (on the bench in the LFH). Still, there was some minor foaming, which was reduced after ~10 minutes. Sometimes the rIgG powder was spread over the vial walls, if this was seen, then the vial would be gently manipulated to try to get all the powder into solution. For the final filtration step, the 20 mL syringe was first filled with 1 mL of air and then filled with the rIgG solution (using a metal needle), and an extra mL of air. The last steps were to attach the filter to the syringe and to apply pressure, filtering the solution to a becker, which was immediately covered with aluminium foil.

Freshly prepared rIgG solutions were either pipetted (Eppendorf ep Dualfilter T.I.P.S.® LoRetention) into 10×2 mm pathlength quartz cuvettes (Lightpath Optical, UK) or aliquot (750 μL) into 1.5 mL Protein LoBind Tubes (Eppendorf) for storage (−70°C) prior to use. Eppendorf ep Dualfilter T.I.P.S.® LoRetention® and LoBind® Tubes were sterile and kept in the LFH once the box was opened to avoid contamination. Non-sterile Eppendorf epT.I.P.S.® LoRetention® were used to recover the defrosted aliquots from the Protein LoBind® Tubes to the quartz cuvettes outside the LFH immediately before measurement. The same cuvette was used for each set of experiments to reduce variations that could have been caused by the cuvette itself. Also, the cuvette was always positioned in same cuvette holder (#1) to avoid variations that may arise from the use of a different position in the multi-cell holder.

2.3.1 Sample preparation for native state characterization and to monitor structural stability and aggregation of rIgG under thermal stress

The sample preparation was carried out in triplicate. Three different vials of rIgG (10.3 mg), with the same lot number (#SLBM2617V), were reconstituted with 10 mL of PBS buffer and membrane filtered, yielding three freshly prepared solutions with concentrations[h] of: R1=1.3 mg mL^{-1}, R2=0.9 mg mL^{-1}, and R3=1.2 mg mL^{-1} (Table 2.1). Freshly prepared rIgG[i] (1.1±0.2 mg mL^{-1}) were measured (pEEM) over a temperature range of 20−75°C (5°C increments), and again after cooling down to 20°C (13 different temperatures), with 5 minutes equilibration time between each temperature. No pTSFS measurements were made on freshly prepared samples.

Table 2.1: The rIgG solutions were prepared from the same lot number, but from three different vials (R1, R2, and R3), and stored at −70°C. Freshly prepared samples were used for the thermal stress experiment, and samples with one freeze-thaw (FT) cycle were used for the native state experiments. The UV-vis absorbance at 280 and 350 nm measured for the rIgG solutions (buffer subtracted) were used to calculate protein concentration,[95, 96] and aggregation index[99, 100] ($\boldsymbol{AI}_{\%} = \left(\frac{\boldsymbol{A_{350}}}{\boldsymbol{A_{280}-A_{350}}}\right) \times \mathbf{100}$) at 20°C.

	EEM Thermal stress			**TSFS Native state**			**EEM Native State**		
	R1	**R2**	**R3**	**R1**	**R2**	**R3**	**R1**	**R2**	**R3**
Lot number	#SLBM2617V			#SLBM2617V			#SLBM2617V		
Sample preparation	07/04/17	12/04/17	21/04/17	07/04/17	12/04/17	21/04/17	07/04/17	12/04/17	21/04/17
Measurement date	07/04/17	12/04/17	21/04/17	27/06/17	04/07/17	11/07/17	14/07/17	18/07/17	21/07/17
UV-vis @280 nm	0.3749	0.2452	0.3374	0.3668	0.2449	0.3350	0.3713	0.2453	0.3393
UV-vis @350 nm	0.0025	0.0009	0.0021	0.0049	0.0048	0.0062	0.0047	0.0049	0.0076
Sample concentration (mg mL^{-1})	1.3	0.9	1.2	1.3	0.9	1.2	1.3	0.9	1.2
Aggregation index (%)	0.7	0.4	0.6	1.4	2.0	1.9	1.3	2.0	2.3

For the native state characterization, rIgG samples were slowly defrosted overnight at 4−8°C, taking care to ensure that there were no ice crystals present before transferring into the cuvettes. Protein concentration in solution was not changed by the defrosting process[j] (Table 2.1). Native state rIgG was measured (pEEM and

[h] Protein concentration was calculated from the absorbance spectra at 280 nm[95] $\left(C = \frac{A_{280}}{E_{280}^{1mg/mL} \times l}\right)$ where E is the absorption extinction coefficient, $E_{280}^{1mg/mL} = 1.4$ for the rIgG,[96] and l is the cuvette pathlength (0.2 cm). Absorbance spectra were collected at 20°C (Figure A-1).

[i] Freshly prepared rIgG solutions were immediately thermally stressed, avoiding any structural changes that could have been caused by the freeze-thaw cycle. Various factors affect protein structure/particle size during freeze and thaw procedures, such as the speed of freeze and thaw, which would be influenced by the position/number of samples stored in the freezer.[43, 97, 98]

[j] The freeze-thaw process did not cause a change in protein concentration, meaning that the use of LoBind tubes was effective to avoid protein loss. However, there was a clear increase in insoluble

pTSFS) at 11 different temperatures (10°, 15°, 17°, 20°, 23°, 25°, 27°, 30°, 33°, 35°C, and cooled down to 10°C), with 5 minutes equilibration time between measurements. The measurements undertaken at 10°C were compromised by condensation on the external walls of the cuvette[k] and were not used for further data analysis. The thermal fluctuation induced by the 15−35°C temperature range should be enough to cause simple structural fluctuation, without changing rIgG native state conformation. This structural fluctuations were likely to be caused by changes in local motion and flexibility,[101, 102] which would be mainly reflected in a change in Tyr-to-Trp FRET efficiency, leading to spectral changes.

Initial studies in the development of ARMES[57, 88] showed that the accurate removal of Rayleigh scatter from the pEEM made this a more efficient measurement mode for ARMES in comparison with pTSFS. Thus, pEEM was the best option to collect data regards structural changes and aggregation profile at the same time. However, after the analysis of the rIgG native state measured with pEEM and pTSFS we could conclude that pTSFS was indeed a better option for extracting information regards small structural changes (Chapter 4). Still, the thermal stress experiments were not repeated with pTSFS because it was latter shown (Chapter 5) that there was no real advantage in collecting/analysing a full pMDF spectrum over a simpler 2D spectral analysis.

2.4 Data analysis and chemometric methods

Data analysis was performed using the PLS_Toolbox ver. 8.2.1 (Eigenvector Research Inc.), MATLAB ver. 9.1.0 (The MathWorks Inc.), and in-house written codes (FluorS). MDF data had to first be pre-processed to make the data suitable for data analysis.

aggregates (as calculated by the AI). UV-vis analysis of the defrosted rIgG (Figure A-1A) showed changes in the non-absorbing region of the absorbance spectra (>320nm), indicating increased particle size, and aggregation.

[k] The room temperature of ~18/20°C was enough to create a condensation in the external walls of the cuvette during equilibration time. In addition, the need to open the machine to manually change the polarizers made it impossible to control the condensation on the external walls of the cuvette. Once the temperature was increased to 15°C, the difference between experimental and room temperature was not enough to cause the same phenomenon.

2.4.1 Instrumental correction

The emission correction factor was calculated as previously described[103] using a Spectral Fluorescence Standard Kit (Sigma, product No. 69336), certified by the Federal Institute for Material Research and Testing (BAM, Germany). Excitation correction factor was calculated with a concentrated solution of Rhodamine B sealed in a triangular cell (Agilent, part No. 6610021700). EEM data were collected over an excitation and emission range of λ_{ex}=200−600 nm (Rhodamine B), and λ_{em}=300−770 nm (Table A-1) using 1 nm step increments. Slit width of excitation and emission monochromators were 10 nm, scan rate 120 nm min^{-1}, integration time 0.5 s, and the PMT voltage was set to 510 V for the 10 replicate measurements of Rhodamine B. For the 10 replicate measurements of the Spectral Fluorescence Standard kit, slit width of excitation and emission monochromators were 5 nm, scan rate 120 nm min^{-1}, integration time 0.5 s, and the PMT voltage was set to 590 V. The correspondent blank (ethanol) was measured 10 times. Due to the restriction in emission wavelength that the dyes provided in the kit could be measured (300−770 nm), the pEEM emission spectra had to be reduced to λ_{em}=302−450 nm for the instrumental factor correction to be applied. The excitation and emission instrumental factor correction (Figure 2.1) was only implemented after the pEEM[91] and pTSFS[92] rIgG native state[l] were already published, and these papers did not include a correction for instrumental response.

[l] There were no changes in Trp classes for rIgG in its native state (Table A-2). Thus, once we were not comparing emission from different Trp classes with those describe in the literature, instrument correction was not trivial for the analysis of the native state.

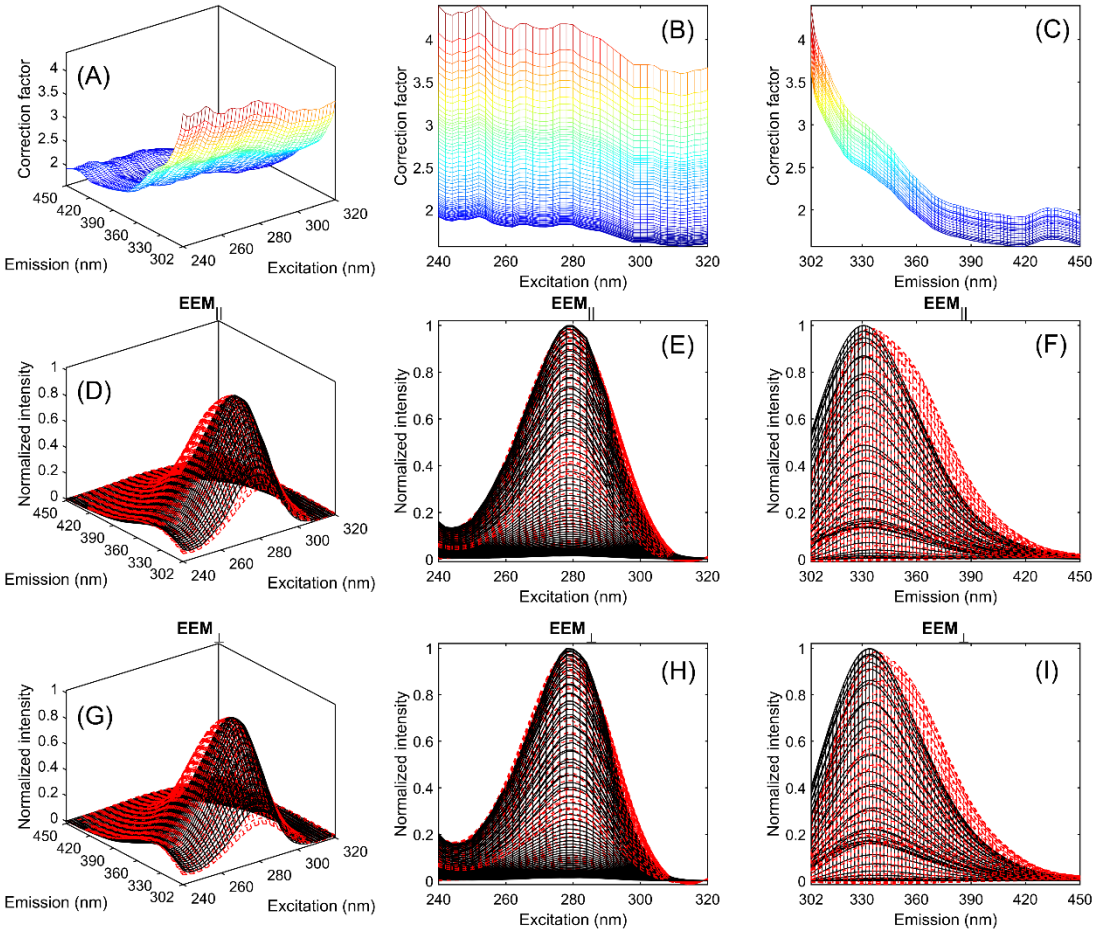

Figure 2.1: (A−C) Instrumental correction factor calculated from the non-polarized EEM spectra for the Cary Eclipse used here. Normalized pre-processed (D−F) pEEM$_{\parallel}$ and (G−I) pEEM$_{\perp}$ spectra for the rIgG measured at 20°C (red spectra) before, and (black spectra) after instrumental factor correction. Instrumental correction was used to correct spectral shape.

2.4.2 G-factor correction

MDF_{HV} and MDF_{HH} measurements were only collected to calculate the G-factor ($G=I_{HV}/I_{HH}$). G-factor was calculated with the raw[60] polarized MDF_{HV} and MDF_{HH} spectra,[m] which was then denoised[n] using a one component PARAFAC model.[o] These G-factors without noise were used to correct the raw MDF_{VH} spectra into the perpendicular $MDF_{\perp}$ spectra (Figure 2.2).[91] The MDF_{VV} spectra did not require any

[m] For the EEM and TSFS rIgG native state data, G-factor was calculated for the full emission space (Figure 2.2). However, for the EEM thermally stressed rIgG, the G-factor was calculated from the instrumental factor corrected EEM_{HV} and EEM_{HH}, with λ_{em}=302−450 nm, and used to correct the instrumental factor corrected EEM_{VH}.

[n] The use of a Savitzky-Golay smoothing function was not able to adequately remove noise from the MDF_{HV} and MDF_{HH} spectra prior to G-factor calculation. An alternative would be to use the average of multiple pMDF spectra to remove noise, however, the time necessary to collect the data made this approach prohibitive. Thus, the best alternative was to use PARAFAC modelling to denoise the calculated G-factor.

[o] TSFS G-factor was arranged into t-EEM to allow the use of PARAFAC for denoising purposes. After that, the denoised t-EEM G-factor was re-transformed into TSFS.

G-factor correction and are known as the parallel MDF∥ spectra. Normal unpolarized MDF spectra were not collected, but calculated from the polarized spectra and defined as total unpolarized MDF_T spectra ($MDF_T=MDF\parallel+2\times MDF\perp$).[60]

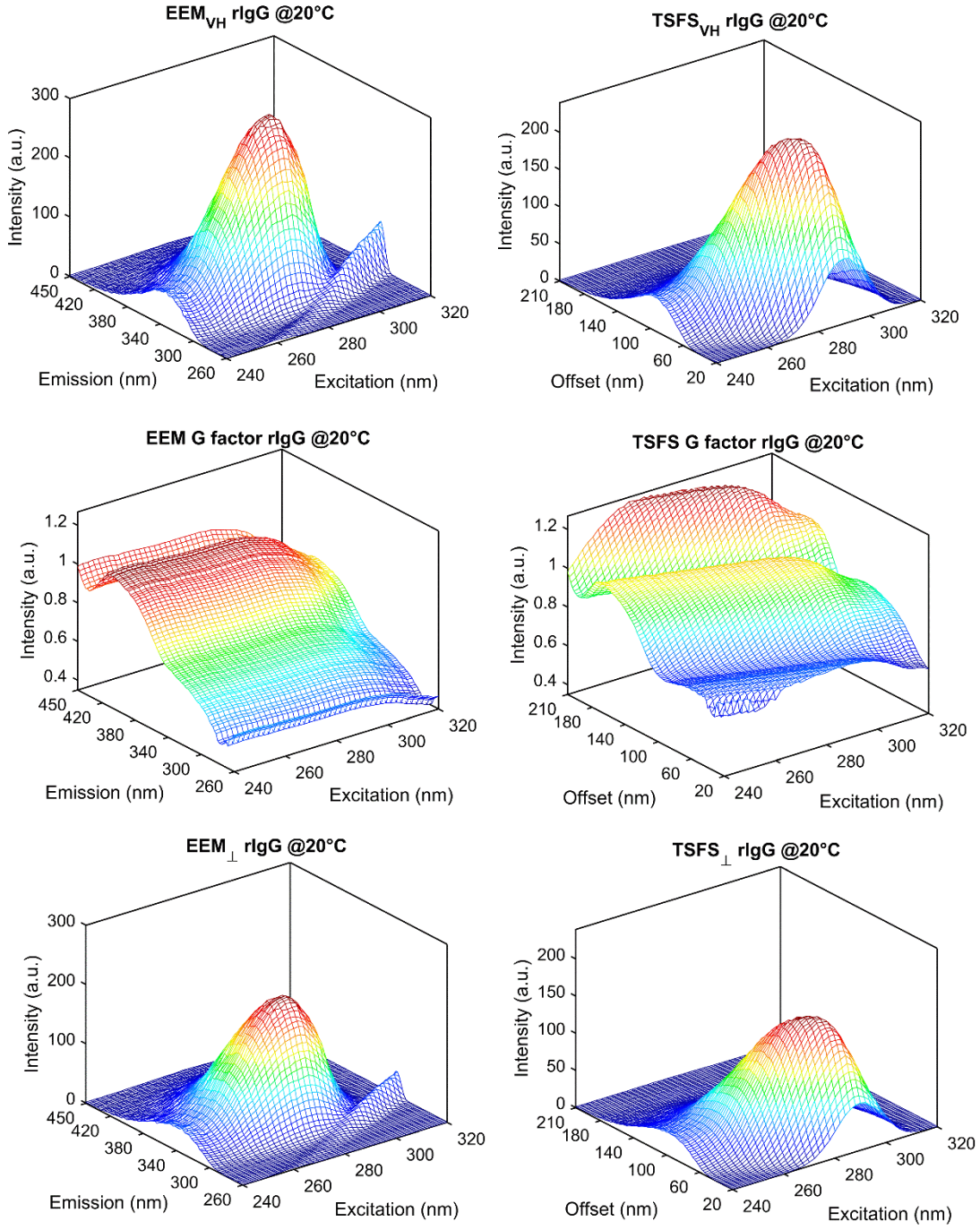

Figure 2.2: (Left) Raw EEM_{VH} spectra of rIgG at 20°C, G-factor correction plots for EEM and the resulting G-factor corrected perpendicular polarization, $pEEM_{\perp}$. (Right) Raw $TSFS_{VH}$ spectra of rIgG at 20°C, G-factor correction plots for TSFS spectra, and the resulting G-factor corrected perpendicular polarization, $pTSFS_{\perp}$.

The TSFS G-factor was converted into a t-EEM layout to enable a comparison between EEM and TSFS G-factors. Figure 2.3 shows the G-factor calculated for the

EEM (red spectra) and TSFS, either calculated as TSFS data and subsequently converted into t-EEM (black spectra), or first converted into t-EEM before the G-factor was calculated (dashed blue spectra). There are some variations at λ_{em}<340 nm, which were caused by structural differences between EEM and TSFS data.

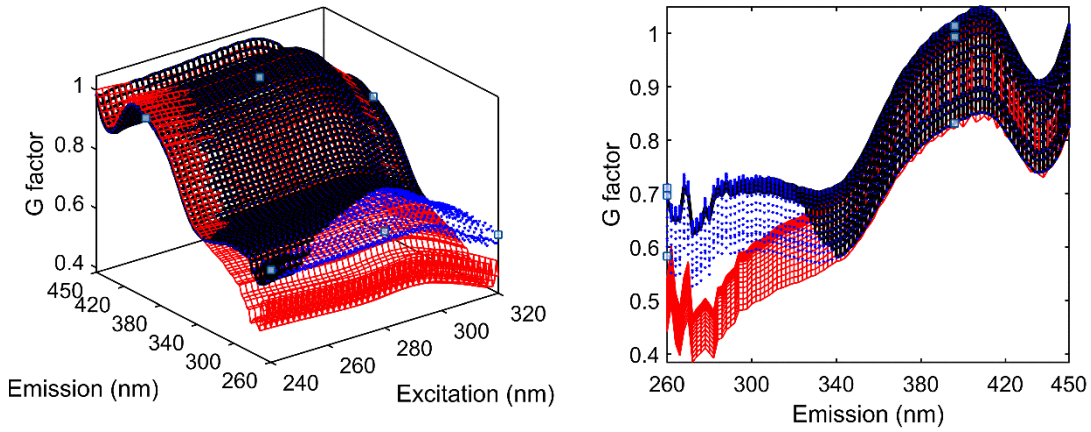

Figure 2.3: Comparison of the EEM (red) and t-EEM (black) G-factor for the rIgG measured at 20ºC calculated with the raw EEM/TSFS$_{HV}$, and EEM/TSFS$_{HH}$ datasets. Alternatively, t-EEM G-factor was calculated with the t-EEM$_{HV}$ and t-EEM$_{HH}$ (dashed blue spectrum) to have the full emission space G-factor.

2.4.3 Inner filter effect (IFE) correction

The use of 2 mm excitation pathlength cuvettes reduced IFE,[94] however, there is still some IFE occurring due to the high optical density at 280 nm (A_{280}=0.32±0.05, 2 mm pathlength) of the ~1.0 mg mL^{-1} rIgG solutions (Table 2.1). IFE is much more significant in the analysis of pMDF data because of the reduced fluorescence intensity derived from the use of polarizers.[1] The IFE correction was performed according to the absorbance-based approach (ABA).[22] The ABA method uses the measured absorbance (A_λ) at each pair of excitation (λ_{ex}) and emission (λ_{em}) wavelengths to convert the observed fluorescence intensity (F^{obs}) into the corrected fluorescence intensity (F^{corr}), as Equation 12.

$$F^{corr}_{\lambda ex,\lambda em} = F^{obs}_{\lambda ex,\lambda em} \times 10^{\left(\frac{A_{\lambda ex}+A_{\lambda em}}{2}\right)} \text{ ; if } F^{obs}_{\lambda ex,\lambda em} > LOR\ F_{\lambda em,\lambda ex} \qquad \textbf{Equation 12}$$

$$LOR\ F_{\lambda em,\lambda ex} = F_{blank(\lambda em,\lambda ex)} + 10 \times SD(F_{blank(\lambda em,\lambda ex)}) \qquad \textbf{Equation 13}$$

IFE correction was limited to the spectral coordinates which had fluorescence intensities that were above the limit of reporting (LOR, Equation 13).[50] LOR is the minimum amount of analyte that could be quantified. The LOR at each $\lambda_{em}/\lambda_{ex}$ coordinate was calculated for each polarized measurement from the standard deviation

of 10 blank (PBS buffer) replicate measurements for each polarization mode for the EEM and TSFS techniques (Figure A-2).

2.4.4 Multi-way decomposition

The first challenge was to select a single multi-way decomposition method that could be applied to both pEEM and pTSFS data. TSFS structure is intrinsically different from EEM and requires proper understanding before subjecting them to chemometric methods.[71] For EEM, the shape of each fluorophore in one spectral mode should be invariant to changes in another spectral mode[p] (Figure 2.4C−F), making the EEM data trilinear. However, it varies for TSFS (Figure 2.4A/B), which behaves as a non-bilinear matrix.[q] TSFS lack of trilinearity can be easily solved[73] by representing the TSFS data into an EEM like layout (t-EEM, where $\lambda_{em}=\lambda_{ex}+\Delta\lambda$).[71, 74]

EEM three-way arrays can be decomposed using either bilinear or trilinear methods such as MCR[104] or PARAFAC,[85] respectively. PARAFAC is one of the most popular curve resolution methods[79] that requires the three-way array data to be trilinear in nature. Ideal EEM three-way array data generally fulfil this trilinearity requirement, however, for real-world data, this could be affected by Rayleigh and Raman scatter, IFE, and interaction between fluorophores.[83] Non-trilinearity caused by Rayleigh and Raman scatter, and IFE can be corrected, however, deviations generated by interacting fluorophores cannot be corrected in complex proteins[r] (Figure 2.4E). MCR has more flexibility in terms of modelling, and offers the advantage to switch between bilinear, partial trilinear, and fully trilinear models.[88] However, bilinear methods could have rotational ambiguities that are intrinsically associated with MCR resolution.[71, 74, 81, 82, 104, 105] In addition, the use of different constraints and augmentation modes can yield different MCR-ALS solutions from the correct solution.[82, 88]

[p] In an ideal case, where the fluorophores do not interact with each other,[68, 85] for the EEM, every slice of the spectra has a different intensity, but the same shape. For the complex rIgG proteins, where multiple fluorophores interact, the shape of the emission spectrum changes (Figure 2.4E) when we excite both Tyr and Trp fluorophores ($\lambda_{ex}\leq 290$ nm), or selectively excite only Trp fluorophores ($\lambda_{ex}>290$ nm), breaking the trilinearity and variability required for PARAFAC analysis.

[q] For the TSFS, every slice of the spectra has a different intensity and shape (Figure 2.4A/B).

[r] In addition, the fluorophores interacting in a FRET process break the variability requirement, making it very difficult, or even impossible, to resolve individual fluorophores with PARAFAC analysis.

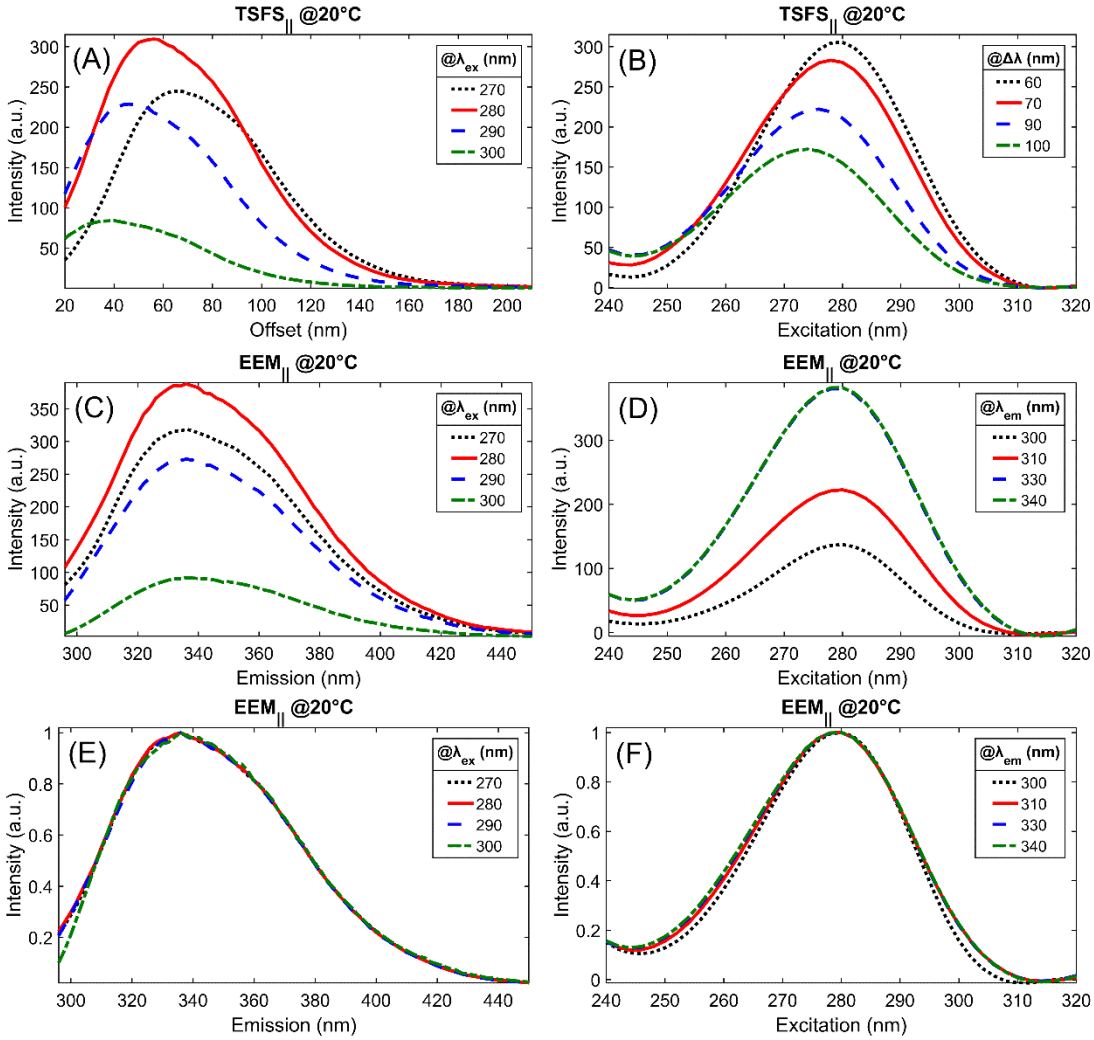

Figure 2.4: Blank subtracted emission spectra of rIgG at 20°C at different λ_{ex} (270, 280, 290 and 300 nm) collected with (A) pTSFS‖, and (C) pEEM‖. Excitation spectra collected at different (B) $\Delta\lambda$ (60, 70, 90, and 100 nm), and (D) λ_{em} (300, 310, 330, and 340 nm). (A/B) For the pTSFS‖ data, every slice of the spectra has a different shape and intensity. (C/D) For the pEEM‖ data, every slice of the spectra has a different intensity. (E/F) Normalized pEEM‖ shows that the shape remains intact for λ_{ex}<300 nm (Tyr + Trp emission spectra) and λ_{em}>300 nm (Tyr + Trp excitation spectra). Emission spectra show a change in shape for λ_{ex}>290 nm (Trp emission spectrum).

Three-way TSFS data could be directly decomposed by bilinear methods like multivariate curve resolution alternating least square (MCR-ALS),[71, 74, 81, 82, 88, 104, 105] allowing the resolution of each fluorophore, if proper data augmentation is performed and suitable designed sample sets are used.[106, 107] However, correct MCR-ALS decomposition of three-way TSFS data is also dependent on the relative intensity, and degree of spectral overlap between fluorophores.[106] If the intrinsic protein emission has a significant overlap between fluorophores, as it is the case for rIgG, the only way of applying MCR-ALS decomposition is if the experimental design introduces variability that could break these dependencies.[74, 104] Generally, this would not be suitable or practical and the trilinearity required for PARAFAC might help when analysing the highly overlapped emission of rIgG. Consequently, TSFS data must be

restructured into t-EEM to generate the trilinearity required for PARAFAC analysis, which may resolve Tyr/Trp spectral overlap in a better way than TSFS/MCR.[75] Also, because TSFS measurements minimize the collection of Rayleigh scattered light,[71, 108] it is less likely that residual Rayleigh scatter[s] would affect data trilinearity once this data is transformed into t-EEM,[73] which could improve PARAFAC model in comparison with the EEM data. Thus, PARAFAC models of TSFS data can be directly compared with those obtained of similar EEM data, to unambiguously determine which measurement method is more suitable for IgG analysis.

Here, we selected PARAFAC as our multi-way decomposition method because of the uniqueness of the solutions in comparison with MCR. However, to guarantee that PARAFAC solutions are unique, pEEM/TSFS data must be pre-processed to remove or, at least, to reduce any deviations from trilinearity.[84, 85, 107]

2.4.5 Data pre-processing for PARAFAC analysis

The best pre-processing methods are those which accurately remove interferent signals (Rayleigh and Raman scatter), correct for systematic bias (instrumental factor, G-factor, and IFE), and leave the analyte signal intact, leading to a trustworthy answer. The number, sequence, and steps used to pre-process a dataset depend on the data itself, and it could be critical for PARAFAC analysis.[84, 85, 109, 110] Data pre-processing has to be implemented with caution and good knowledge of the effects of each step. Improper pre-processing results in spectral distortions, yielding incorrect data. Chemometric analysis of this incorrect data will lead to misleading results. Thus, the pre-processing steps had to be adapted to the pEEM and pTSFS rIgG datasets (Figure 2.5), and multiple methods and combinations were evaluated (Figures 3.1/4.1/5.1) for the specific issues associated with rIgG emission.[91, 92] Data pre-processing was selected to optimize PARAFAC resolution, more details about this selection will be provided in the next chapters.

[s] Depending on the offset, slit width, and the amount of residual scattered light, there can be appreciable amounts of residual scattered light.

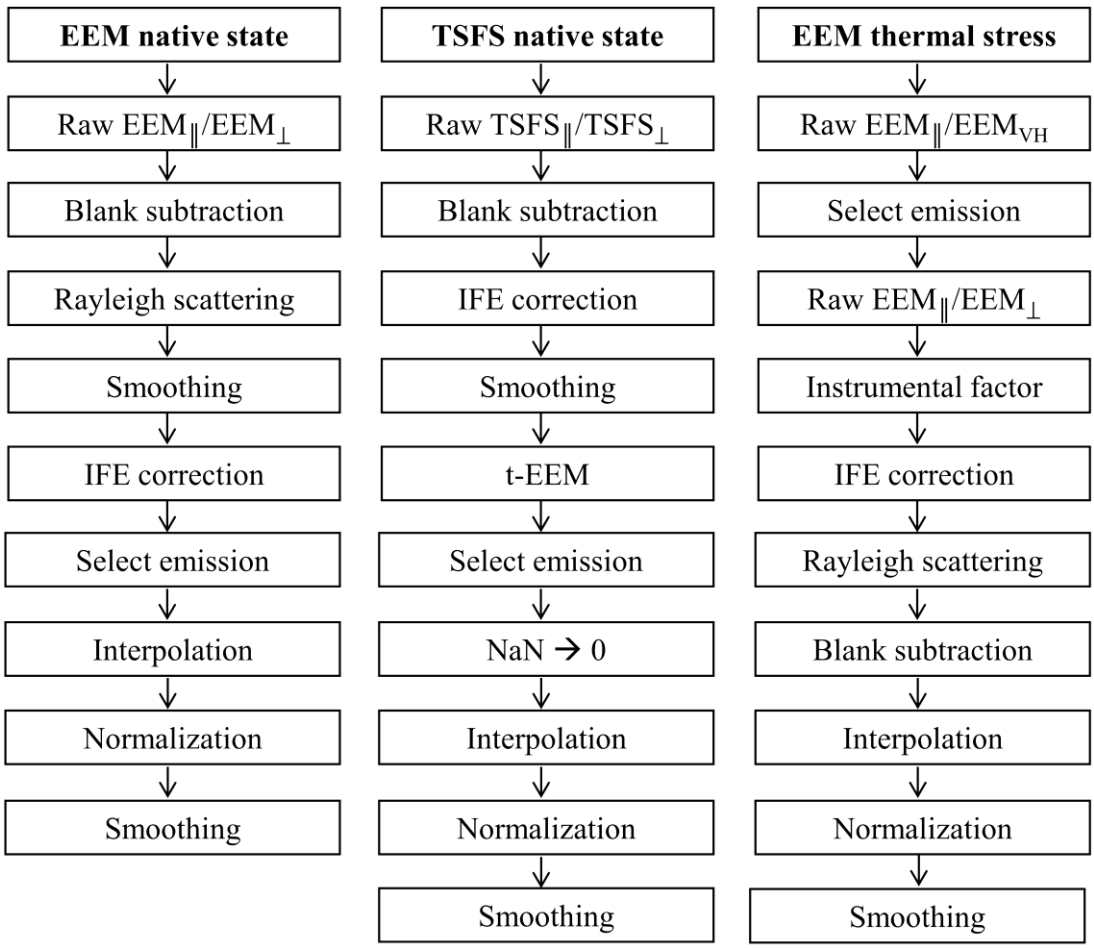

Figure 2.5: Schematic of the selected pre-processing steps applied to (left) pEEM rIgG native state, (middle) pTSFS rIgG native state, and (right) pEEM thermally stressed rIgG. Pre-processing steps for the thermally stressed rIgG had to be modified to optimize instrumental factor correction. Selected emission range for the rIgG native state was 296−450 nm, and 302−450 nm for the thermally stressed rIgG. NaN = missing data, 0 = filled with zeros.

Prior to data analysis, pEEM and pTSFS datasets were pre-processed to reduce IFE, and minimize the effects of Rayleigh and/or Raman scattering. While mathematical procedures had to be applied to remove Rayleigh scatter from the pEEM data,[57, 109-111] Rayleigh scatter is minimized during TSFS measurements by appropriate wavelength offset selection (>10 nm).[52, 54] The use of $\Delta\lambda \geq 20$ nm reduced the collection of Rayleigh scatter, and Raman bands were the main source of scattered light.[22] However, it is not possible to ensure that the Rayleigh scatter was completely removed during pTSFS measurement, and some light scatter will be present (Chapter 4) due to the use of quite wide slit widths (10 nm)[t] coupled with short Stokes shifted emission for λ_{ex}<300 nm. If the sample solution contains particles, then this will

[t] Reducing the slit widths improves the efficiency at Rayleigh scatter removal. However, the use of a Cary Eclipse combined with polarizers meant that we could not reduce the slit widths used here, as it would reduce fluorescence intensity, degrading SNR. Thus, one would have to slow scan speed, and increase PMT voltage to compensate.

increase further due to Mie and Tyndall scattering. Some of the common pre-processing steps (Figure 2.5) were interpolation and Savitzky-Golay smoothing using a second-order polynomial with a 15-point window size[112] to reduce unwanted noise. While smoothing here facilitates PARAFAC analysis, a better approach would have been to average spectra, and use more samples.[u]

The pre-processed pMDF spectra were used to calculate the corresponding anisotropy (r) at each $\lambda_{ex}/\lambda_{em}$ coordinate (Equation 14).[60] This was then used to generate a multidimensional data matrix ($\lambda_{ex} \times \lambda_{em} \times r$) over the full emission space (*aniso*-MDF map). *Aniso*-MDF maps are inherently noisier than the source MDF data, and a 10% threshold had to be used to reduce noise influence. In addition, *aniso*-MDF maps have intrinsically more errors than the MDF measurement, because it is calculated from multiple data (MDF_{VV}, MDF_{VH}, MDF_{HV}, and MDF_{HH}).[2, 3, 57, 88]

$$\bar{r}_{(\lambda ex,\lambda em)} = \frac{MDF_{\parallel(\lambda ex,\lambda em)} - MDF_{\perp(\lambda ex,\lambda em)}}{MDF_{\parallel(\lambda ex,\lambda em)} + 2 \times MDF_{\perp(\lambda ex,\lambda em)}} \quad \textbf{Equation 14}$$

The final pre-processing step before PARAFAC analysis was normalization (by peak maximum),[83] to remove variances related with small concentration differences between the replicate samples, and day-to-day measurements. The normalization step focused PARAFAC resolution on real spectral changes, instead of the simple changes in signal magnitude, facilitating the resolution of the weaker fluorophore contributions.

For PARAFAC analysis, the number of components was selected based on several criteria: the CORe CONsistency DIAgnostic test (CONCORDIA, Equation 15),[83, 113] on how much of variance was explained by the model, and visual inspection of the recovered spectral profiles and residuals. Validation of spectral deconvolution results were performed using split-half analysis.[83-85] PARAFAC analysis was applied with non-negative constraints for all modes (sample, λ_{em}, λ_{ex}) using the best-fitting model as initialization method (from various test models fitted with a small number of iterations).[85, 110]

$$CONCORDIA = 100 \times \left(1 - \frac{\sum_{d=1}^{F}\sum_{e=1}^{F}\sum_{f=1}^{F}(g_{def} - t_{def})^2}{F}\right) \quad \textbf{Equation 15}$$

[u] The collection of several spectra from the same sample at each data point was not feasible due to the time required to collect each 3D pMDF spectra.

First, the pMDF$_{\parallel}$ and pMDF$_{\perp}$ spectra were analysed using PARAFAC to see if different components could be resolved, and if there were differences in the recovered loadings (spectra) for the different polarizations. From these initial PARAFAC analyses of the different polarizations, an initial model for protein structural characterization was generated. Second, we validated the characterization model by calculating the anisotropy, and *aniso*-MDF plots assessing how each emitter contributed to the characterization.

The total unpolarized MDF$_T$ spectra was analysed with two main purposes: *(i)* to compare the information extracted from the polarized MDF measurements with the polarization independent MDF$_T$,[22, 38, 60, 91] showing if the polarized measurements were advantageous over the unpolarized ones, and which was the best polarization. *(ii)* To serve as control for chemometric modelling,[22, 38, 60] because the anisotropy maps were calculated based on the total intensity, and not on each individual polarization.

3. Rabbit IgG Native State Intrinsic Fluorescence Characterization Using ARMES – EEM

This chapter is divided into two main sections. The first section shows how the pEEM data had to be pre-processed for accurate photophysical characterization by chemometrics of the rIgG in its native state (second section). The results presented here were published in *Chemometrics and Intelligent Laboratory Systems*.[91]

3.1 Data pre-processing

The pEEM data must be pre-processed to ensure the trilinearity required for PARAFAC analysis. The selection of the steps required were based on the knowledge about what was affecting pEEM trilinearity.[114] While we could not correct for the non-trilinearity caused by FRET, it was possible to remove (or at least reduce) Rayleigh and Raman scatter, and IFE. Thus, to select the best pre-processing method for accurate PARAFAC analysis, we assessed several methods with different sequences of operations for the rIgG collected with pEEM. Additionally, we had to assess the use of interpolation to guarantee that it would not induce any spectral distortion.[91] After evaluating several pre-processing methods, we selected four that seem to solve most of the problems associated with the pEEM data (Figure 3.1).

While Raman scattering was easily removed by a buffer subtraction (Figures 3.2C/3.3B), there is no perfect method for removing Rayleigh scatter, and many methods are described in the literature.[57, 109-111, 115] Rayleigh scatter contamination is greatest at the emission blue edge (Figure 3.2A), requiring certain expertise for pre-processing it in a proper manner. The complete removal of the Rayleigh scatter signal is not trivial as one could unintentionally remove/distort the fluorescence signal at the emission blue edge. This will have a particularly adverse effect on the Tyr emission signal.[v] The first order Rayleigh scatter was modelled as a separate factor by PARAFAC and corrected (Figure A-3) as previously described.[57, 91, 111] Due to the spectral range sampled, there was no second order Rayleigh scatter in the pMDF datasets.

[v] Tyr signal is weak and could be heavily overlapped with Rayleigh scattered light. It could be very difficult to adequately separate Tyr and Rayleigh scatter spectral overlap.

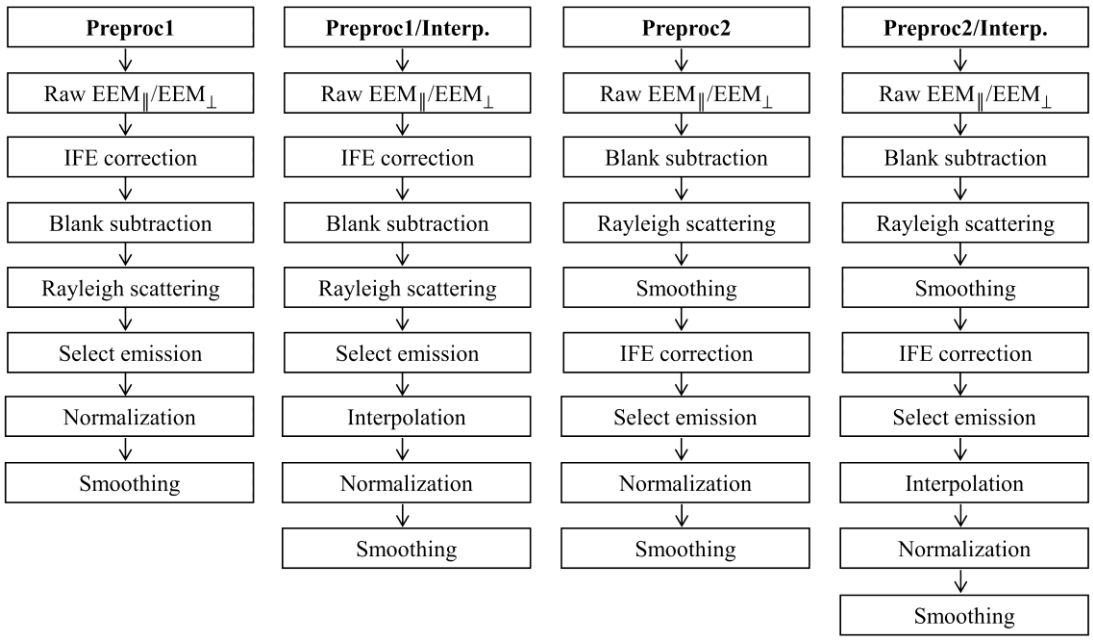

Figure 3.1: Schematic of the different sequences of pre-processing steps applied to the pEEM data.

A smoothing function was used after Rayleigh scatter removal for the data corrected with Pre-processing 2 and Pre-processing 2/Interpolated (Figure 3.3C), but not for Pre-processing 1 and Pre-processing 1/Interpolated (Figure 3.2D). This can be explained by the order of steps used for the different pre-processing methods. When a multiplication-based correction (*e.g.* IFE correction) is the first step, there is an increase in Rayleigh signal (Figure 3.2B), and noise is multiplied, distorting the data.[w] On the other hand, if subtraction-based corrections (*e.g.* buffer subtraction, Rayleigh scatter removal) are at the beginning of the process, the smoothing function can be used to decrease unwanted noise that could be increased by the multiplication-based correction. Thus, the order in which subtraction-, and multiplication-based correction were implemented determined the use or not of a smoothing function after Rayleigh scatter removal (Figure 3.1).

[w] The use of a smoothing function would distort the data even more, affecting PARAFAC model quality.

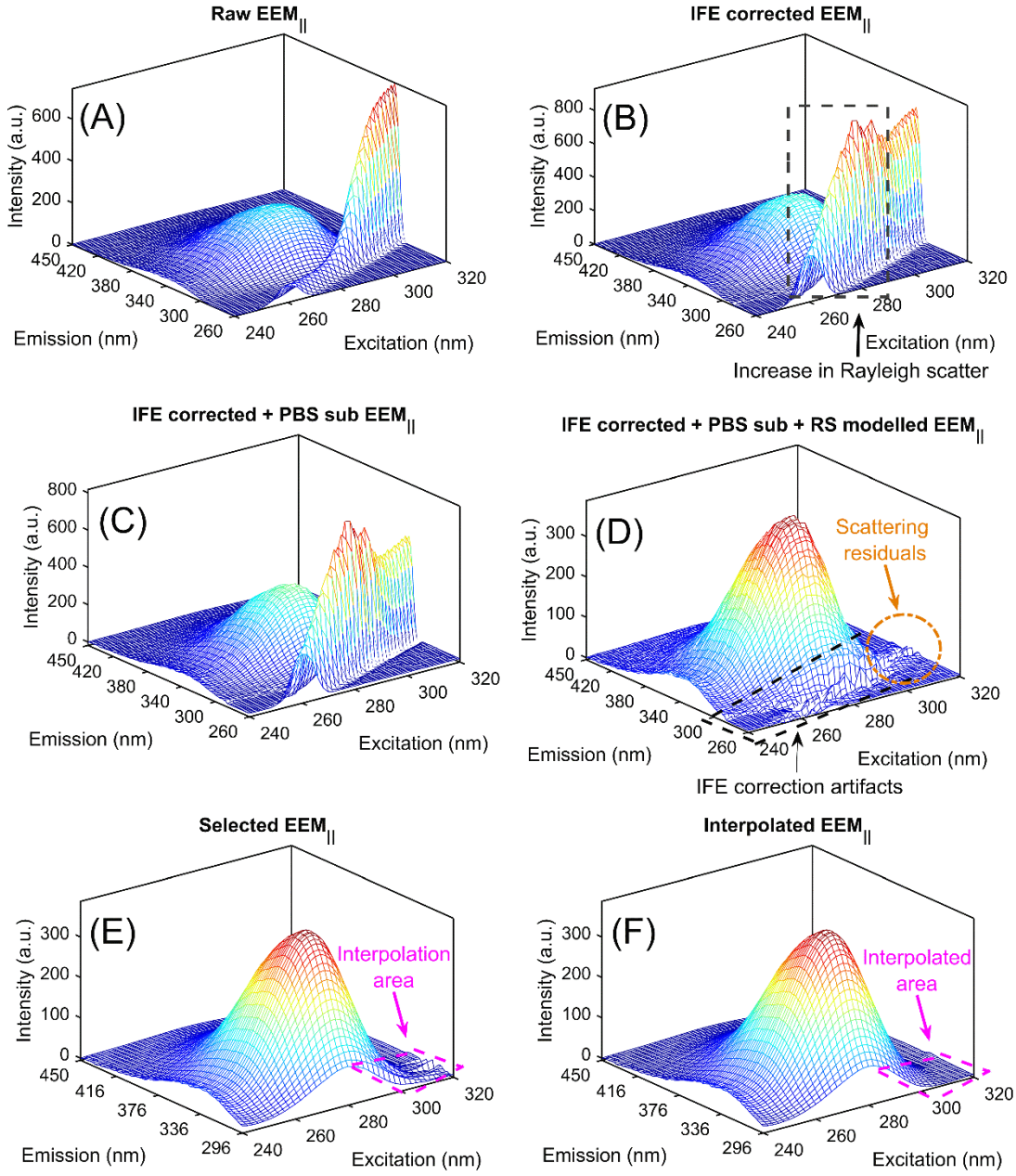

Figure 3.2: Pre-processing 1/Interpolated method. (A) Raw pEEM‖ spectra of rIgG at 20°C. (B) pEEM‖ spectra IFE corrected (IFEc). (C) pEEM‖ IFEc and buffer subtracted. (D) pEEM‖ spectra IFEc, buffer subtracted, and Rayleigh scatter removed. (E) pEEM‖ spectra were cut at λ_{em} 296−450 nm to remove most of the noisy area produced by IFE correction. (F) The pEEM‖ spectra were interpolated and smoothed.

Multiplication-based IFE correction amplified noise (mainly related with residual Rayleigh scatter) present in the pEEM spectra, generating some artifacts at λ_{em}<292 nm region, where emission was weak (Figures 3.2D/3.3D). The spectral region containing these artifacts were thus removed, creating a new dataset, with $\lambda_{em}\geq$296 nm (Figures 3.2E/3.3E). The removed area was mostly related to weak Tyr emission and Trp emission at the blue edge.[3, 22] An interpolation step[109, 110] was necessary to improve residual shot noise removal in the Rayleigh scattering region (Figures

3.2F/3.3F). The pre-processed pEEM∥, pEEM⊥, and unpolarized EEM_T yielded a 3D structure (X) of size 27 samples × 78 λ_{em} × 41 λ_{ex}, which was then normalized by peak maximum and smoothed for PARAFAC analysis. Neither the use of IFE correction nor normalization caused any significant change in PARAFAC resolution (Table A-2) for the pEEM data.

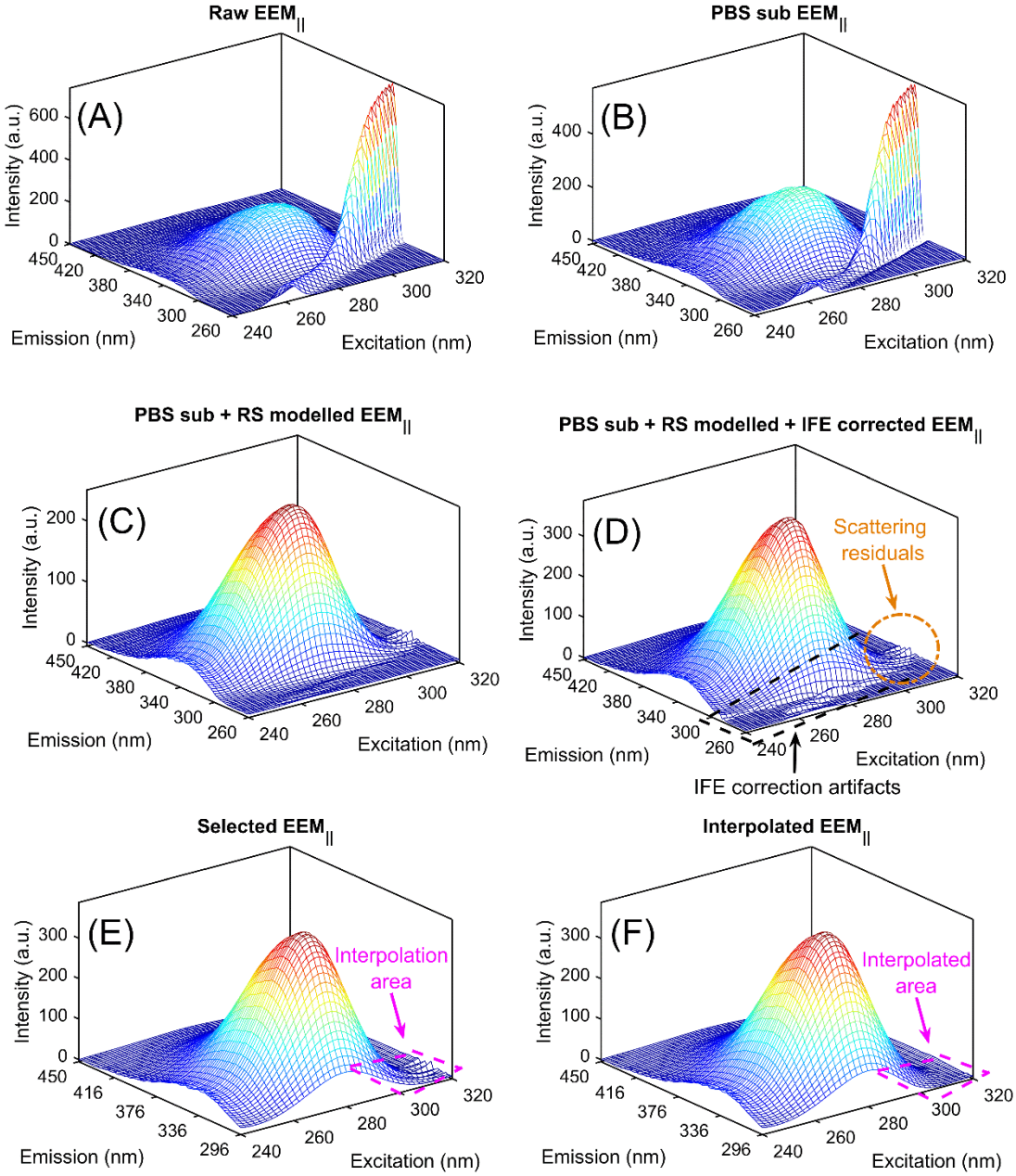

Figure 3.3: Pre-processing 2/Interpolated method. (A) Raw pEEM∥ spectra of rIgG at 20°C. (B) pEEM∥ buffer subtracted. (C) pEEM∥ buffer subtracted, and Rayleigh scatter removed. (D) pEEM∥ spectra IFEc. (E) pEEM∥ spectra were cut at λ_{em} 296−450 nm to remove most of the noisy area produced by IFE correction. (F) The pEEM∥ spectra were interpolated and smoothed.

Polarized EEM∥, EEM⊥, and un-polarized EEM_T rIgG spectral data that underwent the different pre-processing methods were then analysed with PARAFAC.

PARAFAC models for the different polarizations and methods were evaluated to check if: *(i)* the use of different pre-processing methods impacted PARAFAC resolution; *(ii)* to ensure that the use of interpolation did not distort the data; *(iii)* to check if different polarizations resulted in different spectral loadings recovered with PARAFAC; and *(iv)* to see if different fluorophores could be resolved for the rIgG native state.

This was the first time that PARAFAC analysis was used to model such complex proteins,[x] and not like the simpler proteins case,[88] there was a limited knowledge about the exact components[y] which might be resolved. For the complex rIgG structure there was the empirical knowledge of the peak positions for pure Tyr and Trp emissions,[22] and the different classes of Trp within protein structure.[34] However, the expected peak position for pure Tyr and Trp in solution (Figure 1.2), and from different Trp classes more likely do not translate to those present in a complex protein.[z] Thus, this was a long and complicated iterative process to prudently select not only the best pre-processing method for PARAFAC analysis of IgG type proteins, but to understand the recovered components.

3.2 2D spectral analysis

The first step before implementing complex chemometrics on 3D data was to use simpler 2D emission plot to evaluate if the thermal fluctuation used in the experiment was enough to cause structural variations that could be observed as changes in Tyr and Trp emission, and Tyr-to-Trp FRET.[22, 49] To analyse possible changes in Tyr-to-Trp FRET, we subtracted the emission spectra at λ_{ex}=270 nm (excited both Tyr and Trp) from that at λ_{ex}=296 nm (selectively excited Trp). This difference spectrum represents the emission of the Tyr directly excited, which is the fraction that does not come from FRET phenomena with the numerous Trp acceptors.[48, 116, 117] In addition, spectral

[x] The presence of multiple Try and Trp fluorophores makes it hard, or even impossible, to extract individual fluorophores.[34, 47, 65]

[y] Unless there is a significant structural change, it is likely that only an average of all fluorophores presents within protein structure will be resolved. This average could comprise all Tyr and Trp fluorophores together, meaning that only one component would be recovered, or as two separate groups of fluorophores, resolved as two individual PARAFAC components.

[z] The wider, and different emission ranges for the Tyr/Trp mixtures compared to IgG is due to the lack of FRET in the small molecule mixtures, and the inability to replicate the physicochemical environments of the fluorophores within the protein. In addition, the Trp classification was built based on the direct excitation of Trp (λ_{ex}=295 nm),[35] not taking the Tyr/Trp interactions in consideration.

changes observed in the directly excited Tyr with an increase in temperature indirectly reflects a change in Tyr-to-Trp FRET.

Figure 3.4 shows some evidence for weak Tyr fluorescence emission, which is red shifted (λ_{em} ~320 nm) in the unpolarized EEM_T. The Tyr emission was weak (~5%, of the maximum intensity) because it was overlapped with Trp emission and was probably distorted by the Rayleigh scatter removal procedure. The increase in temperature generated some small changes in Tyr emission, which was reflected in a relative standard deviation (RSD) of <6.1% for Pre-processing 1, and <5.7% for Pre-processing 2. These changes were small enough to indicate that there was no significant protein structural change in the 15−35°C temperature range, and the rIgG was in its native state for the temperature range sampled. Thus, any small variations in fluorescence emission was induced by a simple structural fluctuation rather than a significant structural change.

Small changes in the overlaid anisotropy emission plots (Figure 3.4) were limited to the emission blue edge (λ_{em}<320 nm), as observed with a RSD <3.8% at λ_{ex}=270 nm, and <7.4% at λ_{ex}=296 nm, which was caused by residual shot noise from Rayleigh scatter removal, and IFE correction. Based in the emission blue edge, Pre-processing 2 (Figure 3.4C) did a better job removing the artifacts created by IFE correction in comparison with Pre-processing 1 (Figure 3.4A), but neither pre-processing method were able to fully remove residual noise.[aa] The use of interpolation (Figure 3.4B/D) improved residual noise removal, without affecting the area outside the scattering zone. However, there were visible differences within the affected scattering zone (λ_{em}<320 nm), meaning that the anisotropy values within this region were suspect.[57, 88]

[aa] Causing problems for the accurate retrieval of anisotropy at the emission blue edge, which was observed as high anisotropy at λ_{em}<320 nm.

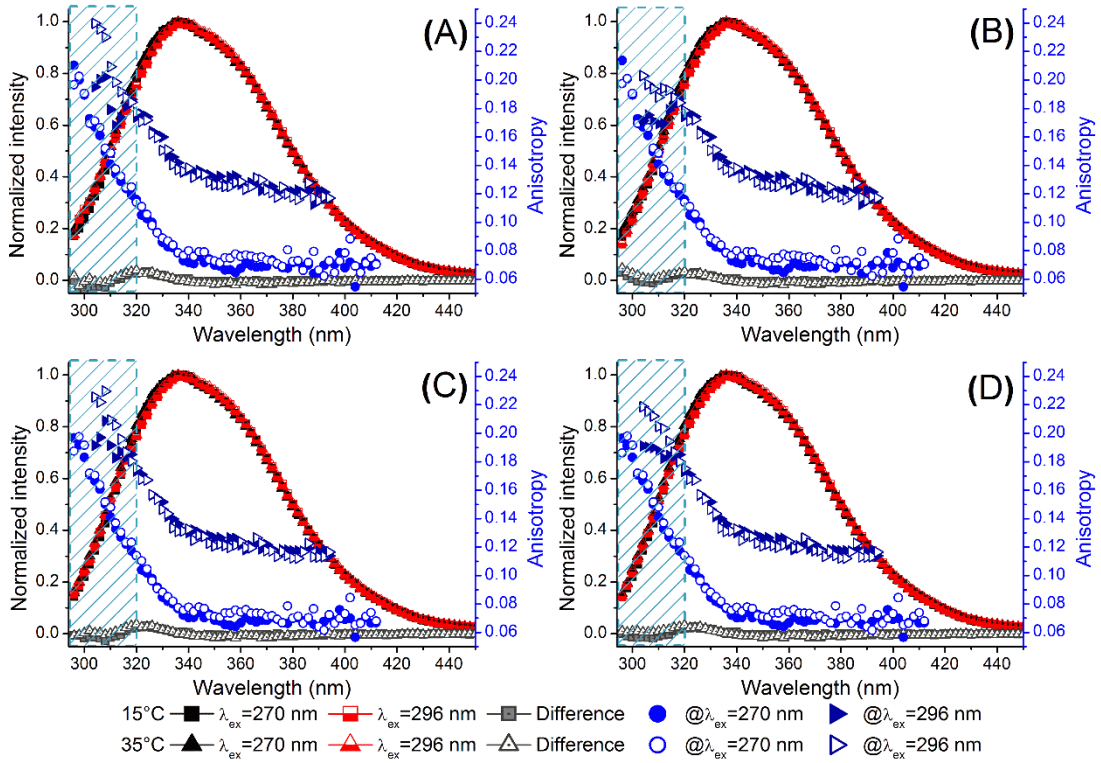

Figure 3.4: Normalized rIgG data corrected with (A) Pre-processing 1, (B) Pre-processing 1/Interpolated, (C) Pre-processing 2, and (D) Pre-processing 2/Interpolated EEM_T emission spectra excited at λ_{ex} 270, 296 nm, and the difference spectra (= λ_{ex} 270 − λ_{ex} 296) at 15°, and 35°C, overlaid with emission anisotropy at λ_{ex} 270 and 296 nm. The shaded boxes represent the emission spectra affected by residual noise.

A second 2D analysis based on Tyr ($\lambda_{ex}/\lambda_{em}$=276/306 nm) and Trp ($\lambda_{ex}/\lambda_{em}$=276/340 nm) emission[39, 118] was used here.[bb] Tyr and Trp fluorescence intensity (I_F) decreased linearly from 15° to 35°C for both pEEM (Figure 3.5A/B). The simultaneous decrease in Tyr and Trp I_F indicated thermally induced quenching, which was slightly faster for $pEEM_{\parallel}$ in comparison with $pEEM_{\perp}$. A decrease in Trp I_F would be expected in the case of changes in Tyr-to-Trp FRET.[39] Once the data were normalized (Figure 3.5C/D), it was clear that there was an increase in Tyr I_F, but there was no real trend observed for Trp. The linear tend observed for the Tyr/Trp ratio (Figure 3.5E) and relative FRET efficiency[39] (Figure 3.5F) confirmed the presence of a simple thermal quenching between 15−35°C.

[bb] Only the data obtained from the rIgG treated with Pre-processing 2/Interpolated is shown (Figure 3.5) as there was no significant variation between the results for the different pre-processing methods.

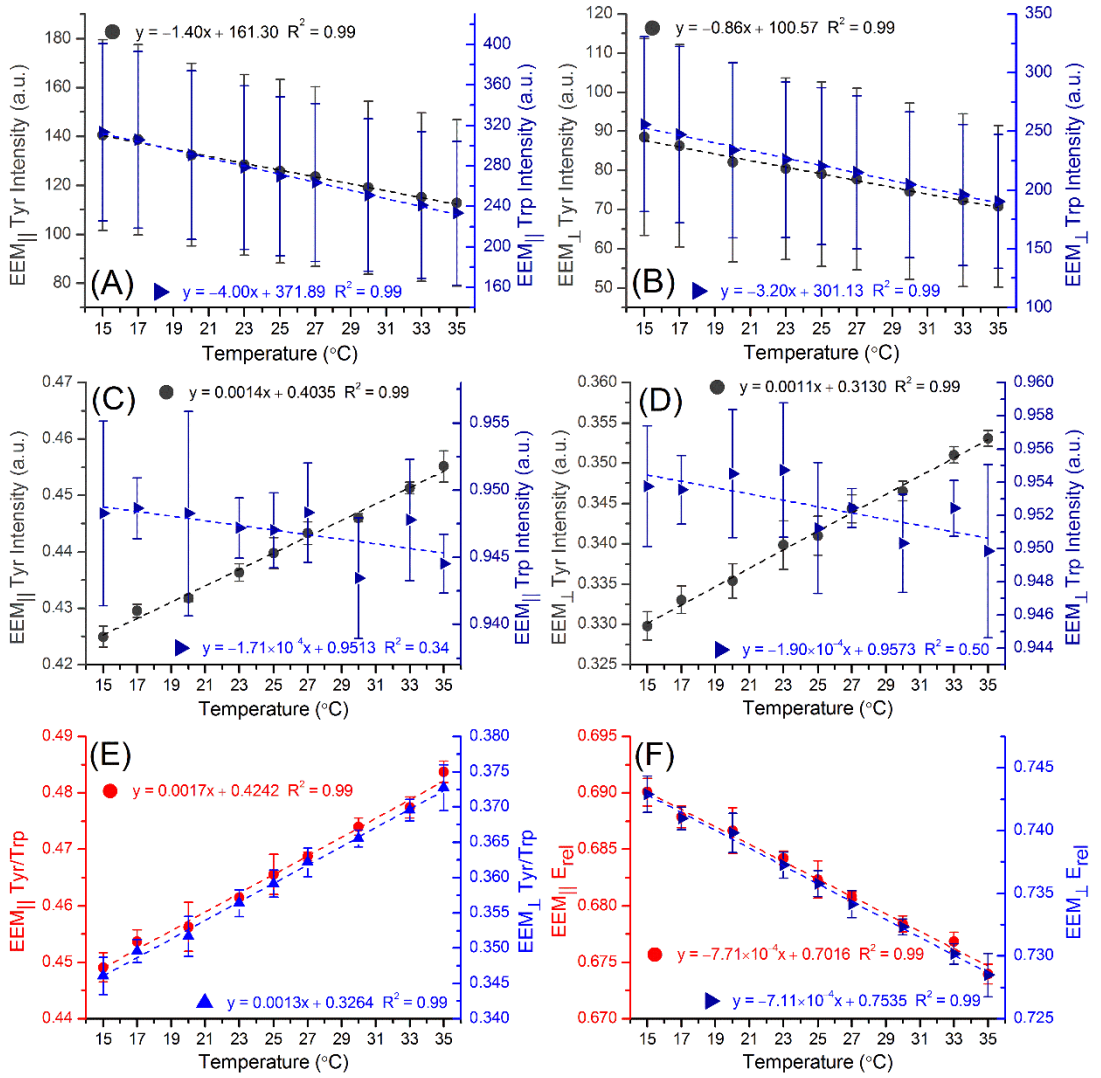

Figure 3.5: There was a decrease in Tyr ($\lambda_{ex}/\lambda_{em}$=276/306 nm) and Trp ($\lambda_{ex}/\lambda_{em}$=276/340 nm) intensities for the (A) pEEM∥, and (B) pEEM⊥ rIgG measured from 15−35°C, caused by thermal quenching. Normalized (C) pEEM∥, and (D) pEEM⊥ showed the effects of Tyr-to-Trp FRET. (E) The Tyr/Trp ratio increases from 15−35°C for pEEM∥ and pEEM⊥. (F) The relative FRET efficiency decreases from 15−35°C for pEEM∥ and pEEM⊥. pEEM data from Pre-processing 2/Interpolated. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

There was no real indication of changes in Tyr-to-Trp FRET process, as there were no significant changes in Tyr spectral emission (Figure 3.4), and the increase in Tyr intensity was not accompanied by a decrease in Trp intensity (Figure 3.5). It could indicate that the temperature range applied in this experiment was enough to cause protein fluctuation,[101] but protein tertiary structure remained intact.[cc]

[cc] It seems like the thermal fluctuation changed the distance between Tyr and Trp fluorophores,[101, 119, 120] without changing Trp environment.

3.3 *Aniso*-EEM maps

Relying on the use of single excitation wavelengths could possibly ignore very valuable information. The selection of excitation wavelength at 270 and 296 nm highlighted spectral changes at the emission blue edge, but it could overlook small changes at the red edge.[121] This is the motivation for studying the full emission space. The *aniso*-EEM maps for the rIgG were heterogeneous (Figure 3.6) with the variations across the full emission space being caused by numerous factors such as: the type and how many fluorophores, where the fluorophore is located within the protein, changes in fluorophore mobility and in intra-molecular FRET, differences in fluorophores environments, and residual shot noise.[1, 88]

Figure 3.6 shows a small difference at λ_{em}>336 nm in the *aniso*-EEM maps obtained from the different pre-processing methods at various temperatures. Even though the Trp emission region (λ_{em}~330−350 nm) was not affected by the different pre-processing methods, there was a clear difference at shorter emission wavelengths (λ_{em}<336 nm). The larger scatter contamination produces a greater degree of shot noise, which was distributed further across the pEEM spectrum. This stochastic noise was not fully removed by any of the correction procedures, and thus causes problems for the accurate retrieval of anisotropy, especially at the emission blue edge. *Aniso*-EEM maps had a more irregular contour for the data corrected with Pre-processing 1 in comparison to those corrected with Pre-processing 2. This was due to the residual shot noise from Rayleigh scatter being amplified in the data corrected using Pre-processing 1 (Figure 3.2B).[dd] Thus, Pre-processing 2 (Figure 3.3E) was more efficient removing residual scattered light, and did not increase IFE correction artifacts and residual shot noise as much as Pre-processing 1 (Figure 3.2E). One should implement the subtraction-based corrections (*e.g.* Raman and Rayleigh scatter corrections) before the multiplication-based correction (*e.g.* IFE corrections) otherwise one gets increased noise contributions.[ee]

[dd] The maximum anisotropy values (Figure 3.7A) were higher and more variable for Pre-processing 1 than Pre-processing 2, with an RSD of 8.90% and 3.88% (for triplicate measurements, from 15−35°C), respectively.

[ee] This issue is especially critical at short wavelength (<250 nm) where the fluorescence emission signal and the system response changes are weak.

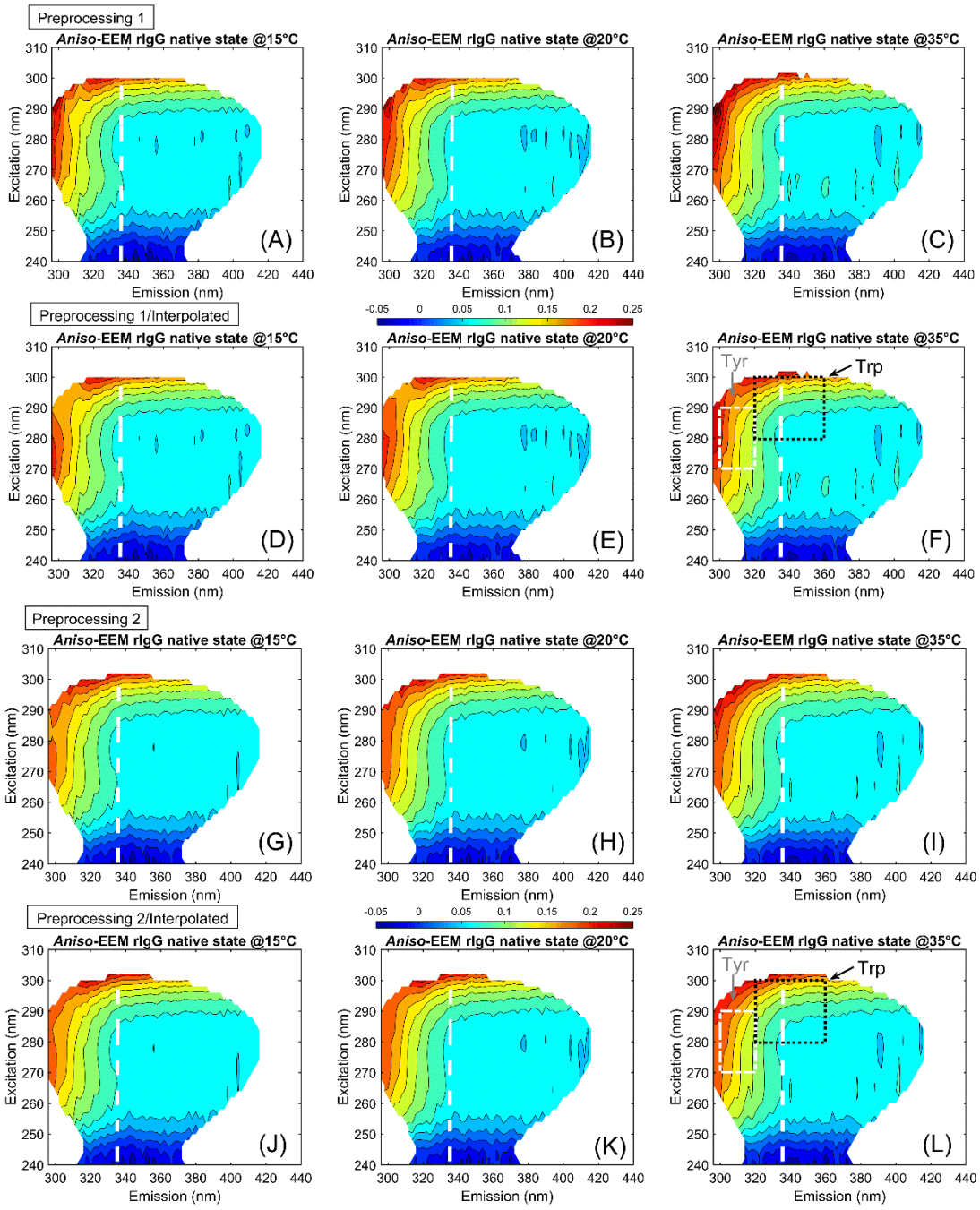

Figure 3.6: *Aniso*-EEM maps for rIgG measured at 15°, 20°, and 35°C, corrected with (A−C) Pre-processing 1, (D−F) Pre-processing 1/Interpolated, (G−I) Pre-processing 2, and (J−L) Pre-processing 2/Interpolated. The white dashed lines show λ_{em}=336 nm line. The colour bars in the middle represent the anisotropy scale.

The interpolation step was used to try to improve light scatter removal. Figure 3.7A shows a decrease in maximum anisotropy values (over the full *aniso*-EEM maps) from the non-interpolated to the interpolated pre-processed data. A significant decrease in maximum anisotropy from the non-interpolated Pre-processing 1 to Pre-

processing 1/Interpolated[ff] suggested that the interpolation step was efficient at removing residual scatter/noise. Pre-processing 2 was more efficient in removing Rayleigh scatter and noise as observed in the smaller differences in anisotropy maximum values between Pre-processing 2, and Pre-processing 2/Interpolated.[gg]

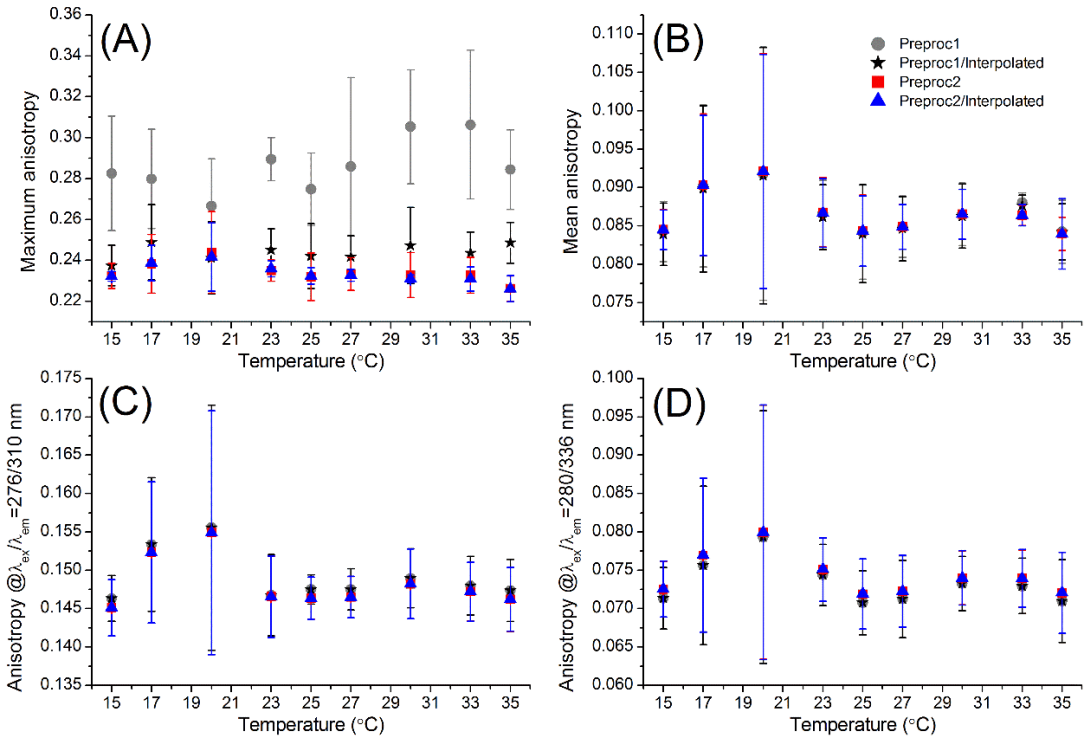

Figure 3.7: (A) Maximum anisotropy, (B) mean anisotropy, (C) Tyr anisotropy ($\lambda_{ex}/\lambda_{em}$=276/310 nm), and (D) Trp anisotropy ($\lambda_{ex}/\lambda_{em}$=280/336 nm) for each pre-processing method. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

Figures 3.6/3.7 clearly indicated that the interpolation step improved Rayleigh scatter and noise removal. However, care must be exercised here because proving that interpolation does not distort the true emission is not easy in FRET dominated systems like proteins.[hh] Thus, we must guarantee that whichever pre-processing method is selected that this is fixed during the analysis, and that the anisotropy values in these regions be understood to be estimates rather than accurate values.[65]

[ff] The inspection of the *aniso*-EEM maps for Pre-processing 1/Interpolated showed a reduction in contour irregularity and anisotropy values in the scatter region (RSD=4.90%) in comparison with the non-interpolated Pre-processing 1 (RSD=8.90%).

[gg] The differences between the *aniso*-EEM maps for the non-interpolated Pre-processing 2 (RSD=3.88%) and Pre-processing 2/Interpolated (RSD=3.15%) were less significant because there was less residual scatter/noise in Pre-processing 2.

[hh] As it was shown in the in initial development of ARMES,[57, 88] if we were working with an ideal case (non-interacting small molecules), without the presence of FRET, it would be possible to verify if the extracted spectra match that of the individual fluorophores or mixture of fluorophores.

There was no significant changes in mean anisotropy values (Figure 3.7B) over the full *aniso*-EEM maps for the different pre-processing methods.[ii] This indicated that overall, the degree of structural change was minimal, and the lack of differences in the mean anisotropy among the pre-processing methods could lead to a mistaken assumption that interpolation did not affect the data. However, anisotropy is intrinsically heterogeneous over the full emission space (Figure 3.6) and it is not correct to use maximum and/or mean anisotropy values to assess the degree of changes caused by each individual pre-processing method, requiring alternative assessment methods to quantify the degree of change. An alternative to remove the heterogeneity related to the full emission space was to select two single anisotropy points for the Tyr (at $\lambda_{ex}/\lambda_{em}$=276/310 nm) and Trp (at $\lambda_{ex}/\lambda_{em}$=280/336 nm) fluorophores (Figure 3.7C/D). Still, there were no significant changes in anisotropy values for the Tyr[jj] and Trp[kk] fluorophores for the different pre-processing methods.

To further assess the effects of the different pre-processing methods in the *aniso*-EEM maps a similarity index (SimI)[122] was calculated (*aniso*-EEM map from R1 at 15°C was used as the reference spectrum) for all the samples (Figure 3.8). The SimI plot calculated for the full emission space (Figure 3.8A) showed a better similarity between the *aniso*-EEM maps for Pre-processing 2 (RSD=22.06%) and Pre-processing 2/Interpolated (RSD=21.50%) in comparison with Pre-processing 1 and Pre-processing 1/Interpolated (RSD=35.71%). The reduced similarity for Pre-processing 1 and Pre-processing 1/Interpolated can be largely attributed to the residual shot noise from the scatter contamination.

[ii] Mean anisotropy values had an RSD of 7.67% for Pre-processing 1, 7.04% for Pre-processing 2, 7.79% for Pre-processing 1/Interpolated, and 7.03% for Pre-processing 2/Interpolated (for triplicate measurements at all temperatures).

[jj] There were no significant differences between the Tyr anisotropy values for Pre-processing 1 and Pre-processing 1/Interpolated (RSD=4.28%), and for Pre-processing 2 and Pre-processing 2/Interpolated (RSD=4.40%).

[kk] The Trp anisotropy values did not show any significant differences between Pre-processing 1 and Pre-processing 1/Interpolated (RSD=9.10%), Pre-processing 2 (RSD=8.88%), and Pre-processing 2/Interpolated (RSD=8.90%).

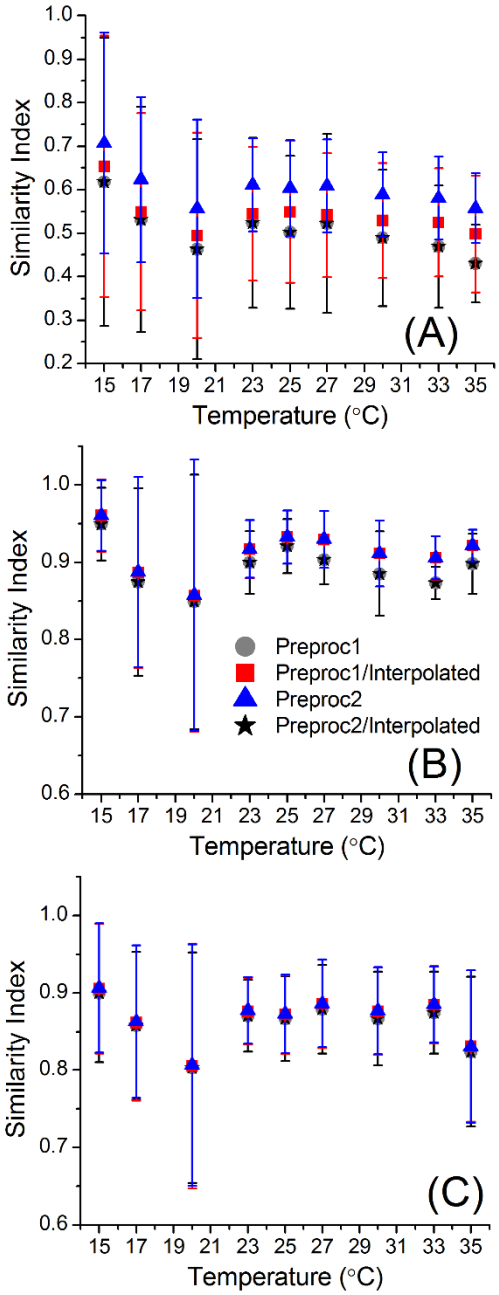

Figure 3.8: SimI calculated for (A) the full emission space, (B) Tyr region ($\lambda_{ex}/\lambda_{em}$=270─290/300─320 nm), and (C) Trp region ($\lambda_{ex}/\lambda_{em}$=280─300/320─360 nm) for each pre-processing method. SimI value =1 indicates identical maps. SimI values ≥0.9 could be considered to have no significant differences from the reference spectrum. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

The small dip at 20°C (Figure 3.8) is a genuine effect since the data were collected on three different days, with the same interval between defrosting and measuring, using the same measurement settings, with three different stock solutions prepared from a different source vial (but with the same lot number). This drop (and increase in error bars) was possibly caused by a change in local motion at 20°C, but at present we have no strong evidence to support this observation.[101, 102] Changes in protein structure flexibility are unlike to cause a significant change in the emission profile, but

it could cause quantifiable changes in anisotropy. A minimum SimI value was observed at 20°C, followed by a slightly increase in SimI at 23°C, remaining constant up to 35°C, indicating a stable structure. This effect requires additional investigation as the IgG used here was polyclonal in origin with only ~80% monomer and rabbit IgG composition is known to vary.[120]

To better evaluate the changes in Tyr and Trp emission in the *aniso*-EEM maps two areas were selected (Figure 3.6L): Tyr ($\lambda_{ex}/\lambda_{em}$=270−290/300−320 nm), and Trp ($\lambda_{ex}/\lambda_{em}$=280−300/320−360 nm). There were very small differences between the *aniso*-EEM maps selected for the Trp (RSD ~8.60%) and Tyr (RSD ~7.65%) regions. SimI analysis of the specific Tyr and Trp emission regions showed that there was no difference in the *aniso*-EEM maps due to the pre-processing methods outside the Rayleigh scattering region. Residual Rayleigh scatter and shot noise are related to greater errors related with anisotropy measurements, causing a poor discrimination of *aniso*-EEM map analysis. This combined with minimal structural changes and the relatively high noise spectral measurements implemented here make it a less useful method for assessing small changes in rIgG structure.

3.4 PARAFAC analysis

Anisotropy analysis had suggested that there were some possible changes in Trp emission. We had hypothesized that the thermal fluctuations between 15−35°C would cause enough spectral change to enable the use of PARAFAC to resolve individual fluorophore emission from rIgG in its native state in a similar way to previously reported for HSA.[1] We were aware that protein emission does not vary linearly because of FRET and thus, we cannot assume that the trilinearity, variability, and additivity required[85] for successful and chemically meaningful PARAFAC analysis are present. Nevertheless, for small structural variations it might be reasonable to expect that the structural changes were enough to generate significant emission intensity fluctuations (due to changes in quenching rates for different fluorophores) without significant variations in spectral shape.

Each different pre-processed polarized and un-polarized EEM data of all samples (9 temperatures × triplicate measurements) were used to generated different PARAFAC models. Based on the general rIgG structure (Figure 1.1), we had hoped to recover at least three different components: one from Tyr, a second from Trp located

in a more hydrophobic environment (λ_{em}~330 nm), and a third one from Trp more solvent exposed (λ_{em}>340 nm).[34, 35] Furthermore, changes in Trp emission are not restricted to environmental factors, and three photochemical processes are related with Trp emission in IgG: *(i)* Trp emission from directly excited Trp, producing emission with a high anisotropy; *(ii)* Trp-to-Trp homo-FRET, resulting in lower anisotropy;[123] and *(iii)* Tyr-to-Trp hetero-FRET, also resulting in lower anisotropy values. Thus, one might expect to see this reflected in the presence of one or two additional components with a considerable score.

PARAFAC analysis for all the different pre-processing methods (Table 3.1, Figure 3.9) only yielded two components (PFC1 and PFC2), with the contribution from PFC2 being very small, and sensitive to pre-processing. The two components recovered with PARAFAC were largely related to differences in the excitation spectra. The small number of samples and the relatively noisy data are also likely to be a big problem with resolving anything else.[ll] A four-component model was required to extract changes in emission spectra (Table A-3), but the analysis of the residuals and validations of the models showed that increasing the component number up to four did not produce better fit models.

Table 3.1: Comparison of the two-component PARAFAC model parameters and components obtained for the normalized polarized $EEM_{\parallel}$, $EEM_{\perp}$, and unpolarized EEM_T of the rIgG native state with the different pre-processing methods.

	Pre-proc1			**Pre-proc1/Interp.**			**Pre-proc2**			**Pre-proc2/Interp.**		
	$EEM_{\parallel}$	**$EEM_{\perp}$**	**EEM_T**	**$EEM_{\parallel}$**	**$EEM_{\perp}$**	**EEM_T**	**$EEM_{\parallel}$**	**$EEM_{\perp}$**	**EEM_T**	**$EEM_{\parallel}$**	**$EEM_{\perp}$**	**EEM_T**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 336	280/ 336	278/ 336	280/ 336	280/ 336	278/ 336	280/ 336	278/ 336	280/ 336	278/ 336	278/ 336	278/ 336
C1 Fit model (%)	99.73	99.70	99.85	98.74	99.71	99.65	99.75	99.74	99.85	99.34	99.75	99.84
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	296/ 342	296/ 346	296/ 346	298/ 336	296/ 346	298/ 342	296/ 342	294/ 346	298/ 346	298/ 338	296/ 346	298/ 346
C2 Fit model (%)	0.27	0.30	0.15	1.26	0.29	0.35	0.12	0.26	0.15	0.66	0.25	0.16
Variance explained (%)	99.97	99.99	99.99	99.98	99.99	99.98	99.98	99.99	99.98	99.98	99.99	99.98
CONCORDIA	95.22	88.71	97.02	96.92	86.80	69.21	96.84	91.53	99.61	57.96	88.43	95.76
Split-half analysis (%)	99.13	99.67	99.74	98.39	99.74	99.82	99.14	99.64	99.82	99.57	99.67	99.86

Some of the reasons why we could only recover two components were: *(i)* the temperature range used here generated very small structural changes, without significant spectral changes; *(ii)* the Tyr-to-Trp FRET process was strong, but nearly

[ll] Data quality is key here with small spectral changes.

constant (Figures 3.4/3.5) over the 15−35°C temperature range; *(iii)* structural differences, real or generated by differences in Rayleigh scatter, between the samples were more relevant than small changes in Tyr-to-Trp FRET process that would allow the recovery of directly excited Tyr emission; *(iv)* the emission spectra of all different Trp classes were heavily overlapped and the resolved emission spectrum represented an average of buried and partially exposed/exposed Trp fluorophores present within the rIgG structure (λ_{em}~336 nm); and, *(v)* low sample number and poor SNR.

PFC1 was virtually identical for all polarized and unpolarized EEM (SimI=0.949, RSD=1.78%) corrected with all the different pre-processing methods (Figure A-4). This coupled with the fact that the emission was relatively depolarized (~0.1 for most the emission), and the excitation anisotropy spectra (Figure 3.9) were distinctive[mm] indicated that it originated mostly from FRET rather than direct excitation/emission of the fluorophores. From an analysis of the loadings it was clear that PFC1 was the same for all the different pre-processing methods (Figure 3.10), representing an average of all Tyr and Trp fluorophores. The excitation spectra was very similar to that of Tyr,[45, 124] and the emission was that of Trp. Thus, PFC1 largely represents the Tyr-to-Trp hetero-FRET process, with probably a minimal contribution from directly excited Tyr[nn] at the emission blue edge (in the region where the anisotropy increases). The low calculated anisotropy values recovered for PFC1 (~0.05 at λ_{em} >340nm) could be related to the fact that PFC1 largely represents the Tyr-to-Trp FRET. The somewhat higher anisotropy values for the non-interpolated pre-processing methods (Figure 3.9A/C) could be assigned to the presence of residual Rayleigh scatter and shot noise, which also impacted on component recovery.[22]

[mm] In the case of a freely diffusing small molecule, the excitation anisotropy spectrum of Tyr should be positive from $\lambda_{ex}\geq$260 nm with a maximum increase up to 290 nm.[45, 124] For the Trp, the excitation anisotropy spectrum should not be constant, due to the overlapping excitation of two transition states. A minimum anisotropy value at ~λ_{ex} 290 nm should be observed when the maximum excitation is due to the ^1L_b state, and a maximum at ~λ_{ex} 300 nm indicates the ^1L_a excited state of Trp.[22, 36, 38, 45, 60, 124] ^1L_a is the main Trp emitter in proteins, with only a small contribution from the ^1L_b dipole in the Trp excitation spectrum.[125, 126]

[nn] The presence of directly excited Tyr was also indicated by the fact that the emission of PFC1 was >0 at λ_{em}=300 nm, as it would be expected if pure Trp fluorophores were represented by this component.

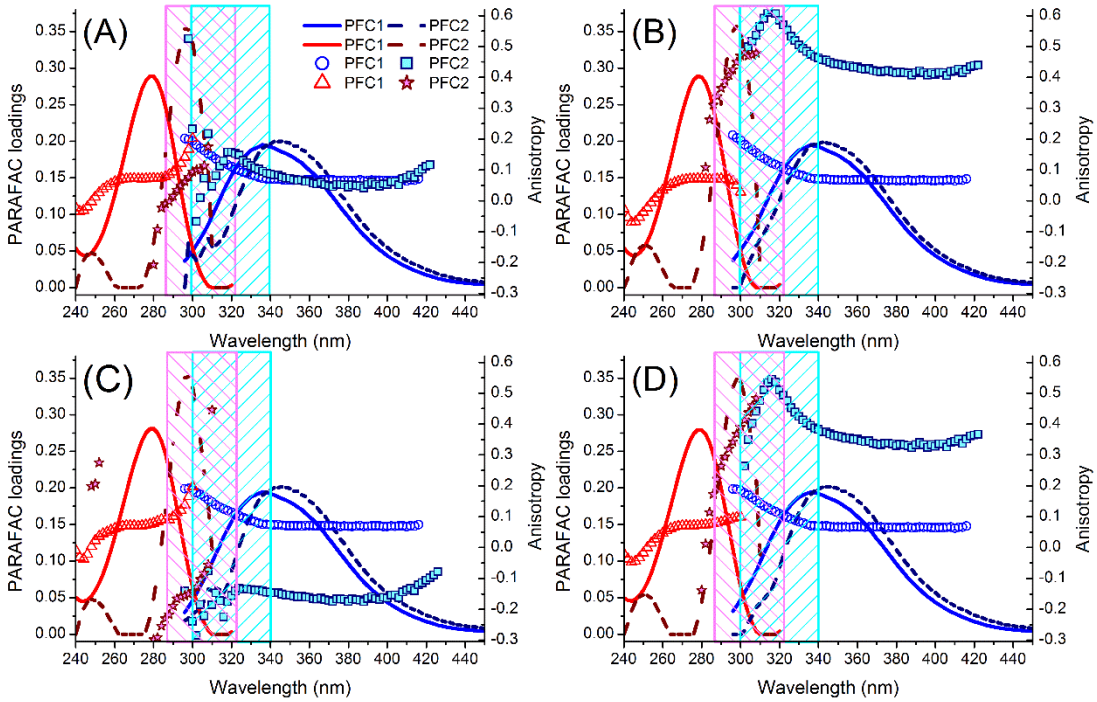

Figure 3.9: Comparison of PARAFAC modelling of the rIgG native structure for the EEM_T datasets with the different pre-processing methods. PFC1 and PFC2 (A) Pre-processing 1, (B) Pre-processing 1/Interpolated, (C) Pre-processing 2, and (D) Pre-processing 2/Interpolated emission and excitation profiles recovered for unpolarized EEM_T, overlaid with the component anisotropy recovered at 20ºC. The anisotropy values were calculated for each component at $\lambda_{ex/em}$= 280/336 and 298/346 for PFC1, and PFC2, respectively. The shaded boxes represent the spectral regions most affected by scatter in the EEM.

The very weak second PARAFAC component, while certainly present, was not clearly resolved, and its properties (*e.g.* anisotropy) were very sensitive to pre-processing. Thus, it was hard to unambiguously assign PFC2 to any specific emission or to use this component for any quantitative analysis purpose. Nevertheless, it does look like that PFC2 was related to directly excited Trp emission because of the high anisotropy at longer wavelengths, and the loadings recovered (Table 3.1, Figure 3.9). This would suggest that the directly excited Trp resolved by PFC2 were most likely located in the hyper variable Fab region (Figure 1.1).[oo] The interpolation step heavily impacted on the recovered anisotropy at short emission wavelengths (Figure 3.9) because of the effects on $pEEM_{\parallel}$ (Table 3.1). The use interpolation caused a ~5-fold increase in PFC2 scores (Figure 3.10, Figure A-5), which can be attributed to the fact that the scatter signal present in $pEEM_{\parallel}$ was much greater than $pEEM_{\perp}$ or EEM_T. This means that the remaining shot noise was very significant and indistinguishable from emission and was thus, included in the data for PARAFAC analysis. This then also contributed to the increased anisotropy values in these regions and generated a lot of

[oo] Where the Trp and Tyr fluorophores were located within a distance from each other that would not result in Tyr-to-Trp FRET interactions.

the variability observed in the *aniso*-EEM maps. Thus, the component anisotropy recovered for $\lambda_{em} \leq 340$ nm was unreliable, but at the longer emission wavelengths should be sensible.

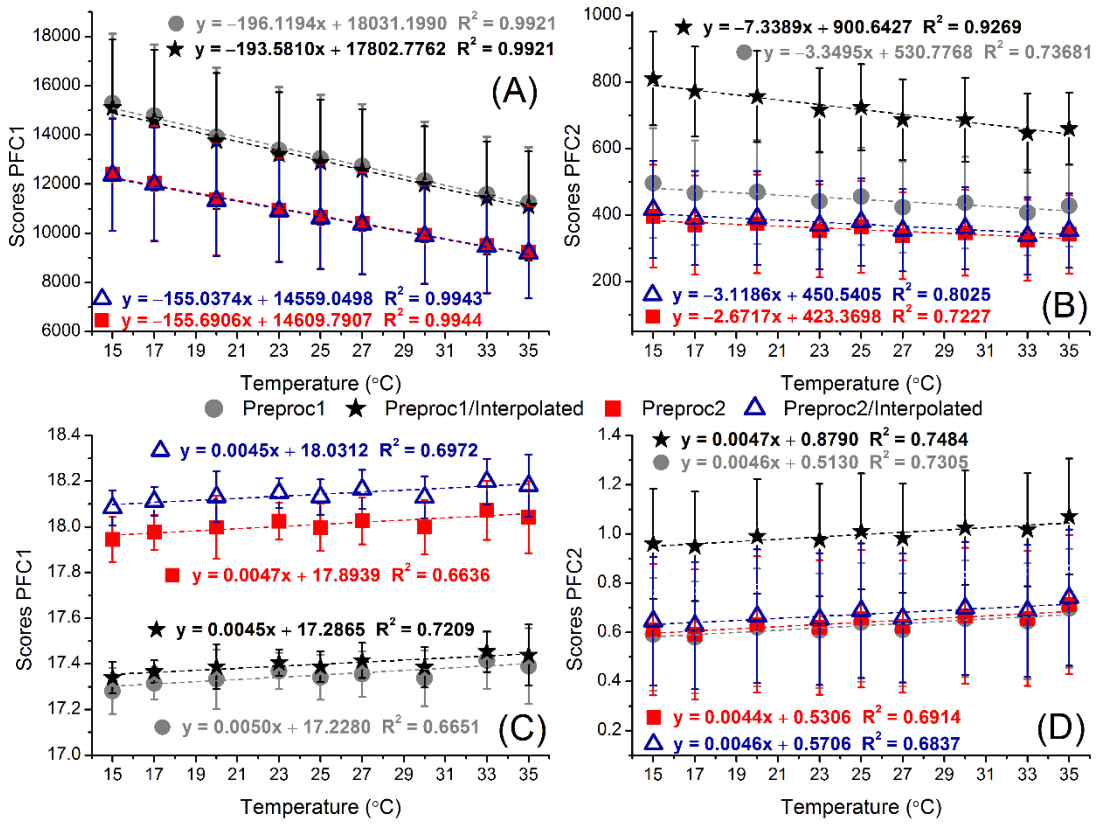

Figure 3.10: Scores plots of the two-component PARAFAC models for the (A/B) non-normalized and (C/D) normalized unpolarized rIgG EEM_T datasets. (Left) PFC1 and (Right) PFC2 scores for the various pre-processing methods used to correct the data. P-values for the regression fits in A and B were <0.05 indicating that the changes were statistically significant. For C and D, the regression fits had P-values of <0.001. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

Our capacity to identify any significant photophysical or structural changes was limited by the small differences between the different pEEM data (Figure A-4), pre-processing methods (Figure 3.9), low scores (Figure 3.10, Figure A-5), relatively high noise (Figures 3.2/3.3), and the number of samples used. Nevertheless, if the PARAFAC modelling process was robust one might expect that the results from the different pre-processing methods should give the same result within reason. In fact, all four pre-processing methods generated virtually the same PARAFAC resolution (Figure 3.9), with small differences in model validation (Table 3.1). It is important to note that all substantial differences in PARAFAC loadings were mainly restricted to the interpolated area (Figure 3.9). The scores plot obtained for the non-normalized EEM_T data (Figure 3.10A/B) showed the scores for both PFC1 and PFC2 decreasing linearly, which indicates that the main change in spectral emission was simply caused

by fluorescence quenching due to an increase in thermal motion with the increase in temperature.[17, 31, 42] Normalizing the data removed this quenching effect, and no changes in PFC scores were observed in the PARAFAC models (Figure 3.10C/D, Figure A-5). This lack of change indicates that there were no significant structural changes that could generate a significant spectral change.

3.5 Conclusions

In this chapter we showed the first attempt to use ARMES combined with PARAFAC analysis for the photophysical characterization of rIgG. The use of a small temperature range (15−35°C) was not enough to cause a significant structural change (IgG unfolds at ~60°C)[17] and the fluorescence changes over this thermal fluctuation were mainly related to simple thermal quenching with very small changes in spectral shape. The difficult in resolving pure Tyr and different classes of Trp restricted PARAFAC analysis, which could only resolve two components, and the second one was very low (<1.3%) and sensitive to pre-processing method.

The combination of a small structural change with the low number of samples used, poor SNR, and incomplete Rayleigh scatter removal resulted in a poor PARAFAC resolution, with low component number. From these results, it was clear that if we were to resolve emission components more precisely from pEEM data we have to collect data with much less Rayleigh scatter, and improved SNR. This would require the use of much more expensive spectrometer with improved stray light rejection (double emission monochromators), and/or the use of smaller slits, but none of these options were practical nor feasible with the Cary Eclipse found in the laboratory.[2, 62]

The next step is to explore the use of pTSFS measurements as an alternative to reduce the influence of Rayleigh scatter.

4. Rabbit IgG Native State Intrinsic Fluorescence Characterization Using ARMES – TSFS

From the previous chapter, it was clear that the presence of Rayleigh scatter in the data collected with pEEM caused significant problems to the use of ARMES/PARAFAC analysis for complex proteins. Here, we evaluated if the use of pTSFS, by minimizing Rayleigh scatter, could improve PARAFAC analysis of native state rIgG emission. First, this chapter shows how the pTSFS data had to be pre-processed to allow the use of PARAFAC for photophysical characterization of the rIgG native state. Second, the results obtained from PARAFAC analysis for the pTSFS and pEEM data were compared to select the best pMDF method for IgG type proteins. The results presented in this chapter were published in *Chemometrics and Intelligent Laboratory Systems*.[92]

4.1 Data pre-processing

The use of an optimal offset ($\Delta\lambda$)[52, 54] for the pTSFS measurements[69] reduces the collection of Rayleigh scatter,[pp] which did not require any Rayleigh scatter correction, and Raman scatter is likely to be the main source of light scattering. Thus, the first pre-processing step required for the pTSFS data was the Raman scatter removal (buffer subtraction), followed by IFE correction (Figure 4.1B). The residual noise amplification in the weak Tyr region was less evident in pTSFS (Figure 4.1) than in pEEM (Figures 3.2/3.3) data. The next step was to rearrange the pTSFS data into an EEM like layout (t-pEEM) generating matrices of 27 samples × 136 λ_{em} × 41 λ_{ex}[71, 74] (Figure 4.1C). The weak emission bellow $\lambda_{em}\leq$292 nm was mainly due to noise, which was amplified by IFE correction. To both eliminate IFE correction artifacts and to make the pTSFS data equivalent to the pEEM data, the t-pEEM data was reduced to λ_{em}=296−450 nm, generating matrices of 27 samples × 78 λ_{em} × 41 λ_{ex} (Figure 4.1E).

[pp] The TSFS emission spectral range of $\Delta\lambda$=20−210 nm yielded matrices of 27 samples × 96 $\Delta\lambda$ × 41 λ_{ex} (Figure 4.1A).

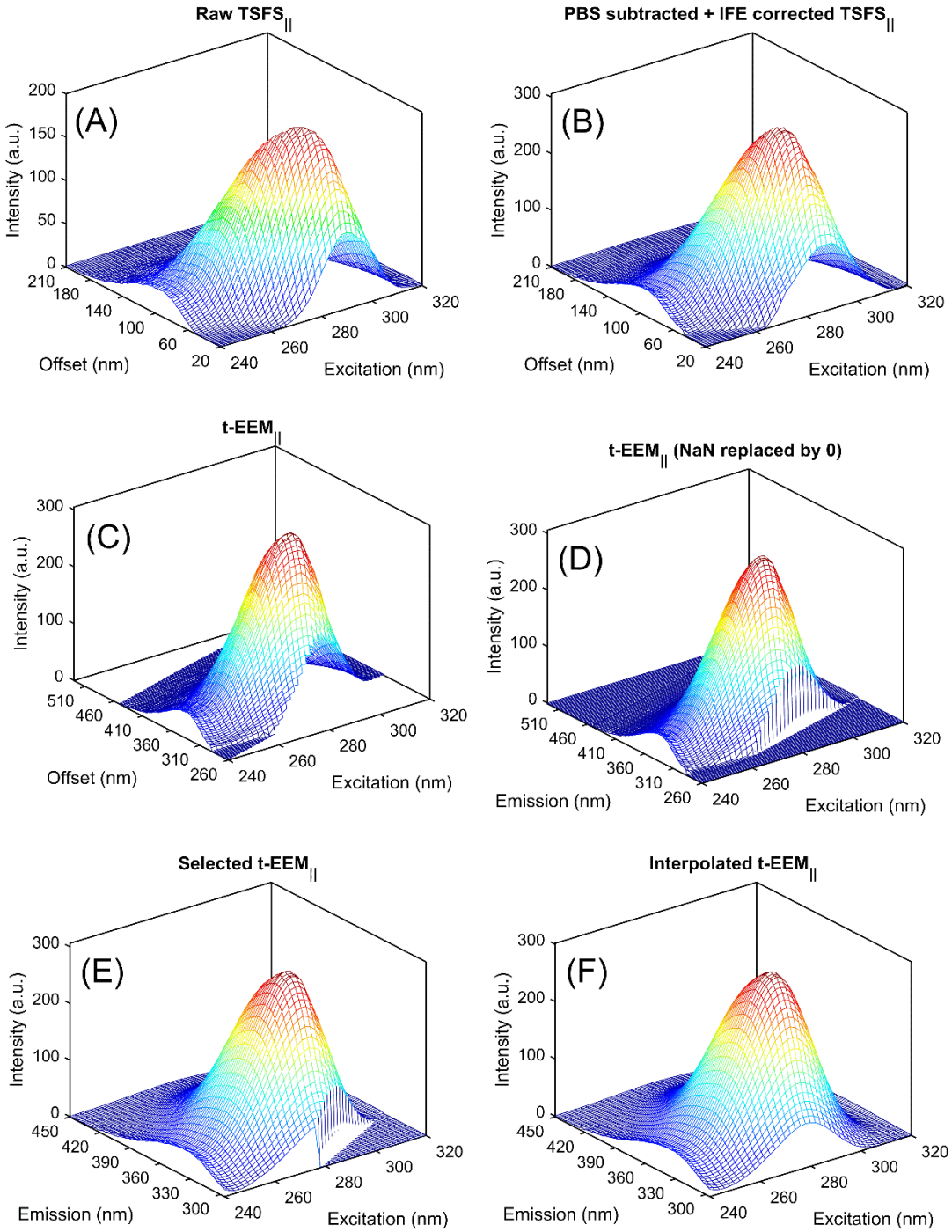

Figure 4.1: (A) Raw pTSFS‖ spectra of rIgG at 20°C. (B) pTSFS‖ buffer subtracted, and IFE corrected. (C) pTSFS‖ data was converted into t-pEEM‖. (D) t-pEEM‖ spectra were cut at λ_{em} 296−450 nm to remove most of the noisy area produced by IFE correction. (E) The areas outside the first Rayleigh scattering were replaced by zeros. (F) The t-pEEM‖ spectra were interpolated and smoothed.

The critical pre-processing steps for t-pEEM data were not related to light scattering correction but associated with an area with no experimentally acquired information in the t-pEEM spectra (Figure 4.1C). The t-pEEM data as collected had a large number of coordinates (λ_{ex}, λ_{em}) with no intensity data (~30%),[74] the absent spectral data region (ASDR). The ASDR was reduced (Figure 4.1E) to a much smaller

region (~8%) when the spectra were edited down to the λ_{em}=296−450 nm range used for data analysis, but still could have a negative impact on PARAFAC modelling. TSFS is relatively under used, and there are only a few studies using t-EEM data combined with PARAFAC analysis. These studies used a combination of missing data and/or zeros to deal with the ASDR.[73, 74, 127, 128] These same studies considered the use of interpolation and/or extrapolation for addressing this issue, but the large amount of missing data (>95%) combined with spectral characteristics of the fluorophores prevented its use. To select the best way of pre-processing t-pEEM data, we evaluated the impact of using three different methods to deal with the ASDR: missing data, insertion of zeros, and interpolation.

The use of missing data to deal with the ASDR makes PARAFAC free to estimate a continuous profile for the spectral components. Unfortunately, these solutions can easily be distorted, leading to discontinuities and sharp peaks,[129] due to the presence of IFE correction induced artifacts, residual light scatter and noise, in the short wavelength emission regions close to these missing values. These factors are exacerbated here by the poor polarizer transmittance and instrument performance in the 250–300 nm region, and the relatively low sample number. The other solutions for facilitating PARAFAC modelling were either to replace the ASDR with zeros (Figure 4.1E)[130] or use interpolated values (Figure 4.1F). However, while using zeros may facilitate PARAFAC convergence,[130] here the ASDR, contains the short Stokes shifted emission (both Tyr and Trp) region. This means that imposing a zero value, artificially distorts the emission data being used for modelling, and this can lead to incorrect spectral profiles being recovered from the PARAFAC models.[110] The use of interpolation has been previously proven to obtain chemically meaningful solution for EEM data,[110] however, it has never been used to reconstruct t-EEM data or that from protein emission.[qq] As with the other methods, areas with low SNR and residual scatter close to the ASDR will adversely affect PARAFAC. One issue with interpolation of protein emission is validating that the reconstructed emission is the true emission and that artifacts are not introduced.

Once the t-pEEM data were pre-processed with one of the three methods, a smoothing function was applied to reduce unwanted noise. The smoothing function

[qq] The use of interpolation for addressing the ASDR issue has been explored,[71, 73, 74] but with a very different sample system.

used in Chapter 3 (Appendix 7.4.1) was not able to handle missing data, the area filled with zeros, or to completely smooth excitation mode. The Savitzky-Golay code was modified (Appendix 7.4.2) to improve smoothing in the excitation mode and to enable its use for the data with missing data and zeros. The pre-processed data were used to calculate anisotropy and to build the *aniso*-t-EEM maps. The final step prior to PARAFAC analysis was to normalize the spectra (by peak maximum).

Polarized t-EEM$_{\parallel}$, t-EEM$_{\perp}$, and unpolarized t-EEM$_T$ rIgG native state data pre-processed with missing data, filled with zeros, and interpolated were analysed with PARAFAC. PARAFAC models for the different pre-processing methods and polarizations were assessed to identify if different emitters could be resolved, and if there were differences in the recovered loadings. Once sensible models and components had been generated, anisotropy spectra for the recovered components were calculated. PARAFAC resolution of t-pEEM[92] were then compared with that previously obtained from the same samples measured with pEEM[91] (Table A-4).

4.2 2D spectral analysis

2D spectra were used to quickly evaluate the spectral information obtained in the t-EEM (Interpolated) and EEM (Pre-processing 2/Interpolated) measurements, considering the differences between Trp and Tyr emission, Tyr-to-Trp FRET, and the amount of spectral variations produced over the 15−35°C temperature range.[91] The Rayleigh scatter removal used in the pEEM data resulted in a red shift in Tyr emission peak maxima (Figure 4.2, grey spectra) in the EEM$_T$ (316 nm) in comparison with the t-EEM$_T$ (312 nm). The shape of the 2D t-EEM spectra was very similar to that obtained from EEM measurements, however, the Tyr emission was stronger in the t-EEM. Tyr fluorescence[48] corresponded to <5% of the maximum intensity of the EEM$_T$ (Figure 4.2C/D), but ~13% of the t-EEM$_T$ (Figure 4.2A/B) spectra. This suggested a bigger contribution of Tyr fluorophores in t-EEM spectral emission, which should facilitate its recovery using PARAFAC.

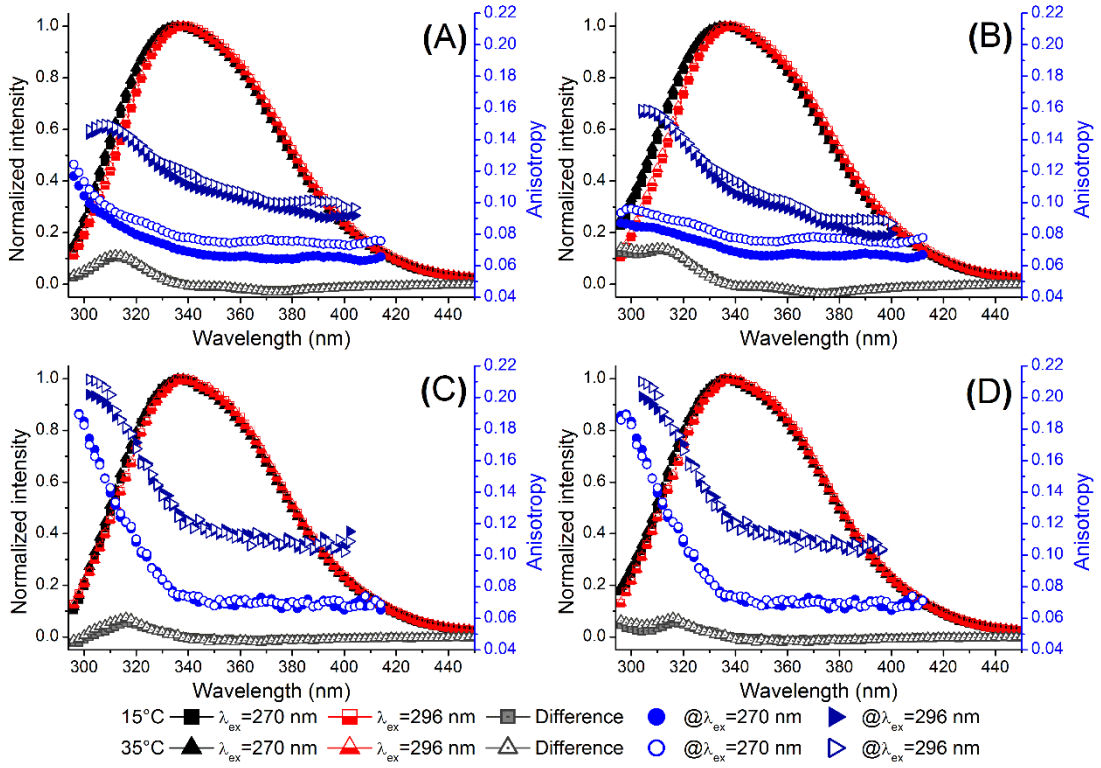

Figure 4.2: Normalized (Top) t-EEM$_T$ and (Bottom) EEM$_T$ emission spectra excited at λ_{ex} 270 and 296 nm and the difference spectra (=λ_{ex} 270 − λ_{ex} 296) at two different temperatures (15° and 35°C), overlaid with the emission anisotropy at λ_{ex} 270 and 296 nm. (A, C) Not IFE corrected. (B, D) IFE corrected using UV-Vis spectra collected at 20°C before MDF measurements.

The better quality of the t-EEM data was also suggested by the smaller anisotropy values obtained at short emission wavelengths (λ_{em}<320 nm). The anisotropy values recovered for the t-EEM data (Figure 4.2B) agree better with those expected from a system where there is extensive Tyr-to-Trp FRET, and where at least some of the Trp fluorophores are directly excited yielding relatively high anisotropy.[22] These we expect to be the Trp residues located in the hydrophobic environments and thus, emit at the blue side of the band.

There were some small, yet significant differences, with a RSD of <5.4 (t-EEM) and <5.0% (EEM), restricted to the emission blue edge when the protein was excited at 270 and 296 nm. However, IFE correction, created an artifact that can be seen at λ_{em}<310 nm in the t-EEM$_T$ and EEM$_T$ difference spectra (Figure 4.2B/D), which may have an impact on PARAFAC analysis. The IFE correction step increased fluorescence intensity in the weak Tyr emission region, and because this region was compromised by Rayleigh scatter in the EEM data, EEM not IFE corrected (Figure 4.2C), and IFE corrected (Figure 4.2D) were very similar. However, this Tyr emission was still very weak, and IFE could have been overestimated.[54] This means that there is still some ambiguity with any measured data in this region.

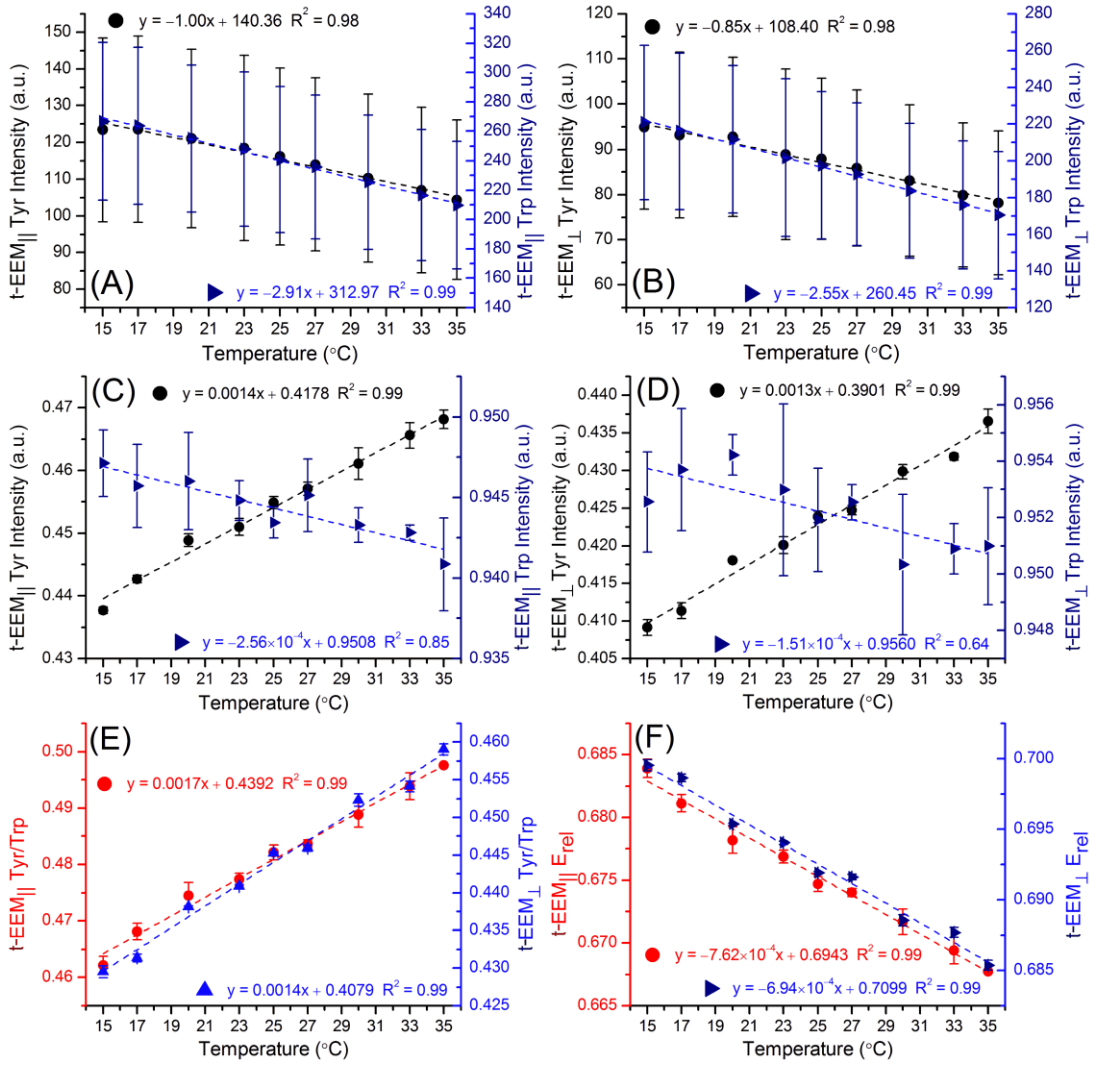

Figure 4.3: (A/B) There was a decrease in Tyr ($\lambda_{ex}/\lambda_{em}$=276/306 nm) and Trp ($\lambda_{ex}/\lambda_{em}$=276/340 nm) intensities for the t-pEEM from 15−35°C, caused by thermal quenching. (C/D) Normalized t-pEEM showed the effects of the Tyr-to-Trp FRET process. (E) The Tyr/Trp ratio increases from 15−35°C. (F) E_{rel} decreases from 15−35°C. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

As observed for the pEEM data (Figure 3.5A/B), Tyr ($\lambda_{ex}/\lambda_{em}$=276/306 nm) and Trp ($\lambda_{ex}/\lambda_{em}$=276/340 nm) I_F decreased linearly with the increase in temperature in the t-pEEM data (Figure 4.3A/B). While the same trend was observed between the different polarizations for both pEEM and t-pEEM data, the increased decay rate for the parallel in comparison with the perpendicular polarizations was less pronounced for the t-pEEM data. This smaller difference between polarizations for t-pEEM in comparison with pEEM highlighted the influence of Rayleigh scatter. Figure 4.3 shows that there was an increase in Tyr fluorescence intensity for the normalized data, however, we could not confirm that there were any real changes in Tyr-to-Trp FRET and Trp classes. Thus, the t-pEEM seems to give a better picture of the Tyr-to-Trp FRET process because of the lower Rayleigh scatter influence. The presence of a

simple thermal quenching was confirmed by the linear trend observed for the Tyr/Trp ratio (Figure 4.3E) and the relative FRET efficiency (Figure 4.3F).

The 2D spectral analysis indicated what we might be able to resolve with PARAFAC. The first step is to examine how appropriately the pTSFS measurements removed any scattered light contamination compared to the conventional pEEM based approach (Chapter 3). The most sensitive way of doing this prior to PARAFAC analysis is to first look at the anisotropy maps with focus on the blue edge of the emission space, which is highly affected by Rayleigh scatter.

4.3 *Aniso*-MDF maps

The anisotropy variation across the rIgG emission space represented in the *aniso*-MDF maps was caused by a variety of factors such as: type and number of fluorophores present in the protein, fluorophore location in the protein, local fluorophore motion, changes in FRET, and variations in the physicochemical environment.[1] Rayleigh scattered light contamination will also be evident as regions of abnormally high anisotropy at the emission blue edge. This is clearly shown in Figure 4.4 where the *aniso*-EEM map is much more heavily distorted as evidenced by the much higher anisotropy at the emission blue edge, which also means that the weak Tyr signal is masked. The differences between the *aniso*-t-EEM maps that were generated from data which was interpolated, used the missing data approach, or filled with zeros, were mainly due to reconstruction of part of the Tyr emission region in the interpolated method (Figure 4.4, black box).

Visual inspection of the full emission space showed some variation in *aniso*-t-EEM maps with the increase in temperature (Figure 4.5), which were virtually identical for the datasets whether interpolated or not. Previously, the *aniso*-EEM maps only showed small changes for λ_{em}>336 nm (Figure 3.6), but the changes in *aniso*-t-EEM maps were not restricted to the Trp region. For the *aniso*-EEM maps, emission blue edge was affected by Rayleigh scatter and residual shot noise, obscuring the weak Tyr signal. Since the t-pEEM datasets were much less affected by Rayleigh scatter, we could better observe changes at λ_{em}<336 nm (*i.e.* Tyr emission), indicating that, here, PARAFAC analysis could potentially yield a significant component related to Tyr emission.

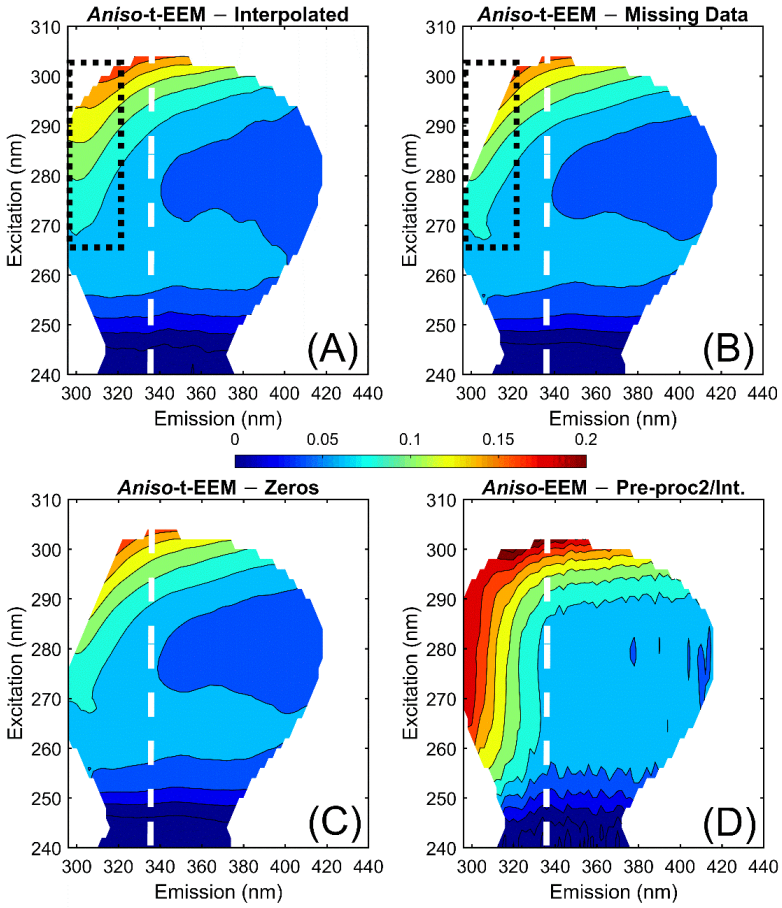

Figure 4.4: (A/B/C) *Aniso*-t-EEM and (D) *aniso*-EEM maps (corrected for Rayleigh scatter) of rIgG at 20°C. The white dashed lines mark λ_{em}=336 nm excitation lines. The black box shows the Tyr region reconstructed by interpolation. The colour bar on the middle represents the anisotropy scale.

An examination in the change in *aniso*-t-EEM maps over the temperature range (Figure 4.5) did not show any significant differences (ANOVA, $p>0.05$) between the *aniso*-EEM and *aniso*-t-EEM maps in the Trp emission region, λ_{em}>336 nm. For both EEM and t-EEM maps, there were only small changes in anisotropy at λ_{em}>336 nm. The influence of residual shot noise and Rayleigh scatter could be observed when the anisotropy at a single excitation/emission wavelength combination was selected. Selecting a single anisotropy point at $\lambda_{em}\geq$336 nm should give similar values from both *aniso*-EEM and *aniso*-t-EEM maps because, technically, this area was not heavily affected by Rayleigh scatter. The selected Trp anisotropy (Figure 4.5H) value increased by only ~3% from the *aniso*-EEM to the *aniso*-t-EEM. On the other hand, a single point at λ_{em}<336 nm should result in a bigger gap between *aniso*-EEM and *aniso*-t-EEM values. The anisotropy values for the selected Tyr emission (Figure 4.5G) showed the biggest difference (>60%) between the *aniso*-EEM and *aniso*-t-EEM, confirming the influence of residual light scatter in the *aniso*-EEM maps.

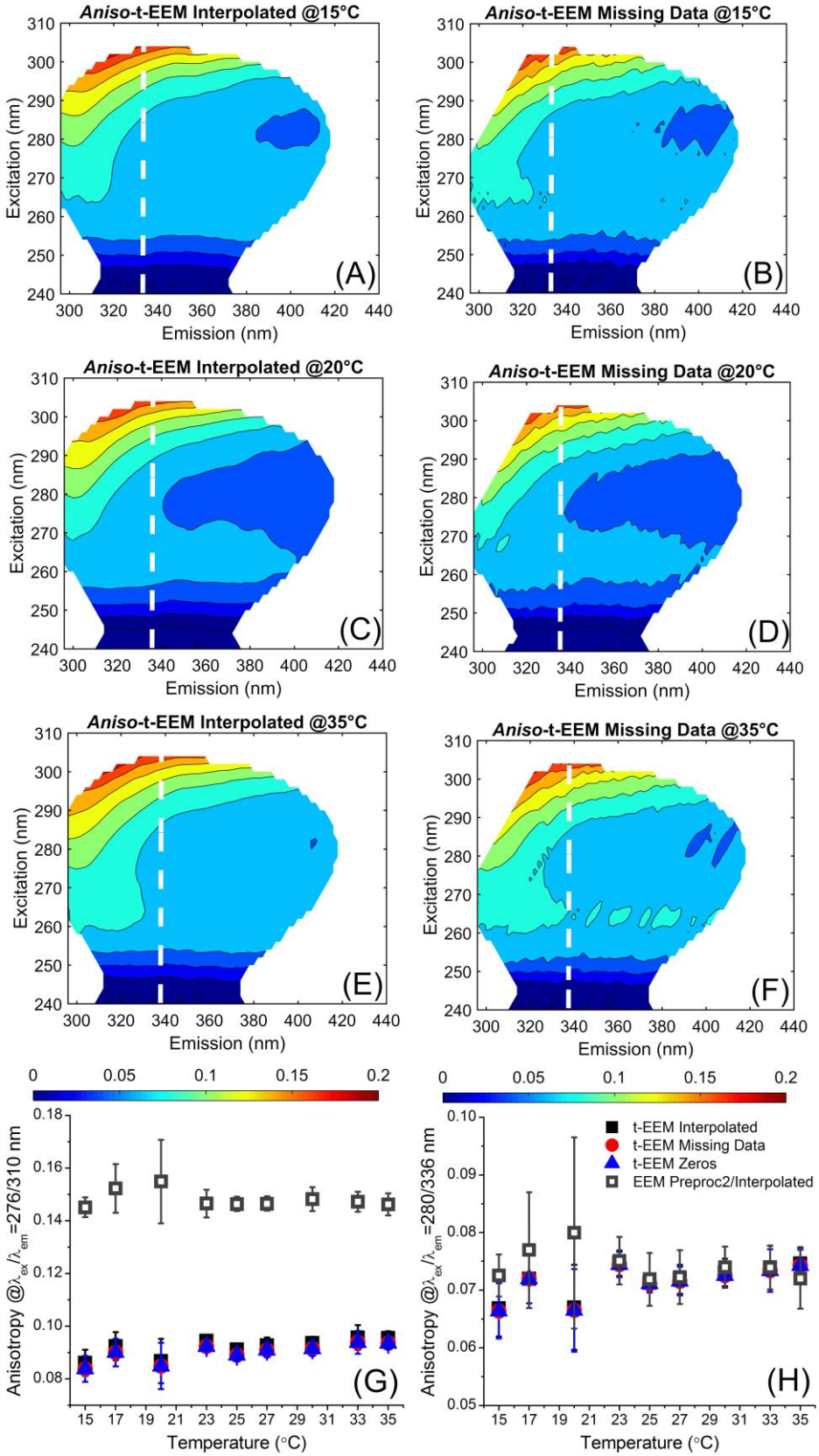

Figure 4.5: *Aniso*-t-EEM maps for rIgG collected at (A/B) 15°C, (C/D) 20°C, and (E/F) 35°C. *Aniso*-t-EEM maps for the data corrected with (A/C/E) interpolation, and (B/D/F) missing data (<8% of data points). The white dashed lines mark λ_{em}=336 nm excitation lines. The colour bars on the bottom represent the anisotropy scale. (G) Tyr anisotropy at $\lambda_{ex}/\lambda_{em}$=276/310 nm, and (H) Trp anisotropy at $\lambda_{ex}/\lambda_{em}$=280/336 nm. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

SimI[122] was used to better asses the overall degree of change in *aniso*-MDF maps (Figure 4.6). The trend shown in Figure 4.6 for the pTSFS was virtually identical to that for the pEEM. The key result was the dip observed at 20°C.[rr] Changes in anisotropy were greater at lower temperatures (≤20°C), with a significant difference between replicate measurements.[ss] This was caused by a combination of concentration differences (up to ~31%) coupled with varying Rayleigh scatter (mainly for the pEEM, amplified by IFE correction). Although Rayleigh scatter was reduced in the pTSFS measurements, there was still some residual scattered light present in the pTSFS spectra.

SimI analysis over the full emission space clearly showed that *aniso*-t-EEM maps (RSD=7.7 and 10.2% for the interpolated and missing data/zeros datasets, respectively) were more reproducible than *aniso*-EEM maps (RSD=21.50%). The improved reproducibility of the interpolated *aniso*-t-EEM data compared to the missing data or added zeros data was due to the reconstructed ASDR. For the Tyr/Trp specific emission regions (Figure 4.6B/C) this variation was noticeably reduced, for the *aniso*-EEM (RSD=7.6 and 8.6% for the Tyr and Trp regions, respectively) and *aniso*-t-EEM maps (RSD~5.9 and 6.6% for the Tyr and Trp regions, respectively). This suggests that more of the temperature induced emission changes were concerned with Trp emission.

[rr] The reproducibility of this small dip in SimI at 20°C for the EEM and t-EEM data collected three weeks apart (Table 2.1) indicated that this effect was real.

[ss] Although all the experimental procedures were carefully carried out, with every step meticulously planned, there is no way of determining that the differences observed between the replicate measurements were caused by the protein itself, or by any other factor. The use of polyclonal rIgG from multiple vials contributed to these variations.

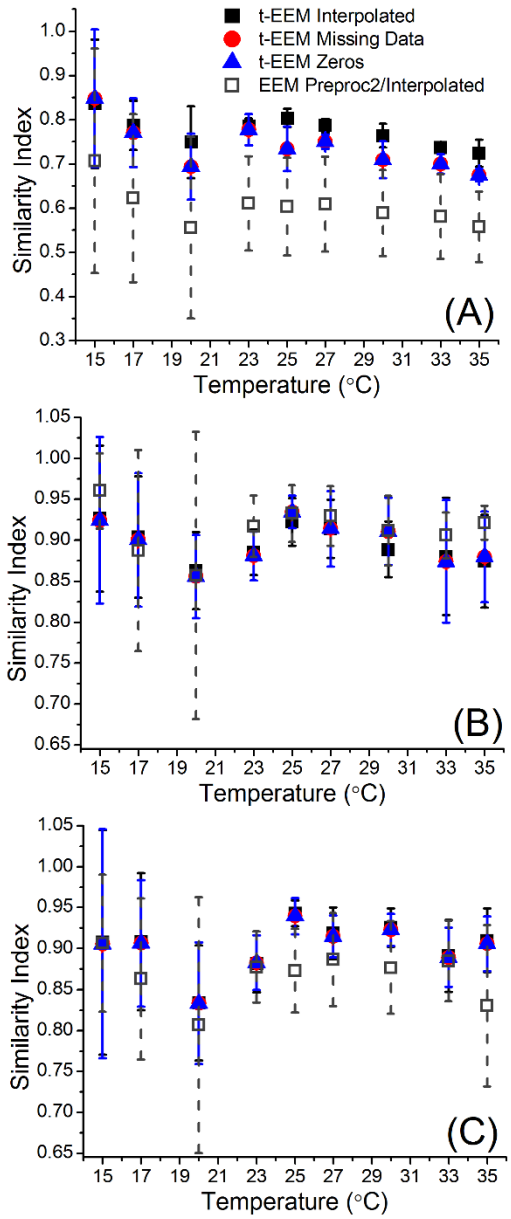

Figure 4.6: Plots of SimI values calculated for the various *aniso*-MDF maps over the 15−35°C temperature range. In each case, the reference spectrum was the MDF collected at 15°C from the first replicate sample (R1). (A) The full emission space, (B) the Tyr emission region ($\lambda_{ex}/\lambda_{em}$=270−290/300−320 nm), and (C) the Trp emission region ($\lambda_{ex}/\lambda_{em}$=280−300/320−360 nm). Error bars generated from the standard deviation from triplicate measurements of the independent samples.

It was clear that the *aniso*-MDF maps obtained from the pTSFS measurements were much superior to those from pEEM because of the reduced scattered light contamination at short emission wavelengths (<320 nm). Here the big issue is that it is expected to resolve two components with PARAFAC analysis from the pTSFS data, but larger sample numbers, and SNR improvement are also needed.

4.4 PARAFAC analysis

Meaningful PARAFAC deconvolution of intrinsic fluorophore emission from proteins is inherently challenging because of several factors: the non-linearity associated with Tyr-to-Trp FRET process, emission and Rayleigh scatter overlap, particularly at short wavelengths, large differences in relative emission contributions between fluorophores, and the large numbers of fluorophores involved which only show small spectral variation (Figures 4.2−4.6). For pEEM measurements,[91] one critical factor that limited the effective application of PARAFAC to resolve Tyr emission contribution was the Rayleigh scatter. Here, the better quality of the t-pEEM data at short emission wavelengths resulted in a stronger Tyr emission (Figure 4.5H), which should facilitate its recovery by PARAFAC. However, the combination of TSFS and PARAFAC analysis is not widely used,[71, 73, 74] and requires first that the data is converted into an EEM like layout. The problem related to the use of t-EEM and PARAFAC was to determine what is the impact of the different methods for dealing with the ASDR (*i.e.* emission blue edge) on PARAFAC resolution.

PARAFAC modelling was implemented under equivalent conditions for the different pre-processing methods and the model results then compared (Table 4.1, Figure 4.7). There was an improvement in data quality, but Table 4.1 (and Table A-4) shows that in all cases we could only resolve two components (as suggested by CONCORDIA and split-half analysis), and that these were very similar. Independently of the pre-processing method used, the same two components were resolved by PARAFAC (Figure 4.7, Figure A-6), the major component being associated with Trp emission (PFC1, ~92−97%), and a second weaker one mainly associated with Tyr emission (PFC2, >3−7%). Analysis of the scores (Figure 4.8, Figure A-7) for the un-, and normalized data clearly shows that the emission was quenched, and that this mainly involved the Trp emission.[tt] The increased thermally induced quenching of Trp, changed the Tyr/Trp ratio (Figure 4.3E) facilitating the resolution of the two components (Figure 4.7). Although the same changes in Tyr/Trp

[tt] This could be confirmed by the 2D spectral analysis of Trp and Tyr shown in Figure 4.3, which shows that the decay rate for Trp quenching was nearly three times faster than the Tyr decay rate. The normalized scores plot (Figure 4.8C/D) shows that only PFC2 has a linear fit, which was also observed for the Tyr emission (Figure 4.3C/D), confirming that PFC2 was related to the weak Tyr emission. The lack of linear fit for PFC1 scores and for the Trp emission (Figure 4.3C/D) clearly indicated that there was no real change in protein structure that could have caused a significant spectral change.

ratio were observed for the pEEM data (Figure 3.5), PARAFAC was not able to resolve a pure Tyr emission component (Figure 3.9).

Table 4.1: Comparison of the two-component PARAFAC model (50 repetitions) parameters and components obtained for the normalized polarized t-EEM$_{\parallel}$, t-EEM$_{\perp}$, and unpolarized t-EEM$_{T}$. The spectral data were corrected using interpolation, missing data (7.9%), or zeros. Samples were measured over the 15−35°C temperature range (9 temperatures × triplicate measurements).

	Interpolated			**Missing Data**			**Zeros**		
	t-EEM$_{\parallel}$	**t-EEM$_{\perp}$**	**t-EEM$_{T}$**	**t-EEM$_{\parallel}$**	**t-EEM$_{\perp}$**	**t-EEM$_{T}$**	**t-EEM$_{\parallel}$**	**t-EEM$_{\perp}$**	**t-EEM$_{T}$**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/338	280/342	280/340	280/338	280/340	280/340	280/336	280/340	280/338
C1 Fit model (%)	96.99	92.60	95.03	96.55	95.17	96.46	96.35	96.07	96.23
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	278/298 (278/310[a])	278/314	278/312	320/300 (278/300[a])	276/312	278/300 (278/310[a])	276/306	276/306	276/306
C2 Fit model (%)	3.01	7.40	4.97	3.45	4.83	3.54	3.65	3.93	3.77
Variance explained (%)	99.99 ($\pm4\times10^{-9}$)	99.99 ($\pm3\times10^{-8}$)	99.99 ($\pm1\times10^{-8}$)	99.99 ($\pm2\times10^{-5}$)	99.99 ($\pm3\times10^{-9}$)	99.99 ($\pm2\times10^{-9}$)	99.58 ($\pm1\times10^{-5}$)	99.70[b] ($\pm9\times10^{-8}$)	99.66 ($\pm8\times10^{-8}$)
CONCORDIA (%)	29.25 (±0.12)	31.09 (±0.52)	23.87 (±0.38)	99.80 (±0.04)	63.27 (±0.004)	65.98 (±0.003)	85.80 (±0.46)	76.53 (±0.002)	78.59 (±0.004)
Split-half analysis (%)	99.97	99.96	99.95	99.80	99.03	99.61	56.84	99.72	85.67
Iterations	7 (±16)	22 (±32)	9 (±19)	3424	1242 (±1)	2119 (±2)	4 (±3)	4 (±0)	4 (±0)
Time (s)[c]	7 (±1)	10 (±1)	8 (±1)	47 (±21)	22 (±2)	30 (±2)	5 (±1)	7 (±1)	6 (±1)

[a] PARAFAC components without artifacts. [b] CONCORDIA values of local minima models were removed. One local minima model was found. [c] Time required for generating the two-component PARAFAC model.

The same two components were resolved for the t-EEM data, however, there were substantial changes in the profiles recovered for each fluorophore, particularly for the weaker mostly Tyr emitting PFC2 (Figure 4.7, Figure A-6), and this is a problem for the photophysical interpretation of the PARAFAC models. These changes were related to the pre-processing methods used to deal with the ASDR and their effect on the short-Stokes shifted emission region. PFC1 (Trp) was virtually identical in terms of the recovered spectral profiles (Figure A-6), except for the models filled with zeros. Filling the ASDR with zeros restricted PARAFAC to an abnormal solution which led to an underestimation of the emission blue edge of the Trp fluorophores.[130] The use of zeros facilitated PARAFAC convergence (lowest number of iterations, Table 4.1), however, it could cause a premature deconvolution, increasing the variability of the spectral profiles of the calculated models.[uu] Consequently, the use of zeros slightly

[uu] This was the case for the polarized t-EEM$_{\perp}$ data, in which we obtained one local minima model where PFC2 ($\lambda_{ex}/\lambda_{em}$=290/306 nm) was not in agreement with the other 49 models ($\lambda_{ex}/\lambda_{em}$=276/306 nm). The single local minima model was removed from our analysis (Table 4.1). In addition, filling the missing data with zeros restricts PARAFAC resolution to the region where $\lambda_{ex}/\lambda_{em}>0$.

underestimated both Trp and Tyr relative contributions (Figure 4.7E/F) compared to the other two methods tested here.

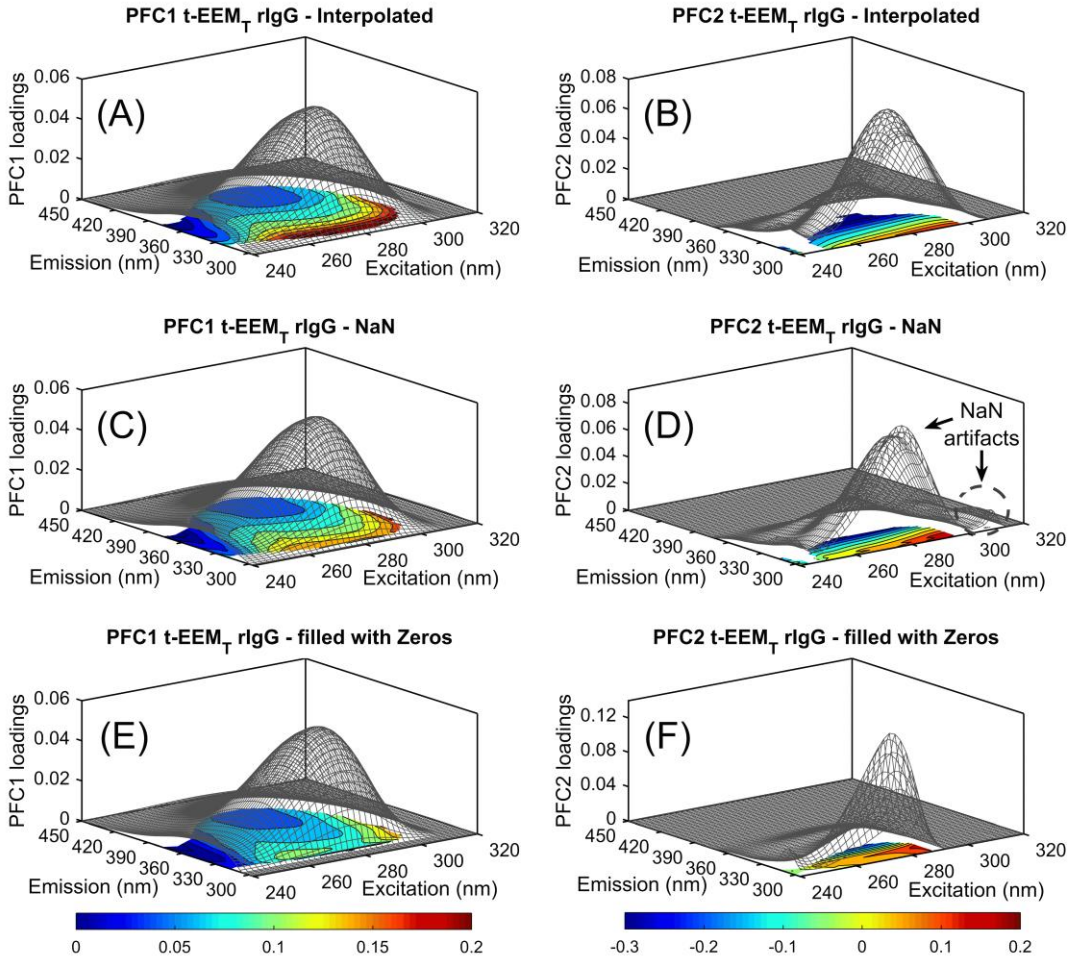

Figure 4.7: Comparison of PARAFAC modelling of the t-EEM$_T$ data corrected with (A/B) interpolation, (C/D) missing data (NaN), and (E/F) zeros. PFC1 (Trp) was virtually identical for the datasets with (A) interpolation, (C) NaN, and (E) zeros. Although PFC2 resolved the Tyr signal for the (B) interpolated, (D) NaN, and (F) zeros, there were slight differences caused by the presence of NaN and zeros. t-EEM$_T$ landscapes overlapped with *aniso*-t-EEM maps for rIgG measured at 20°C. The colour bars on the bottom represent the anisotropy scale.

When the missing data approach to ASDR was implemented, PARAFAC was free to better estimate the continuous shape of Trp and Tyr emission. This was because it did not restrict or modify the data, and thus better extracted spectral profiles which might be a truer representation of the actual emission. The missing data method did, however, create some artifacts at the emission blue edge (Figure 4.7D) which was amplified by the IFE correction as shown in the excitation slab corresponding to PFC2 emission, and this was particularly severe for the parallel polarized light data (Figure A-6). The use of interpolation significantly decreased the required computational time (Table 4.1) and led to a resolution that appears to be spectrally acceptable. Thus, the use of interpolation to reconstruct the ASDR could overcame the problems caused

using missing data and zeros. However, CONCORDIA values were reduced for the interpolated data compared to the other methods, which was caused by extension of emission into the ASDR and thus increased emission overlap (Figure A-6). One might expect that this suggests that three components should be recoverable by PARAFAC (Tyr, Trp from hydrophobic locations, and Trp externally located on the protein). However, the three component PARAFAC model (Table A-4) did not show any improvement in the model performance quality parameters.

Differences in quality and model parameters for the t-EEM data pre-processed with missing data and interpolation methods were associated with sample spectral characteristics, and particularly the intrinsic anisotropy of the emission from the large protein molecule, and the degree of FRET. The comparison of the t-EEM$_T$ with the t-pEEM results (Table 4.1, Figure A-6) makes these effects clear. For rIgG, the parallel polarization measurements seem to be more directly sensitive to Tyr-to-Trp FRET process (*i.e.* lower PFC2 scores, Figure A-7), and to the presence of residual scatter.[vv] It was clear that, in the t-pEEM$_\parallel$ models, there was an element of the spectral profile that could be assigned to Rayleigh scatter[ww] (PFC2 blue shifted). In the t-EEM$_T$ and t-pEEM$_\perp$ models there was a clear shoulder at λ_{em} ~350 nm (Figure 4.7, Figure A-6) due to Trp emission which resulted in a slightly higher score (~5−7%) for PFC2. However, there were no significant differences between the polarized and unpolarized t-EEM PARAFAC models using the ASDR filled with zeros, which is surprising, as there are real differences between the polarized and unpolarized spectra[22] of rIgG. Thus, this observation rules out the use of zeros to correct the ASDR for intrinsic protein fluorescence analysis.

[vv] Due to a combination of factors such as the short fluorescence lifetime of Tyr (<4 ns) coupled with the long (26 ns) rotational correlation time of rIgG.[131]

[ww] The Rayleigh scatter contamination was more evident, but not restricted to, the t-pEEM$_\parallel$ models. There will still be Rayleigh scatter in the t-pEEM$_\perp$, but it should be very weak. The pTSFS$_\perp$ measurement offered the best t-pEEM$_\perp$ quality data for modelling Tyr and Trp emission because it was less affected by residual scatter, generating the most reasonable estimate of the true contributions of each fluorophore. Still, PFC2 resolved for the t-pEEM$_\perp$ data did not represent a pure Tyr emission and there were some associations with Trp emission present.

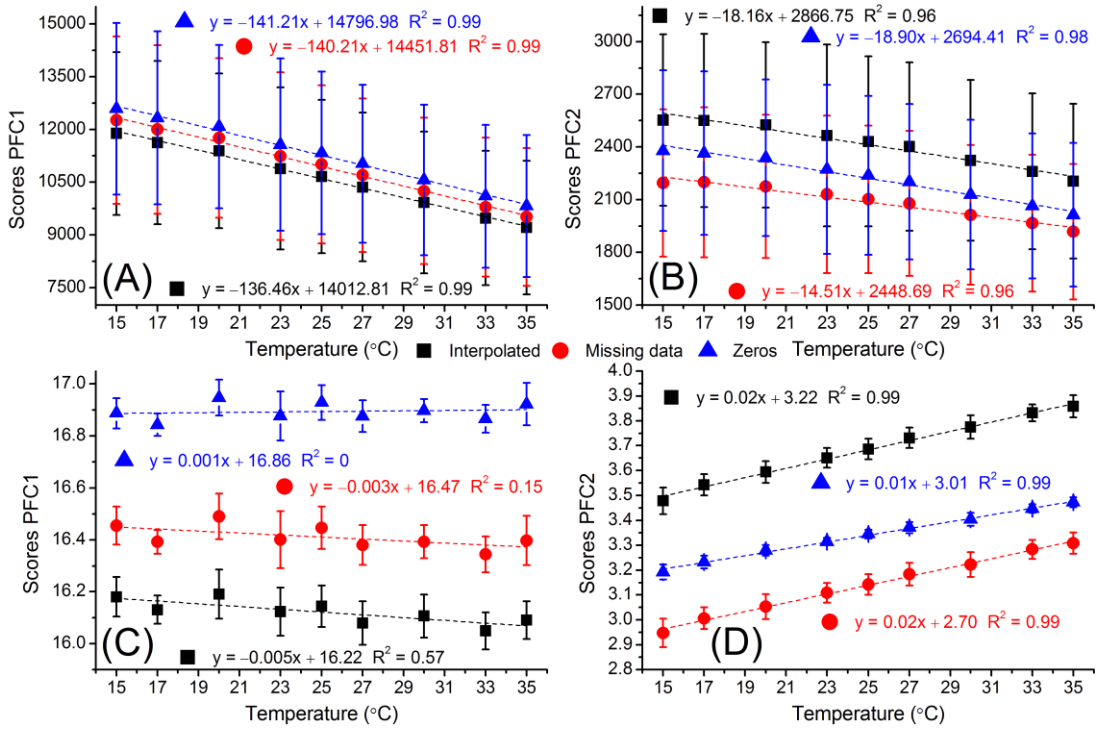

Figure 4.8: Scores plots of two-component PARAFAC models for the: (top) non-normalized, and (bottom) normalized unpolarized t-EEM$_T$ datasets with the different ASDR treatment methods. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

For better analysis of protein structural change, it would be useful to be able to fully resolve Trp and Tyr emission,[1, 48, 132] to provide both spectral profiles and relative contribution values. Analysis of the PFC1 (Trp) and PFC2 (Tyr) scores from the un-normalized t-EEM$_T$ data from each method showed a linear decrease with increasing temperature due to quenching (Figure 4.8A/B). The trends were the same for all three data handling methods with no significant differences (ANOVA, $p>0.05$) in PFC1 and PFC2 scores trends, indicating that any of the PARAFAC scores could be used to monitor structural/concentration changes (r Pearson>0.99). The big error bars in Figure 4.8A/B were due to differences in protein concentration between the replicate samples (~31%), which is why the normalized data is better for investigating structural changes.

Normalization reduced these errors, and this was reflected in the smaller error bars (~1%) obtained for PARAFAC scores (Figure 4.8C/D). While scores trends for PFC1 (Trp) showed that there were no significant structural changes in the rIgG between 15° and 35°C, PFC2 scores (Tyr) slightly increased with increasing temperature (Figure 4.8D). A major difference between the scores for the different pre-processing methods used to deal with the ASDR was that the interpolated data usually showed higher PFC2

scores.[xx] This increase in scores would indicate that the use of interpolation was the best option for generating protein fluorescence data suitable for PARAFAC analysis to monitor subtle changes in protein structure and behaviour via changes in the Tyr/Trp emission ratio.

The use of different methods for dealing with ASDR resulted in significantly different PARAFAC outputs (spectral profiles and scores). The insertion of zeros and missing data underestimated the relative contribution of the Tyr emission resolved with PARAFAC. These methods were prone to generate spectral artifacts, mostly at the emission blue edge where the noise and light scatter contamination was still a problem. Another disadvantage related to the use of missing data was the high computational time required for this method, making it unreasonable for online or inline applications. The use of interpolation to reconstruct the ASDR was the most suitable option, generating the best quality PARAFAC models based on spectral profiles and scores recovered. However, caution was required for the interpretation of this data, particularly the spectral profiles, because the validation of PARAFAC output from intrinsic emission is still unproven (but known for small molecule examples). The general spectral trends observed in the scores were similar between the different pre-processing, but the magnitudes were significantly different. Thus, when using scores plots to follow protein changes it is advised to use a combination of two ASDR approaches (missing data and interpolation) to guarantee a useable model outcome for analytical purposes.

4.5 Inner filter effect (IFE) impacts in pMDF PARAFAC analysis

IFE is more pronounced at the emission blue edge,[50] the Tyr signal could be improved via IFE correction, facilitating the recovery of this weak emitting component by PARAFAC. However, there were no significant differences for the pEEM data not corrected (Figure 4.9A, Table A-5), or corrected (Figure 4.9B, Table A-6) for IFE, and the Tyr component was still not clearly resolved. This observation increased the initial suspicion that the Rayleigh scatter removal necessary in the pEEM data was affecting Tyr emission. The IFE correction step was supposed to correct for differences in protein emission caused by different light scatter, but the residual scatter was likely

[xx] PFC2 scores for the t-pEEM$_{\perp}$ were higher for the interpolated (~7%) than the missing data (~5%), suggesting that the interpolation method improved the resolution of Tyr emission.

overcorrected,[54] and the scores plots for the two-component PARAFAC models still showed a clear difference between the replicate samples (Figure 4.10B/D).

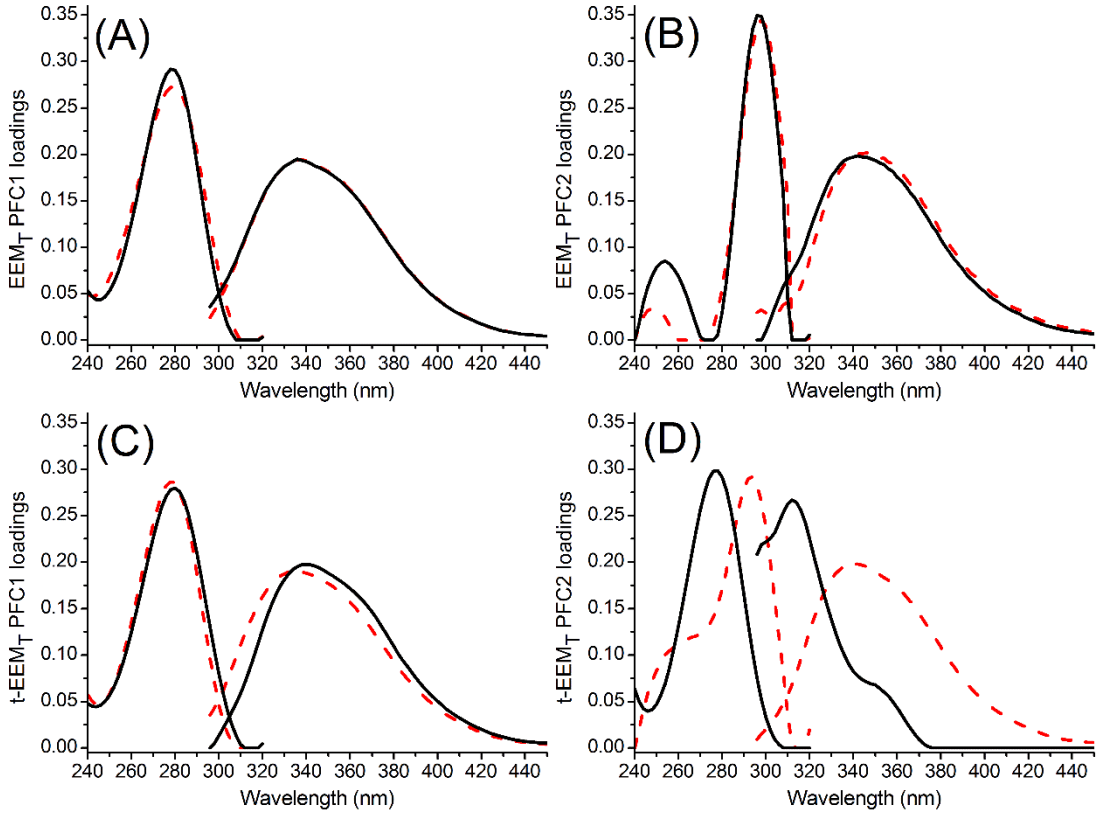

Figure 4.9: Comparison of PARAFAC modelling of the (A/B) EEM$_T$ and (C/D) t-EEM$_T$ datasets not corrected (dashed red line), and corrected (black line) for IFE.

On the other hand, PARAFAC models for the non-IFE corrected (Table A-5) and IFE corrected (Table A-6) datasets were different for the t-pEEM data, with IFE correction completely changing PFC2 resolution (Figure 4.9D). PARAFAC resolution of the t-pEEM data not IFE corrected indicated that Rayleigh scatter still influenced protein emission, just as observed for the pEEM data. Independently of the pMDF measurement, PARAFAC analysis of the data not corrected for IFE resolved two components: PFC1 resolving a major Trp contribution, with a small contribution from the weak Tyr emission; and PFC2 representing the directly excited Trp. IFE correction changed PARAFAC resolution for the t-pEEM data: PFC1 clearly resolved a purer Trp contribution (Figure 4.9C),[22] and PFC2 was finally able to resolve the Tyr emission contribution (Figure 4.9D).

Figure 4.10 shows that the differences caused by IFE were clearly observed in the score plots for the two-component PARAFAC models resolved for the EEM and t-EEM. Without IFE correction, the PFC1 *vs.* PFC2 scores plots for EEM (Figure

4.10B) and t-EEM (Figure 4.10E/F) data were mainly separating the replicate measurements, with no clear trend for the temperature increase for the EEM (Figure 4.10A).[yy] Thus, if the pMDF data is not corrected for IFE, PARAFAC analysis is more likely to resolve these differences caused by IFE as a pure component. Once the datasets were corrected for IFE, any spectral differences caused by IFE were presumably removed, and PARAFAC should be able to resolve structural/environmental changes (*i.e.* changes in Tyr-to-Trp FRET and/or in Trp classes). This was observed for the t-EEM, as PFC1 *vs.* PFC2 scores plot were showing a trend with the increase in temperature (Figure 4.10G/H), but not for the EEM data (Figure 4.10C/D).

The analysis of the influence of IFE in pMDF measurements confirmed that the pTSFS measurements were much superior to pEEM because of the reduction in Rayleigh scatter contamination at both the short emission wavelengths (<320 nm) and long excitation wavelengths (≥290 nm).

[yy] Changes in spectral shape caused by IFE were not removed by normalization.

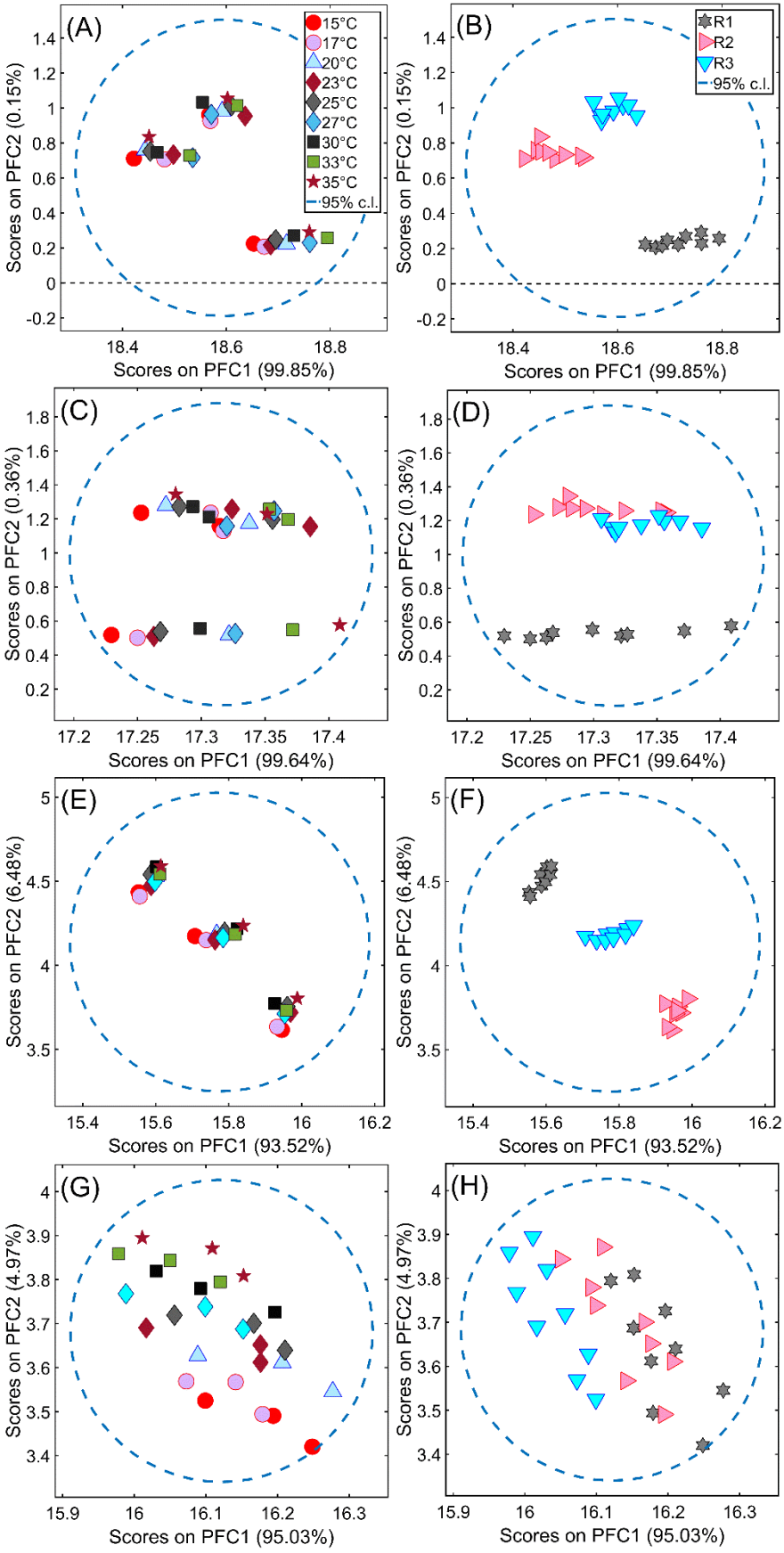

Figure 4.10: PFC1 *vs.* PFC2 scores plots for the two-component PARAFAC model resolved for the pre-processed (A/B) EEM_T non-IFE corrected, (C/D) EEM_T IFE corrected, (E/F) t-EEM_T non-IFE corrected, and (G/H) t-EEM_T IFE corrected rIgG native state emission.

4.6 Conclusions

This chapter showed that the combination of pTSFS measurements, with precisely defined pre-processing, and PARAFAC analysis is the more suitable method for the analysis of intrinsic protein fluorescence without significant structural changes. In comparison with the pEEM study presented in Chapter 3, the pTSFS method has sufficient sensitivity to extract signals from the weak Tyr emission which is key to being able to observe and measure subtle changes in protein structure.[43] However, one must take in consideration that even though the pTSFS measurement is a better choice for the analysis of small structural changes, pEEM might still be preferred when analysing significant changes in protein structure and aggregation, as an extra measurement would be necessary to acquire Rayleigh scatter with pTSFS.

The last step is to thermally stress the rIgG to assess the ability of ARMES to monitor protein structural changes and aggregation. EEM measurement was preferred for the next experiment as protein structural changes are mainly followed by Trp emission, and the Rayleigh scatter can provide useful information regards changes in particle size, and aggregation.

5. Monitoring Structural Stability and Aggregation of rIgG Using ARMES – EEM

The previous chapters provided a starting point from which we can explore the use of pMDF and PARAFAC for further investigations into physically stressed rIgG. PARAFAC might provide a different approach to follow protein structural changes via intrinsic fluorescence emission measurements. Here, thermal stress was selected because it provides large and well-known structural changes, and leads to significant protein aggregation at high temperatures.[17, 42, 72, 97, 133-135] This chapter assesses the applicability of pEEM/ARMES for monitoring structural stability and aggregation of rIgG.

5.1 Large structural changes favor the use of pEEM

IgG type proteins have two unfolding temperatures,[12, 133, 134, 136, 137] the first at ~60/65°C being reversible, followed by an irreversible unfolding at ~70/75°C[zz] which may be accompanied, and/or followed by protein aggregation. No major structural changes are expected up to the first unfolding temperature, limiting PARAFAC to identifying subtle structural changes, such as changes in Tyr-to-Trp FRET process.[aaa] Finally, once the first unfolding temperature is achieved, protein structure starts to change, which should facilitate the recovery of different Trp fluorophores for the first time for complex proteins using pMDF/PARAFAC.

Even though Rayleigh scatter can affect the pMDF data at the emission blue edge, it can be useful to monitor large structural changes, changes in particle size, and aggregation.[56, 62, 63] The use of pEEM can potentially provide information about structural changes and protein aggregation with a single measurement. The use of pTSFS measurements would require an extra data acquisition step to measure changes in Rayleigh scatter, increasing the time required for data collection, and analysis.

[zz] The exact unfolding temperature is highly dependent on IgG structure, buffer composition and pH.[12, 101, 133, 138, 139] For the rIgG used in this study, visible aggregates were formed at 75°C (Figure A-8), which was followed by the formation of a gel-like solution upon cooling down to 20°C.

[aaa] Protein relaxation precedes any significant structural changes, thus, up to the first unfolding temperature, we should be able to see an increase in Tyr and a decrease in Trp signal, caused by a change in Tyr-to-Trp FRET, and an average of all Trp in the native state.

Here, we explore the use of pEEM to try to monitor rIgG structural stability and aggregation caused by thermal stress from 20° to 70°C.

5.2 Data pre-processing

The pEEM pre-processing method had to be adapted to include the instrumental factor correction (Figures 2.1/2.5).[bbb] First, pEEM spectra were corrected for instrumental factor (excitation and emission wavelengths), followed by IFE correction, Rayleigh[ccc] and Raman scatter removal, interpolation, and smoothing to remove residual scatter/shot noise (Figure 5.1).[91] Even though the protein was thermally stressed up to 75°C (Section 2.3.1), the increased optical density (Figure A-8) limited IFE correction up to 70°C, restricting PARAFAC analysis to this temperature point.

Pre-processed $pEEM_{\parallel}$, $pEEM_{\perp}$, and unpolarized EEM_T were arranged in a three-dimensional structure (33 samples × 78 λ_{em} × 41 λ_{ex}), normalized by peak maximum,[83] and subject to PARAFAC analysis. PARAFAC models for the different polarizations were evaluated to identify if different fluorophores could be resolved for the thermally stressed rIgG, and if there were differences in the recovered loadings for each polarization.

[bbb] The use of instrumental factor correction caused a significant change in PARAFAC resolution for the rIgG thermally stressed (Tables A-7/A-8).

[ccc] For the $pEEM_{\parallel}$, Rayleigh scatter saturated (I_F>1000) at ≥70°C.

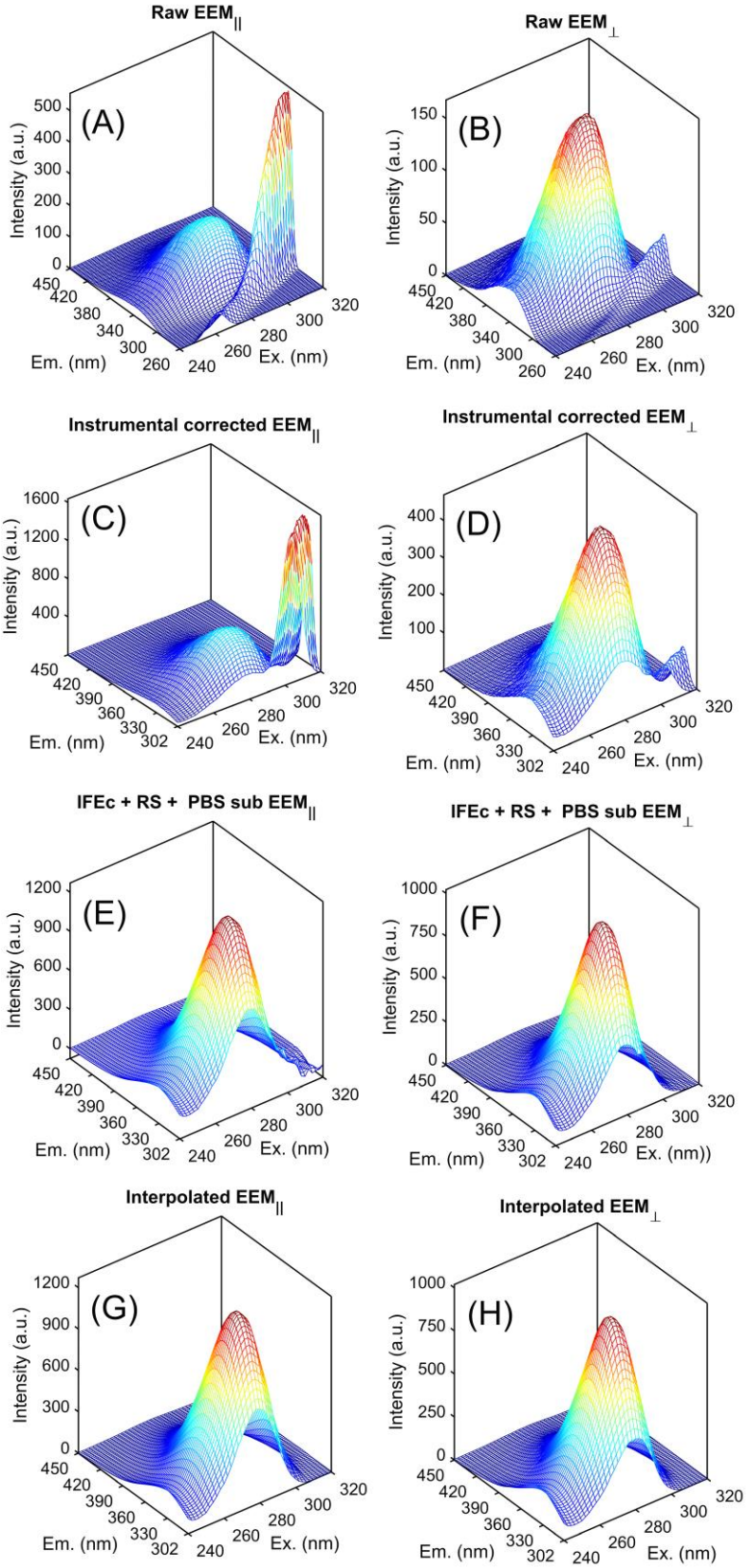

Figure 5.1: (Left) pEEM∥ and (right) pEEM⊥ pre-processing steps for the rIgG measured at 20°C. (A/B) Raw pEEM spectra showing the difference in Rayleigh scatter between pEEM∥ and pEEM⊥. (C/D) pEEM spectra corrected for instrumental response. (E/F) pEEM spectra were corrected for IFE using the UV-vis absorbance spectra of each replicate sample measured at 20°C, followed by Rayleigh scatter removal, buffer subtraction, and smoothing. (G/H) pEEM spectra were interpolated to remove residual scatter.

5.3 2D spectral analysis

A combination of 2D spectral analysis were used to quickly evaluate structural changes caused by the thermal stress. Figure 5.2 indicates that significant structural changes were only observed at higher temperatures (≥65°C), which was preceeded by a change in Tyr-to-Trp FRET at lower temperatures (<65°C).

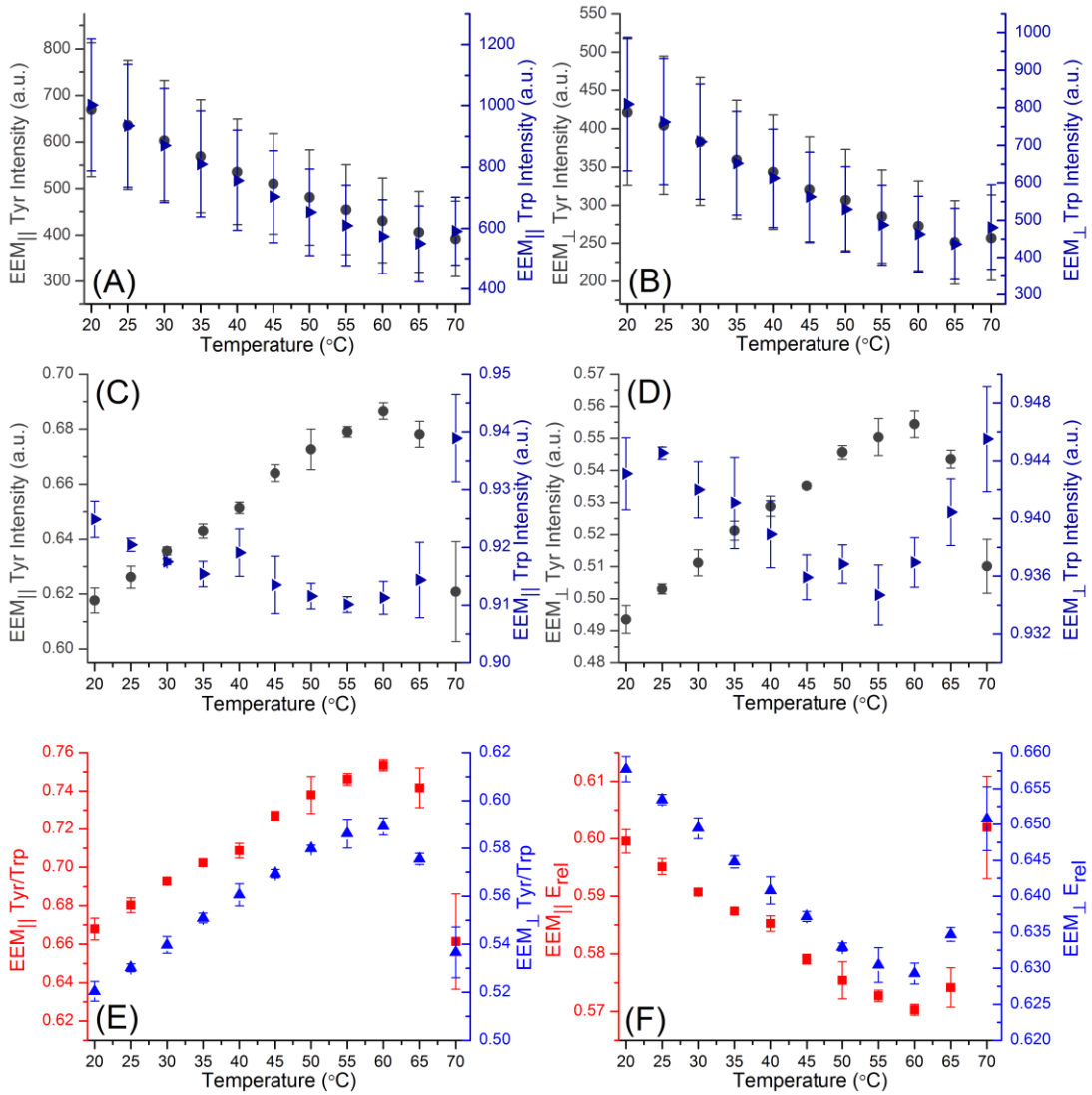

Figure 5.2: There was a decrease in Tyr intensity ($\lambda_{ex}/\lambda_{em}$=276/306 nm) for the rIgG measured from 20−70°C with both (A) pEEM∥, and (B) pEEM⊥. Trp intensity ($\lambda_{ex}/\lambda_{em}$=276/340 nm) decreased up to 60/65°C, increasing at 70°C, due to thermal quenching. Normalized (C) pEEM∥ and (D) pEEM⊥ showed the effects of Tyr-to-Trp FRET process. (E) The Tyr/Trp ratio increased from 20−60°C for pEEM∥ and pEEM⊥, decreasing at 65/70°C. (F) E_{rel} decreased from 20−60°C, increasing at 65/70°C. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

The initial increase in temperature was not enough to cause any significant structural change (Figure 5.2C/D). A change in Trp and Tyr trends was observed when the temperature was increased above 55°C, a sudden change in Tyr/Trp ratio (Figure

5.2E), and relative FRET efficiency (Figure 5.2F) was observed at 65/70°C, indicating a change in protein structure.

Here, it was clear that the use of single excitation/emission wavelengths were a very useful tool to quickly assess large spectral changes. However, the use of a simple 2D spectral analysis did not provide much information regarding small structural changes prior to protein unfolding. Thus, the use of chemometrics is still necessary to analyse the full emission space to try to resolve more details about the structural changes.

5.4 *Aniso*-EEM maps

The analysis of changes in *aniso*-MDF maps was proven to be an useful tool for measuring protein structural and concentration changes in simple proteins, such as BSA and HSA.[1, 3] However, it was clear that subtle structural changes in the rIgG native structure,[91, 92] along with the large number of fluorophores involved, reduced the utility of the *aniso*-MDF maps (Chapters 3/4). Here, we expected that large structural changes caused by the thermal stress would cause more significant changes in the *aniso*-maps.

Visual inspection of the *aniso*-EEM maps (Figure 5.3) showed variations over the full emission space, indicating changes in FRET process, fluorophore location within protein, fluorophore environment,[1] and residual scatter/shot noise as the temperature was increased. There were no significant changes in anisotropy for Tyr (Figure 5.3G) and Trp (Figure 5.3H) emission up to 60/65°C, and even though those changes were not statistically significant, it could indicate the first significant structural changes. Changes in anisotropy at 70°C were likely related to changes in FRET,[22] residual Rayleigh scatter/shot noise, and poor SNR.[ddd]

[ddd] A plausible explanation for the decrease in anisotropy is based on the Rayleigh scatter influence on the pEEM spectra. For the pEEM∥ measurements, the Rayleigh scatter was too strong, and the detector saturated under the instrument settings used, decreasing anisotropy in the affected area.

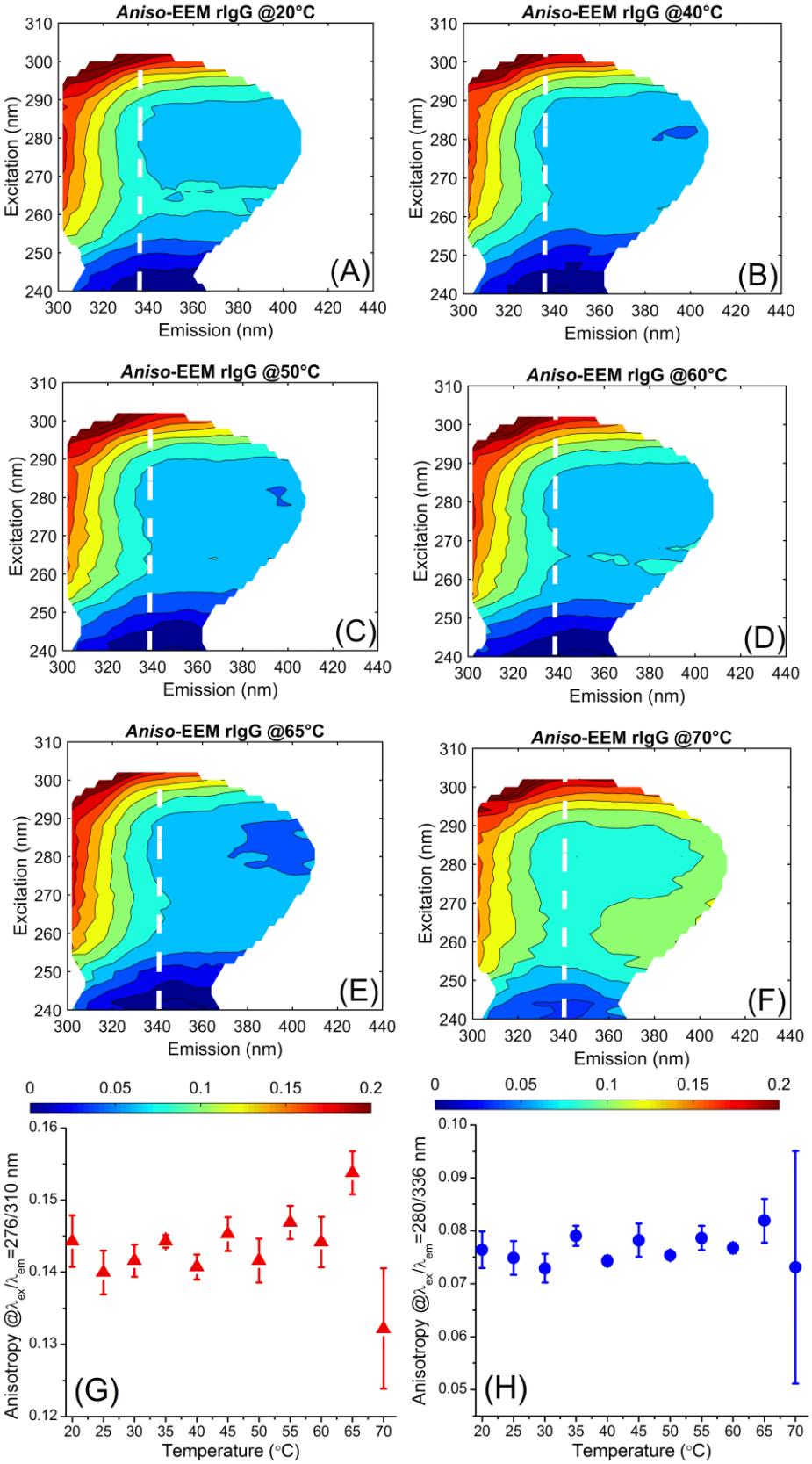

Figure 5.3: *Aniso*-EEM maps for rIgG collected at (A) 20°C, (B) 40°C, (C) 50°C, (D) 60°C, (E) 65°C and (F) 70°C. The white dashed lines mark λ_{em}=336 nm. The colour bars on the bottom represent the anisotropy scale. (G) Tyr anisotropy at $\lambda_{ex}/\lambda_{em}$=276/310 nm, and (H) Trp anisotropy at $\lambda_{ex}/\lambda_{em}$=280/336 nm for the triplicate EEM measurements of thermally stressed (20°−70°C) rIgG. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

SimI[140] was used to further assess protein structural changes observed with the *aniso*-EEM maps (Figure 5.4). SimI from the full emission space showed a significant change in *aniso*-EEM maps with the increase in temperature (RSD=29.85%). However, once we separate the Tyr and Trp regions, it was clear that these changes were mostly related to Trp emission (Figure 5.4B). It is well known that Tyr emission does not change with the increase in temperature,[141] thus, it was not a surprise that there were no significant changes in the *aniso*-EEM maps in this region (RSD=5.18%). On the other hand, changes in SimI for the Trp region (RSD=13.67%) suggested that the thermal stress induced structural changes.

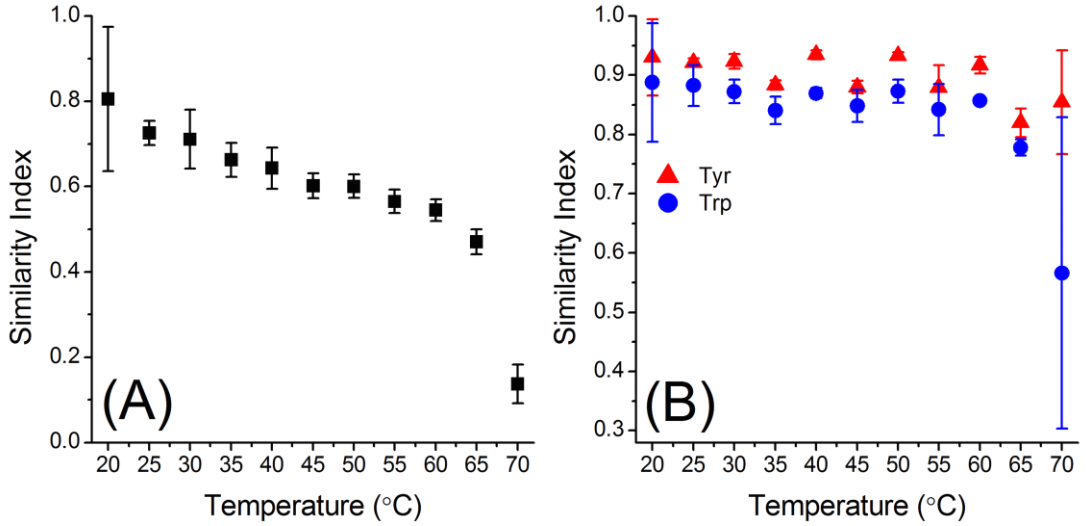

Figure 5.4: Plots of SimI values calculated for the various *aniso*-EEM maps over the 20−70°C temperature range. In each case, the reference spectrum was the EEM collected at 20°C from the first replicate sample (R1). (A) Full emission space; (B) Tyr ($\lambda_{ex}/\lambda_{em}$=270−290/302−320 nm), and Trp emission region ($\lambda_{ex}/\lambda_{em}$=280−300/320−360 nm). Error bars were generated from the standard deviation from triplicate measurements of the independent samples.

Although changes in the Tyr region were present, these were very small and strongly affected by residual scattered light. The presence of large structural changes in the thermally stressed rIgG at higher temperatures indicated that the utility of this method for the analysis of complex protein, such as rIgG, was mainly restricted to the Trp region.

5.5 PARAFAC analysis

One critical factor that restricted the effective application of PARAFAC to resolve individual Tyr and Trp fluorophores contribution from rIgG emission measured with pMDF was the lack of significant structural change.[91, 92] Here, this critical factor was

overcome by the thermal stress, which should facilitate the resolution of more than two PARAFAC components.

Based on the initial analysis (Sections 5.3/5.4), we expected to resolve at least three/four components with PARAFAC: *(i)* one or two components resolving a pure Trp and Tyr contribution reflecting changes in Tyr-to-Trp FRET; *(ii)* an average of all Trp fluorophores in the native state; and *(iii)* a change in the average Trp fluorophores as protein unfolds. However, the presence of Tyr-to-Trp FRET, residual Rayleigh scatter, shot noise, poor SNR, and the low number of samples might still limit PARAFAC resolution.[83] Ideally, PARAFAC would be able to extract all the information regards protein structural changes from the full 20−70°C temperature range with a single model. However, PARAFAC could not handle different processes[eee] at once, and only two valid components were recovered (Table 5.1, Figure 5.5E, and Figure A-9).

Table 5.1: Comparison of the two-, three-, and four-component PARAFAC models obtained for the normalized polarized $EEM_{\parallel}$, $EEM_{\perp}$, and unpolarized EEM_{T} of the data measured over the 20−70°C temperature range.

	$EEM_{\parallel}$	$EEM_{\perp}$	EEM_{T}	$EEM_{\parallel}$	$EEM_{\perp}$	EEM_{T}	$EEM_{\parallel}$	$EEM_{\perp}$	EEM_{T}
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 338	280/ 340	280/ 340	280/ 334	280/ 338	280/ 338	278/ 338	278/ 340	278/ 334
C1 Fit model (%)	55.70	43.31	53.66	90.87	65.94	87.29	51.49	83.05	92.86
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 306	278/ 324	278/ 318	280/ 302	278/ 304	278/ 302	278/ 308	278/ 306	292/ 336
C2 Fit model (%)	44.30	56.69	46.34	8.18	14.55	9.79	44.45	14.12	3.10
C3 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	280/ 344	280/ 340	280/ 342	292/ 332	280/ 344	280/ 302
C3 Fit model (%)	–	–	–	0.94	19.51	2.93	2.22	2.43	2.30
C4 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	–	–	–	280/ 352	294/ 334	284/ 344
C4 Fit model (%)	–	–	–	–	–	–	1.84	0.41	1.73
Variance explained (%)	99.97	99.98	99.98	99.99	99.99	99.99	99.99	99.99	99.99
CONCORDIA (%)	99.75	99.55	99.58	−222.82	−7148	−1476	−5604	−32687	−16648
Split-half analysis (%)	58.80	71.25	59.83	11.00	70.22	29.62	17.65	21.62	6.36

[eee] Thermal quenching, changes in Tyr-to-Trp FRET process, and changes in Trp classes.

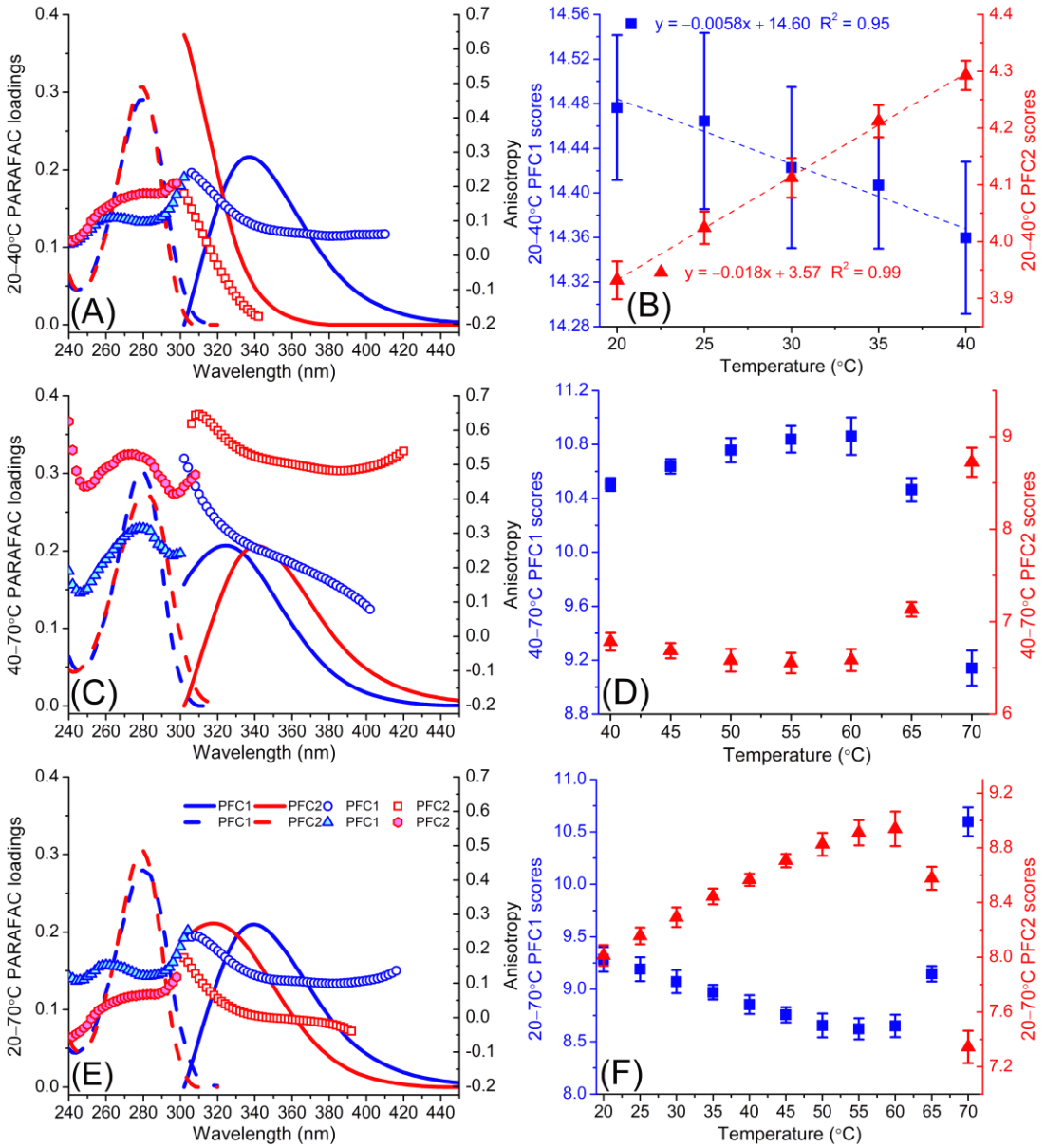

Figure 5.5: Comparison of the two-component PARAFAC (left) loadings and (right) scores for the pre-processed EEM_T data over the (A/B) 20−40°C, (C/D) 40−70°C, and (E/F) 20−70°C temperature range. The anisotropy values were calculated for each component resolved for each temperature range (Table 5.2). Error bars generated from the standard deviation from triplicate measurements of the independent samples.

For the full 20−70°C temperature range, the two-component PARAFAC models (Figure 5.5E), PFC1 resolved an average of all Trp fluorophores in the rIgG structure, with a change in PARAFAC scores trend at 60°C (Figure 5.5F and Figure A-10).[fff] While PFC1 was virtually identical for both pEEM and unpolarized EEM, PFC2 showed some differences between $pEEM_{\parallel}$ and $pEEM_{\perp}$ (Table 5.1, Figure A-9A/D/G). PFC2 resolved an average of all Tyr and Trp fluorophores within the rIgG structure,

[fff] There were no significant changes in Trp maximum emission (Figure A-11) with the increase in temperature. However, a lack of red and/or blued shift in Trp emission does not mean a lack of changes in Trp environment,[22, 34, 47] both phenomena could occur at the same time, resulting in the same averaged Trp emission.

with the parallel polarization resolving a more significant contribution from Tyr emission.[ggg] Three-, and four-component PARAFAC models were not validated by CONCORDIA, and/or split-half analysis, however, one must consider that CONCORDIA is not always appropriate to assess model validity for real-world non-ideal data.[83, 113, 142] Regardless of model validation, three-, and four-component PARAFAC models for the polarized, and unpolarized EEM data were reasonable from the photophysical point of view (Figures A-9/A-10). Thus, it indicated that there was more information available than what a two-component PARAFAC model could extract from the full 20−70°C temperature range.

From the PARAFAC scores plot (Figure 5.5F), we were able to identify two distinct process affecting rIgG structure between the 20−70°C temperature range. The first process was related to changes in Tyr-to-Trp FRET, and the second to changes in protein structure (Trp emission). Thus, the next step was to divide the thermal stress data into two temperature ranges (20−40°C and 40−70°C) in an attempt to separate these two distinct processes, and facilitate PARAFAC to extract more information regards protein structural changes (Table 5.2, Figure 5.5). The main reason for the selection of 40°C as the separation point was because we would still be able to resolve some Tyr contribution at the higher temperature range and remove significant changes in Trp fluorophores at the lower temperature range.

For the 20−40°C temperature range, a two-component PARAFAC model managed to separate the Trp (PFC1) and Tyr (PFC2) contributions (Figure 5.5A). The resolution of these two components were facilitated by the changes in Tyr-to-Trp FRET process, which was confirmed by the scores plot (Figure 5.5B). Although PARAFAC loadings and scores suggested the successful separation of an average of all the Trp and Tyr fluorophores presents in the FRET process, the model could not be validated with CONCORDIA. The low CONCORDIA values were not a surprise,[hhh] and the model was validated based on the split-half analysis.[143] Small differences were observed in model validation parameters, loadings, and scores for PFC2 between

[ggg] Although the maximum emission (306 nm) would suggest that PFC2 for the pEEM$_{\parallel}$ resolved a pure Tyr contribution, the emission was extended to longer wavelengths (>360 nm), indicating that a spectral contribution from Trp fluorophores. The blue shift in PFC2 emission for pEEM$_{\parallel}$ (306 nm) in comparison with the pEEM$_{\perp}$ (324 nm) polarization were related to the polarization itself, as pEEM$_{\perp}$ is less affected by residual scatter.

[hhh] They were, in fact, expected, as the FRET process breaks the trilinearity, and variability requirements for a good PARAFAC model.[85]

the pEEM∥ and pEEM⊥ data (Table 5.2, Figures A-12/A-13). The slightly higher scores and lower CONCORDIA values for PFC2 resolved for pEEM⊥ indicated that this polarization was more influenced by Trp emission.

Table 5.2: Comparison of the model parameters and components obtained for the two-component PARAFAC models obtained for the normalized polarized EEM∥, EEM⊥, and unpolarized EEM$_T$ data over the 20−40°C, 40−70°C, and 20−70°C temperature ranges.

	20−40°C			**40−70°C**			**20−70°C**		
	EEM∥	**EEM⊥**	**EEM$_T$**	**EEM∥**	**EEM⊥**	**EEM$_T$**	**EEM∥**	**EEM⊥**	**EEM$_T$**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 334	280/ 338	280/ 336	278/ 316	280/ 330	278/ 324	280/ 338	280/ 340	280/ 340
C1 Fit model (%)	92.41	90.71	92.47	61.46	84.21	68.85	55.70	43.31	53.66
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 302	278/ 304	280/ 302	280/ 340	288/ 342	280/ 342	280/ 306	278/ 324	278/ 318
C2 Fit model (%)	7.59	9.29	7.53	38.54	15.79	31.15	44.30	56.69	46.34
Variance explained (%)	99.99	99.99	99.99	99.97	99.98	99.98	99.97	99.98	99.98
CONCORDIA (%)	26.01	−56.96	−28.66	99.48	99.41	99.35	99.75	99.55	99.58
Split-half analysis (%)	94.50	99.95	99.69	69.98	88.71	59.54	58.80	71.25	59.83

It was hoped that selecting a temperature range between 40−70°C would allow PARAFAC to resolve Tyr emission and changes in Trp fluorophores, however, only a two-component PARAFAC model was valid (Table A-8, Figure 5.5C/D). Restricting the temperature analysed to the 40−70°C range reduced the influence of the Tyr emission from the Tyr-to-Trp FRET process, as indicated by the anisotropy loadings (Figure 5.5C).[iii] Here, PFC1 resolved the average of all Tyr and Trp fluorophores presents in the rIgG in its native structure. The selection of this temperature range increased the differences between polarized and non-polarized EEM datasets (Table 5.2, Figure A-12). For PFC1, there was a 14 nm blue shift from pEEM⊥ (λ_{em}=330 nm) to pEEM∥ (λ_{em}=316 nm), which was accompanied by a significant increase in scores for pEEM⊥ (Table 5.2, Figure A-13). Even though PFC2 emission loadings were virtually identical for the polarized and unpolarized EEM, there was a red shift (8 nm) in the excitation loadings for the pEEM⊥. This red shift was accompanied by a decrease in scores, indicating that PFC2 recovered for the pEEM⊥ was more influenced by changes in Trp fluorophores.

[iii] The shape of the anisotropy excitation loadings and the increased values for the anisotropy in the emission[36, 45, 65] loadings indicated that there was a lower influence of Tyr emission in the PARAFAC components resolved for the 40−70°C temperature range (Figure 5.5C) in comparison to that resolved for the full 20−70°C temperature range (Figure 5.5E).

5.6 Aggregation analysis using Rayleigh scatter modelled with PARAFAC

Simple 2D UV-vis[97, 99] (Table 2.1) and fluorescence aggregation index (F-AI)[144, 145] are not always suitable for the analysis of changes in particle size and aggregation in complex proteins, particularly at low concentration. Previously, these aggregation measurement tools did not provide much conclusive information regards protein structural changes, and a suitable alternative was to use the Rayleigh scatter collected simultaneously with the pEEM measurements (Figure 5.6).[56, 62]

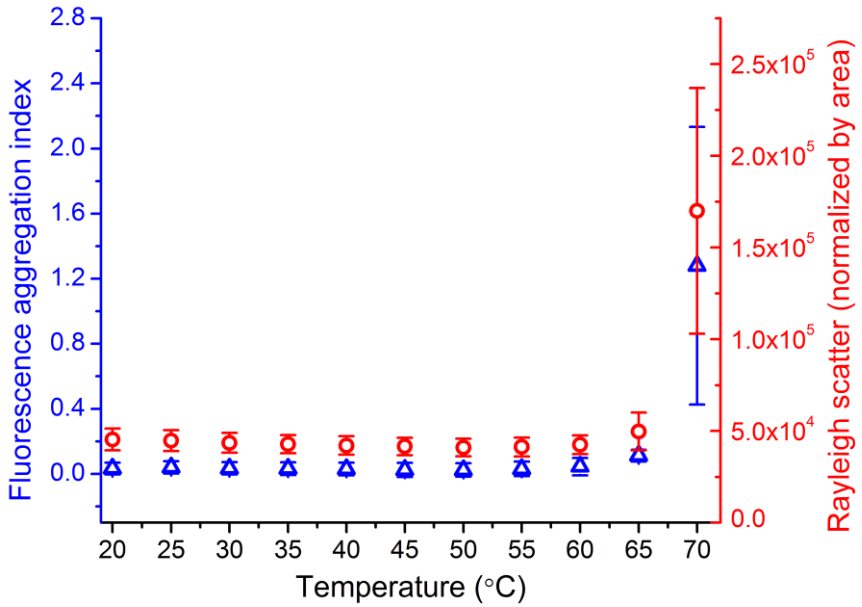

Figure 5.6: Comparison of the plots for the F-AI calculated at λ_{ex}=280 nm ($= \frac{\lambda_{em}280}{\lambda_{em}340}$), and the normalized (by area) first order Rayleigh scatter extracted from the pEEM‖ data over the 20−70°C temperature range. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

The first order Rayleigh scatter removed from the pEEM data[jjj] at the pre-processing stage (Figure A-3) was successfully used to follow changes in particle size for different forms of insulin,[62] and rIgG.[56] However, for the thermally stressed rIgG, there was no significant differences between the use of a 2D F-AI and the full Rayleigh scatter region (Figure 5.6). The same lack of change in particle size below 70°C for the thermally stressed IgG was previously observed with various techniques, such as differential scanning fluorimetry (DSF), static light scattering, intrinsic fluoresce,

[jjj] The whole Rayleigh scattering area was used to calculate the volume of Rayleigh scatter, improving the robustness of the analysis of changes in particle size and aggregation in comparison to the resolution from a single wavelength. The use of the whole Rayleigh scatter region reduces the noise associated with the value recovered from a single wavelength point.

HPLC-SEC, and dynamic light scattering (DLS).[17, 72, 133, 146, 147] These changes in particle size were restricted to temperatures ≥70°C, as a partial unfolding of the protein is a pre-requisite to initiate aggregation, thus, changes in particle size and aggregation are mostly negligible up to the second unfolding transition.[17, 72, 133, 146, 147]

The lack of change in particle size up to 65°C indicated that rIgG structure was stable up to the second unfolding temperature, meaning that it was difficult to cause a change in protein structure.[63] The use of Rayleigh scatter to follow changes in particle size and aggregation is less useful when following small protein changes.[56, 62]

5.7 Conclusions

In this chapter we showed that the use of ARMES combined with PARAFAC analysis can be used for assessing significant structural changes in thermally stressed rIgG. The 20−70°C temperature range was enough to cause changes in the average Trp class at higher temperatures; however, these were accompanied by changes in Tyr-to-Trp FRET process, which highly influenced PARAFAC resolution. The strong presence of Tyr-to-Trp FRET process, poor SNR, and low sample numbers was reflected in a limited PARAFAC resolution.

Independently of the temperature range selected, a two-component PARAFAC model was the only valid model recovered from the thermally stressed rIgG, which did not extract all the available information regards protein stability and structural changes. PARAFAC analysis of different temperature ranges indicate that, for the rIgG, the use of pEEM$_{\parallel}$ should be preferred for the analysis of changes in Tyr-to-Trp FRET process, and the pEEM$_{\perp}$ for analysing changes in Trp.

6. Conclusions

The aim of this project was to determine if the ARMES methodology could provide a better method for measuring changes in the structure of IgG type proteins in solution. This was the first time that pMDF measurements combined with chemometrics analysis were used to analyze such complex proteins. The main challenge was to solve the pre-processing issues to enable the photophysical characterization of IgG using multi-way decomposition methods, such as PARAFAC and MCR. This goal was achieved with two different pMDF measurements: pEEM[91] and pTSFS.[92]

The ARMES method was recently developed for the quantitative[3] and qualitative[1] analysis of proteins. These first studies showed that ARMES was a suitable technique to extract protein signal from a complex media, and to resolve individual fluorophores in simple proteins, such as BSA, HSA,[89] and insulin.[62] However, the use of thin film polarizers, which only transmitted light over 290 nm, prevented the accurate measurement of Tyr and Trp fluorophores. The use of UV transparent wire grid polarizers[57] increased the emission space collected with pMDF, theoretically making it possible to extract Tyr and Trp emission, as the method was validated using a mixture of fluorophores,[88] where FRET and IFE were not present. PARAFAC analysis was selected as the best multi-way decomposition method for the analysis of simple proteins and non-interacting mixtures,[57, 62] we had to be further developed the use of pMDF/ARMES/PARAFAC to suit complex proteins.

The pre-processing of the pEEM and pTSFS data prior to PARAFAC analysis was critical and complex. EEM and TSFS data are intrinsically different (Figure 2.4), requiring different handling to achieve the trilinearity required for PARAFAC analysis.[71, 73, 79, 83, 85] This was the first time that an IgG protein was measured with pMDF and analyzed with PARAFAC, requiring an extensive amount of time and effort to find the ideal pre-processing method. The most complex issues to be solved prior to PARAFAC analysis were the non-trilinearity from Rayleigh and Raman scatter, IFE, and the TSFS structure.

For the pEEM data, independently of the pre-processing method applied, residual Rayleigh scatter, and artifacts introduced by IFE correction caused significant

problems, making PARAFAC results less trustworthy in some cases. Based on PARAFAC resolution, we could conclude that it was better to first implement subtraction-based corrections (*e.g.* buffer subtraction, Rayleigh scatter removal) before any multiplication-based corrections (*e.g.* IFE correction). The use of interpolation in a restricted small spectral region did not have a major impact on components recovery, improving some aspects of PARAFAC models.

We demonstrated that pTSFS is a better measurement method than pEEM for PARAFAC analysis of intrinsic protein fluorescence, particularly for the weak Tyr fluorescence at the emission blue edge. The use of pTSFS can minimize the collection of Rayleigh scatter which caused a lot of issues at the emission blue edge in the pEEM spectra, mostly because of the complexity in removing Rayleigh shot noise.[91] A simple transformation of the pTSFS data into an EEM like layout (t-pEEM) was able to provide the trilinear structure[73] required for PARAFAC analysis. However, this transformation into t-pEEM produced an ASDR (<8%) at the emission blue edge, which was critical for resolving Tyr emission. The assessment of three pre-processing methods to deal with the ASDR problem, missing data, use of zeros, and interpolation, showed that PARAFAC resolution was highly method dependent. In each case PARAFAC separated Tyr from Trp emission, thus, confirming the superior performance of pTSFS over pEEM measurements for subtle structural changes.

For complex proteins, where multiple Tyr and Trp fluorophores are present and interacting with each other, the use of ARMES/PARAFAC could not resolve individual fluorophore populations. Yet, it was possible to photophysically characterize the rIgG, which was used as a baseline for monitoring structural changes. PARAFAC analysis of physically stressed rIgG showed significant differences in the recovered components scores and this provides a different insight into fluorescence emission changes induced by stressing conditions. However, the same trend was observed with the 2D spectral analysis (Figure 5.2) and the scores plots extracted with PARAFAC (Figures 5.5, Figure A-10). The use of pMDF/PARAFAC/ARMES was advantageous for monitoring subtle structural changes, however, a simple 2D spectral analysis was enough to spot large structural changes. The main advantage of collecting pEEM was the possibility to isolate the first order Rayleigh scatter, which was used to monitor changes in particle size and aggregation. Although the use of Rayleigh scatter was previously shown to be useful to monitor changes in particle size

for insulin type proteins,[62] it was not very useful to monitor small changes in rIgG particle size,[56, 63] which was in agreement with the literature for IgG type proteins.[17, 72, 133, 146, 147] Thus, the main advantage of the pEEM measurements for the analysis of stressed IgG was that it gave the possibility of following changes in protein structure, particle size, and aggregation, with a single measurement.[56, 62]

Even though the advantage of the pTSFS over pEEM measurements was to reduce Rayleigh scattering, there was still some residual scattered light in the pTSFS data, which was not a surprise, given the 10 nm slit widths used. In addition, the poor SNR, combined with the low sample number used here, limited PARAFAC resolution, especially at the emission blue edge. Ultimately, this indicates that to extract true and uncontaminated Tyr emission components, and thus accurate scores, one needs to use narrower slits to minimise scatter, which will decrease signal intensity that can only be recovered by increasing acquisition time and PMT voltage. To reduce shot and dark noise, and thus improve SNR, one has to collect multiple spectra from the same sample, and then use the average spectra for modelling, which would also increase acquisition time. However, this is not practical currently for ARMES using conventional scanning spectrometers where four measurements are required (and particularly for rapid or high throughput analytical applications).

The potential use of ARMES/PARAFAC for quality control of IgG is still limited, requiring an improvement in SNR, and a higher number of samples to improve model resolution, which are prohibitive by the time required for collecting a full 3D pMDF spectra with the Agilent Cary Eclipse spectrophotometer found in the laboratory.[62, 89] There are two alternatives that could solve the data quality issues: reduce the time necessary for data collection using the same Cary Eclipse, or use a faster machine that could collect the same amount of data in a reduced time. In addition, future studies must be done using a large amount monoclonal antibody, which should be prepared at once, avoiding any differences that could be related to sample composition, preparation, and manipulation.

The hardest part was to develop the best way of pre-processing the pMDF data to allow the use of trustworthy chemometric analysis. The data quality issue can be resolved with simple changes to the data collection using the Cary Eclipse. Improving data quality will significantly improve chemometrics. The collection time issue could

be addressed by the collection of a simpler 2D measurement. Based on PARAFAC resolution from the pMDF data, the 2D spectra could be collect at λ_{ex} 280 and 296 nm. In addition, there was not much useful information regards structural change in the emission blue edge. Restricting the emission collect to the Trp region will make it possible to collect more samples in a much-reduced period. Collecting more data in a reduced amount of time means that more samples can be collected and averaged, improving SNR and data quality. Further improvements could be achieved by the development of an automatized software to allow the collection of multiple temperature/time points. Increasing the number of data points can facilitate the resolution of subtle structural changes. This software would have to be combined with the use of a single polarization, preferentially the parallel polarization, as it does not require G-factor correction.

An alternative would require the acquisition of a new system, such as the Horiba Acqualog. While the Agilent Cary Eclipse has a single detector channel, the Acqualog system has a CCD fluorescence emission detector, which makes it up to 100 times faster than the current system. Another advantage of the Acqualog is that this is the only spectrophotometer available in the market that can collect UV-vis absorbance and EEM spectra at the same time. The Acqualog software automatically corrects for IFE, Rayleigh, and Raman scatter, reducing the pre-processing complexity. The data collected with the Horiba Acqualog has a continuous sample acquisition, meaning that more samples and data points could be collected in a quicker manner, resulting in a much better data quality than that used in this thesis.

7. Appendices

7.1 Instrumental correction factor

The instrumental correction factor was collected with five different standards (Table A-1), for six different machines at the same time. As the standards have to be quickly measured, it required that nearly the whole NBL group worked together. I was responsible for the sample preparation, the collection of the standards was a group task, and data analysis was carried out by Dr Saioa Elcoroaristizabal. A paper is being prepared on the instrumental correction:

Evaluation of Instrumental Correction in Fluorescence Excitation, Emission and Synchronous Spectra. S. Elcoroaristizabal, D. Melnikau, Y. Casamayou-Boucau, M. Steiner-Browne, A.L. de Faria e Silva, B. Boateng, M. Kyne, F. Gordon, S. Henry, and A.G. Ryder.

Table A-1: Fluorescence excitation wavelength and emission range collected for each dye (and the respective blank).

Dye	BAM certified	λex	Emission range
Fluorescence Spectral Standard A	BAM-F001	280	300−450
Fluorescence Spectral Standard B'	BAM-F002a	315	330−530
Fluorescence Spectral Standard C	BAM-F003	380	390−600
Fluorescence Spectral Standard D	BAM-F004	420	450−700
Fluorescence Spectral Standard E	BAM-F005	550	570−770

7.2 Inner filter effect (IFE) correction

UV-vis absorbance spectra of the thermally stressed rIgG were only measured at the beginning and at the end of the thermal stress experiments. The only way of collecting the correspondent UV-vis absorbance spectra for the sample being stressed at that the same moment would be to have a machine capable of collection pMDF, and UV-vis absorbance spectra at the same time. Unfortunately, the Cary Eclipse could not measure UV-vis absorbance spectra, and there was no such machine available in the laboratory for these experiments. The only solution was to collect the UV-vis absorbance spectra for all the temperatures points with another aliquot in a different day. It is important to keep in mind that we were working with polyclonal rIgG and there was no guarantee that different aliquots from the same protein solution would behave in the same manner. Differences could be cause by the freeze-thaw cycle (the samples were stored at −70°C), by the slightly different composition in the aliquot

itself (the rIgG used here was polyclonal), and by differences in light/time exposure. A simple 5 minute equilibration time was used before collecting each data point, in triplicate measurements, which were then blank subtracted, and averaged, as showed in Figure A-1A/B.

The difference in the maximum absorbance between the replicate samples (Figure A-1A, RSD ~0.8%) solutions was similar to the difference caused by the temperature increase itself (Figure A-2B, RSD ~1.1%). These differences were restricted to the third decimal point, not affecting IFE correction. The difference was more significative for the replicate solutions measured in different days (RSD ~55%) than for the different temperature points collected for the same sample (RSD ~22%) in longer wavelengths (>290 nm), indicating that there was a bigger change in scattered light between the different days than samples that caused by the thermal stress itself. Thus, because the Cary Eclipse did not allow the collection of MDF and UV-Vis absorbance spectra at the same time, IFE correction was done using the UV-Vis absorbance spectra for the exact replicate sample measured at 20°C.

Rayleigh scatter could be heavily overlapped to Trp fluorescence at the excitation red edge, overestimating IFE correction in that region. The use of an interpolation step in the pEEM blank spectra used to calculated LOR (Figure A-2) reduced the risk of an IFE overcorrection.

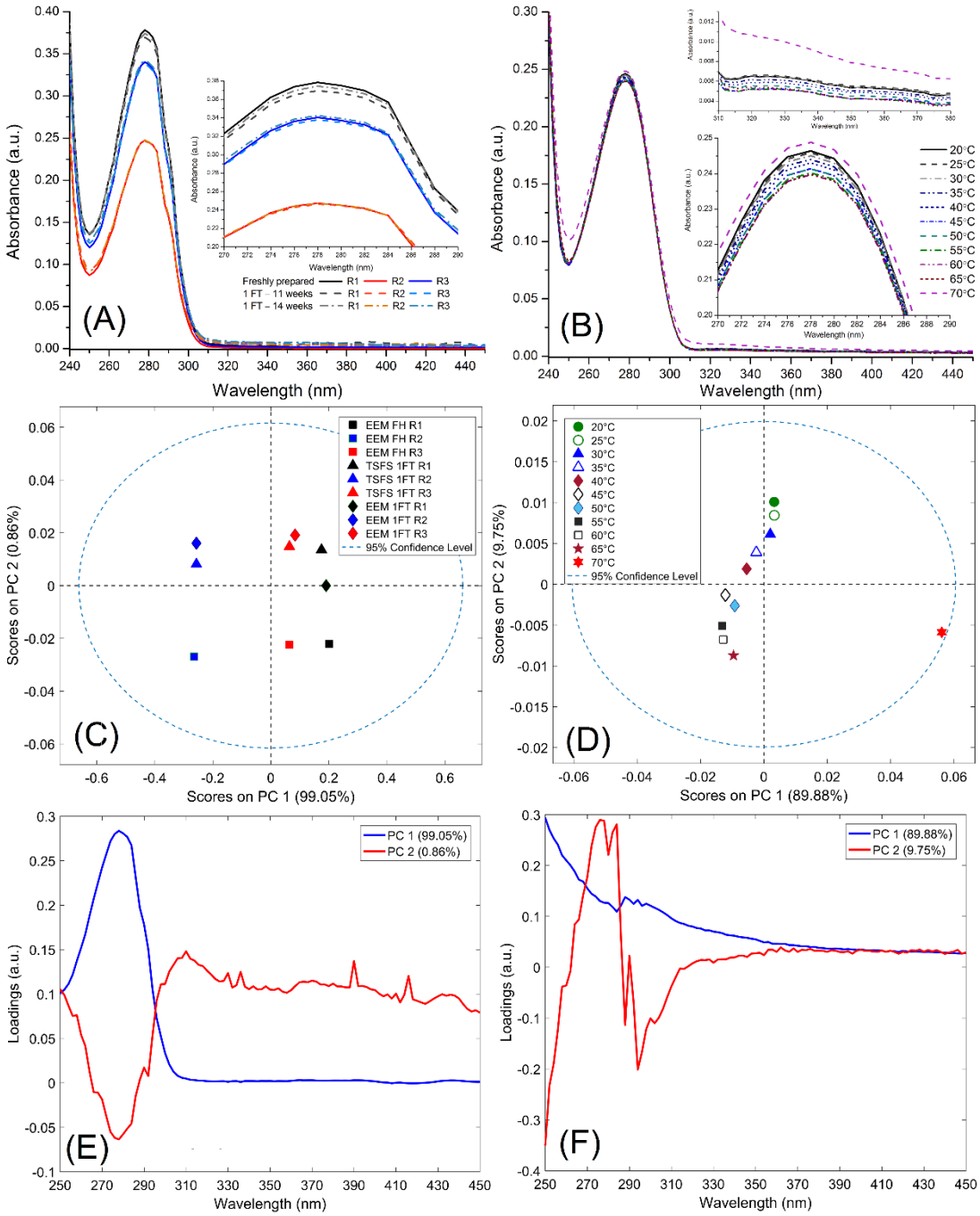

Figure A-1: (A) UV-vis spectra for the rIgG replicate samples (PBS buffer spectrum subtracted) used to calculate protein concentration: R1=1.3 mg mL^{-1}, R2=0.9 mg mL^{-1}, and R3=1.2 mg mL^{-1}. The full lines represent the sample measured with EEM (freshly prepared), the dashed lines with TSFS (11 weeks, −70°C) and the dashed dot lines with EEM (14 weeks, −70°C). Abs=0.31±0.05 (at λ=280 nm, in the 2 mm pathlength) of the rIgG solutions measured with TSFS, and Abs=0.32±0.05 for the samples collected with EEM. Scores plots and loadings for the two-component PCA model using the normalized (by maximum) and mean centered absorbance spectra collected from (C/E) 20°C for all the replicate measurements, and (D/F) 20° to 70°C (R2).

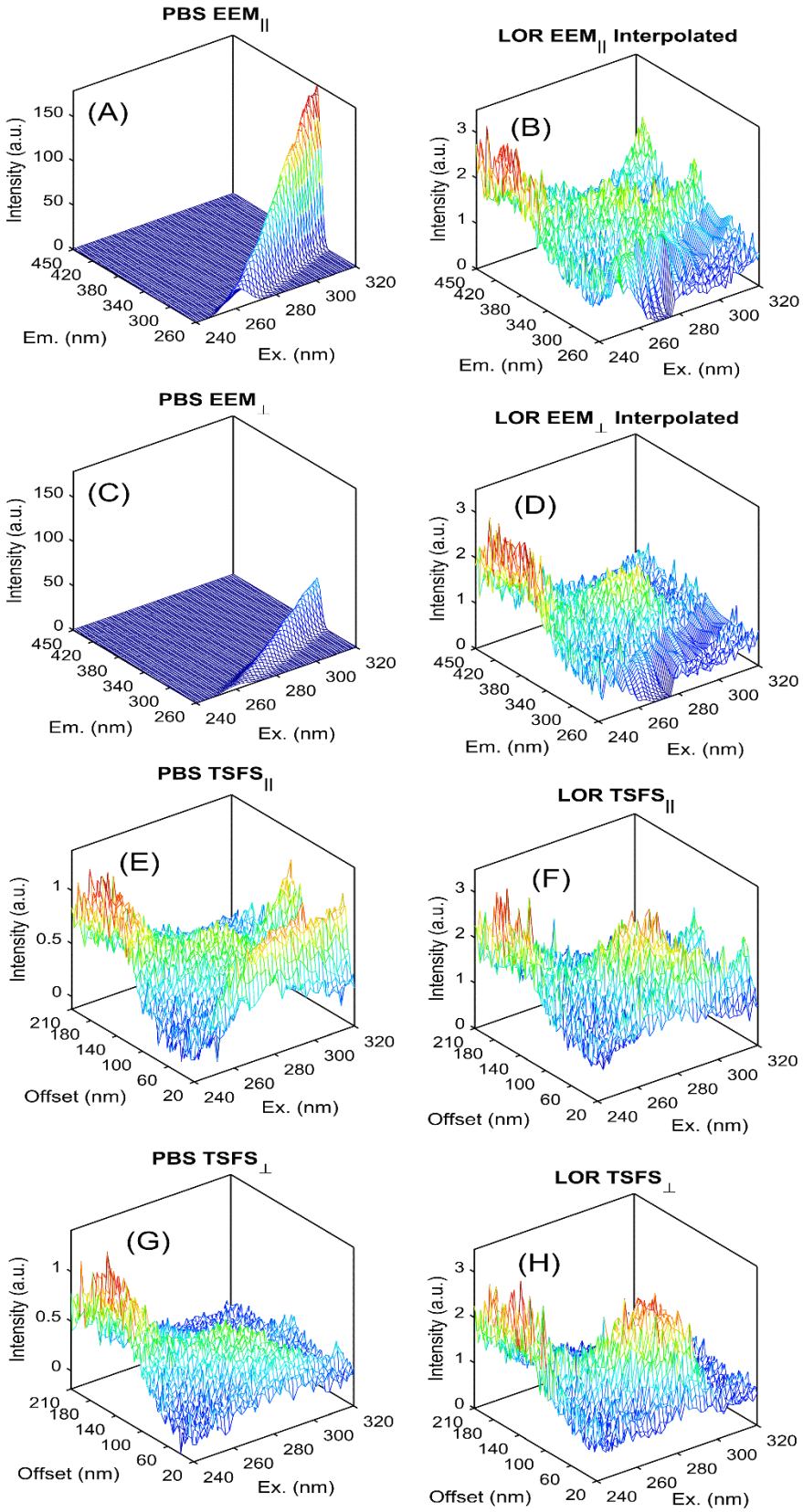

Figure A-2: Polarized (A) EEM∥, (C) EEM⊥, (E) TSFS∥, and (G) TSFS⊥ measurements of PBS solution at 20°C. Limit of reporting (LOR) calculated from the standard deviation of 10 replicate measurements of PBS buffer for the polarized (B) EEM∥, (D) EEM⊥, (F) TSFS∥, and (H) TSFS⊥. The pEEM Rayleigh scatter area was interpolated to remove Rayleigh scatter.

7.3 Rayleigh scatter removal

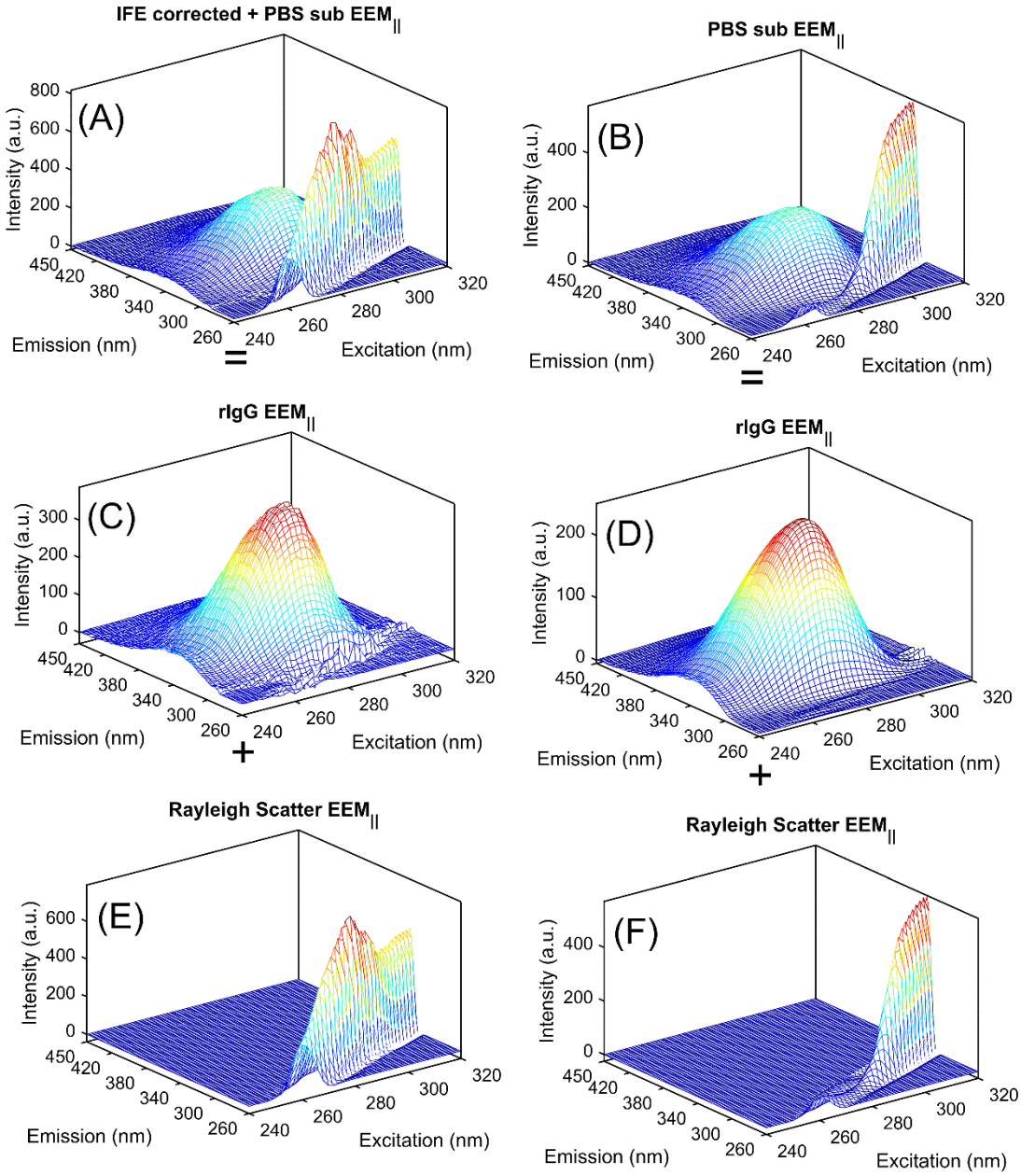

Figure A-3: Illustration of the Rayleigh scatter removal from the rIgG pEEM‖ data corrected with (A/C/E) Pre-processing 1, and (B/D/F) Pre-processing 2. (A/B) The data is modelled with PARAFAC to separate the (C/D) rIgG from the (E/F) Rayleigh scatter.

7.4 Savitzky-Golay smoothing functions

7.4.1 Old Savitzky-Golay smoothing function – pEEM native state

```
function smodata= smoothie(data,width,order)

[n,m,l]=size(data);


for i=1:n
    y=squeeze(data.data(i,:,:));

    [smoy,coeffi]= savgol(y,width,order,0);
    Smodi(i,:,:)=smoy;

end

data.data=Smodi;
smodata=data;

end
```

7.4.2 New Savitzky-Golay smoothing function – pTSFS native state, pEEM thermal stress

```
function smodata= smoothie(data,width,order)

[n,m,l]=size(data);
opt=savgol('options');
opt.tails= 'weighted';
opt.mode={1};

for i=1:n
    y=squeeze(data.data(i,:,:));

    [smoy,coeffi]= savgol(y,width,order,0);
    Smodi(i,:,:)=smoy;

end

data.data=Smodi;
smodata=data;

end
```

Using this new smoothing function slightly improved PARAFAC model validation for the p-EEM data corrected with Pre-processing 2 and Pre-processing 2/Interpolated (Table 3.1, Table A-2). The new smoothing function distorted Pre-processing 1 and Pre-processing 1/Interpolated data due to the increased residual shot noise in these pre-processed data.

7.5 EEM rIgG native state

Table A-2: Comparison of the two-component PARAFAC model parameters and components obtained for the normalized polarized EEM$_{\parallel}$ of the rIgG native state with different pre-processing steps. These results were obtained with the new smoothing function.

	Not IFE corrected	IFE corrected not normalized	IFE corrected normalized	Instrumental + IFE correction
	EEM$_{\parallel}$	EEM$_{\parallel}$	EEM$_{\parallel}$	EEM$_{\parallel}$
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/336	278/336	278/336	280/330
C1 Fit model (%)	99.86	99.40	99.38	98.60
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	300/342	298/340	298/340	296/336
C2 Fit model (%)	0.14	0.60	0.62	1.40
Variance explained (%)	99.98	99.99	99.99	99.99
CONCORDIA (%)	99.02	96.13	98.23	98.54
Split-half analysis (%)	99.69	98.40	99.92	99.75

Table A-3: Selection of number of components using the normalized polarized EEM$_{\parallel}$ modeled by PARAFAC for the different preprocessing methods.

	Pre-proc1			Pre-proc1/Interp.			Pre-proc2			Pre-proc2/Interp.		
Comps.	**4**	**3**	**2**	**4**	**3**	**2**	**4**	**3**	**2**	**4**	**3**	**2**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	278/ 338	280/ 332	280/ 336	278/ 338	280/ 336	280/ 336	280/ 338	282/ 338	280/ 336	278/ 338	280/ 336	278/ 336
C1 Fit model (%)	90.42	79.19	99.73	72.80	71.98	98.74	92.91	50.95	99.75	59.39	62.02	99.34
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 308	280/ 338	296/ 342	280/ 304	280/ 324	298/ 336	280/ 308	280/ 328	296/ 342	280/ 308	280/ 328	298/ 338
C2 Fit model (%)	6.87	20.21	0.27	22.25	26.89	1.26	6.03	48.83	0.12	24.95	37.76	0.66
C3 $\lambda_{ex}/\lambda_{em}$ (nm)	294/ 338	296/ 342	–	294/ 338	298/ 336	–	294/ 342	298/ 344	–	286/ 336	298/ 342	–
C3 Fit model (%)	2.43	0.60	–	3.67	1.13	–	0.90	0.22	–	14.60	0.22	–
C4 $\lambda_{ex}/\lambda_{em}$ (nm)	296/ 338	–	–	298/ 338	–	–	296/ 338	–	–	298/ 338	–	–
C4 Fit model (%)	0.27	–	–	1.27	–	–	0.16	–	–	0.56	–	–
Variance explained (%)	99.98	99.98	99.97	99.99	99.99	99.98	99.99	99.99	99.98	99.99	99.99	99.98
CONCOR-DIA (%)	-1.3×10^6	−76.58	95.22	-2.4×10^4	−280.77	96.92	-6.8×10^4	25.81	96.84	-6.8×10^4	−125.48	57.96
Split-half analysis (%)	62.54	96.02	99.13	15.44	97.66	98.39	96.31	97.55	99.14	51.41	90.08	99.57

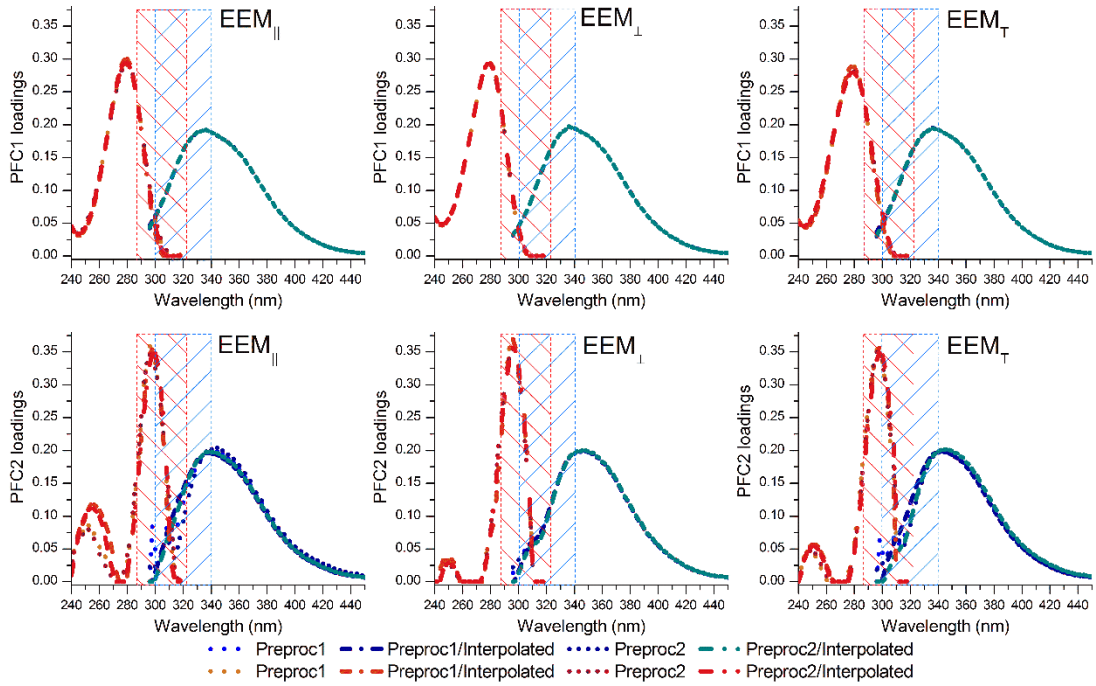

Figure A-4: Comparison of the two-component PARAFAC models for the polarize EEM∥, EEM⊥, and unpolarized EEM$_T$ rIgG data corrected with the different pre-processing methods. The shaded boxes represent the EEM spectral regions most affected by scatter.

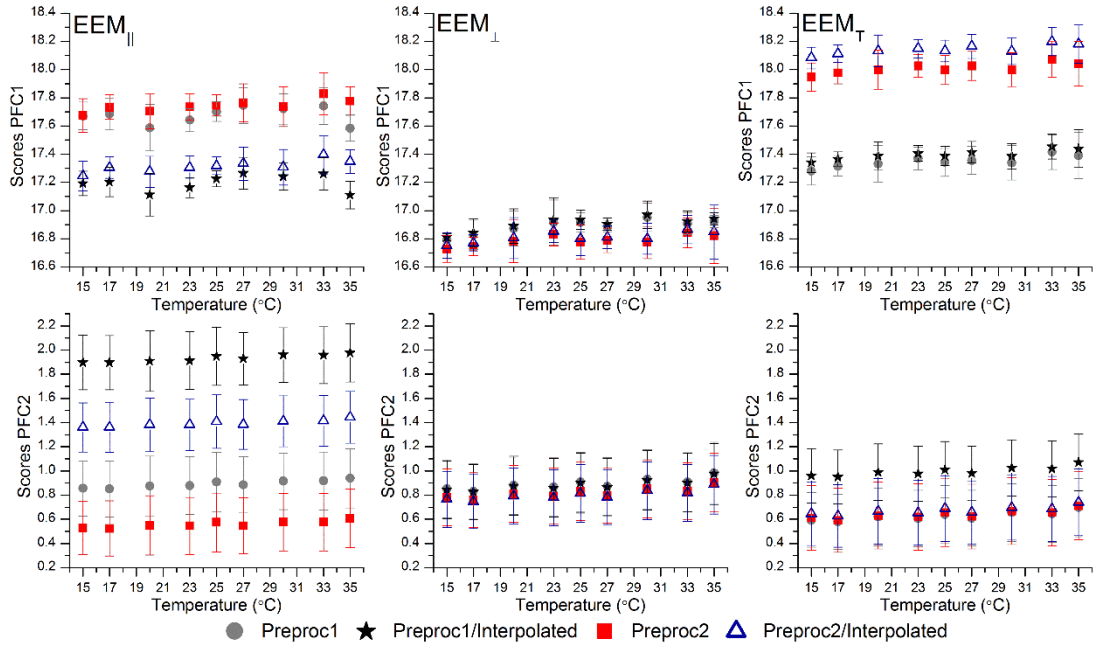

Figure A-5: Scores plots of the two-component PARAFAC models generated for the different pre-processing methods used to correct the rIgG polarized EEM∥, EEM⊥, and unpolarized EEM$_T$ datasets (all normalized). (Top) PFC1, and (Bottom) PFC2 of the PARAFAC models produced with the normalized polarized EEM∥, EEM⊥, and unpolarized EEM$_T$ rIgG datasets.

7.6 TSFS rIgG native state

Table A-4: Selection of number of components using the pre-processed and normalized polarized EEM/t-EEM‖ modeled by PARAFAC (three repetition models). IFE corrected using UV-vis of the replicate sample collected at the same day at 20°C. Polarized t-EEM was interpolated. EEM was pre-processed as previously described.

	EEM‖			**t-EEM‖**			**t-EEM⊥**	**t-EEMT**
Components	**4**	**3**	**2**	**4**	**3**	**2**	**3**	**3**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	278/ 338	280/ 336	278/ 336	280/ 338	280/ 338	280/ 338	280/ 342	280/ 340
C1 Fit model (%)	59.39	62.02	99.38	58.94	83.36	96.99	89.83	91.78
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 308	280/ 328	298/ 340	280/ 340	280/ 340	278/ 298	278/ 314	278/ 312
C2 Fit model (%)	24.95	37.76	0.62	30.82	11.86	3.01	7.03	5.35
C3 $\lambda_{ex}/\lambda_{em}$ (nm)	286/ 336	298/ 342	–	278/ 312	278/ 298	–	276/ 328	276/ 330
C3 Fit model (%)	14.60	0.22	–	9.38	4.78	–	3.15	2.87
C4 $\lambda_{ex}/\lambda_{em}$ (nm)	298/ 338	–	–	298/ 330	–	–	–	–
C4 Fit model (%)	0.56	–	–	0.86	–	–	–	–
Variance explained (%)	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99
CONCORDIA (%)	-6.8×10^4	−125.48	98.23	-2.4×10^6	−387.45	29.14	−281.63	−195.12
Split-half analysis (%)	54.41	90.08	99.91	8.08	23.82	99.97	99.84	99.91

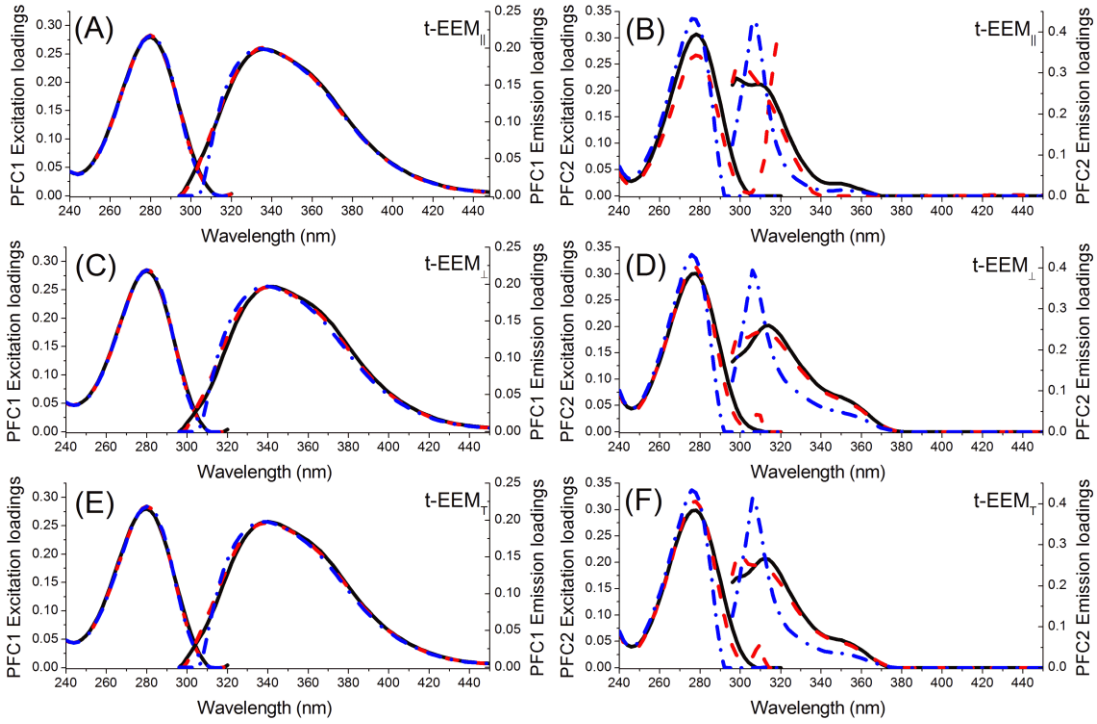

Figure A-6: Comparison of the two-component PARAFAC models for the IFE corrected t-EEM rIgG data pre-processed with interpolating (black line), missing data (dashed red line), and filled with zeros (dash and dotted blue line). PFC1 for the polarized (A) t-EEM‖, (C) t-EEM⊥, and (E) unpolarized t-EEMT rIgG was virtually identical for the dataset with missing data, filled with zeros, and interpolated.

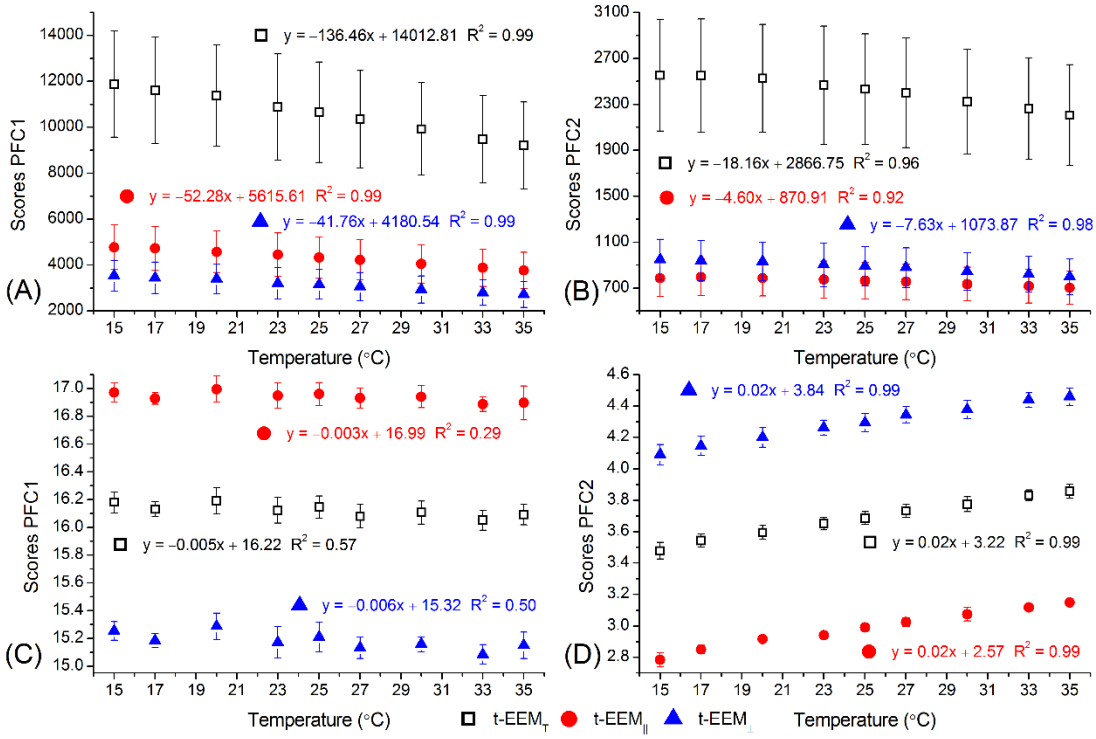

Figure A-7: Scores plots of the two-component PARAFAC models for the (top) non-normalized, and (bottom) normalized unpolarized rIgG t-EEM$_T$, and polarized t-EEM$_{\parallel/\perp}$ interpolated datasets.

7.7 Inner filter effect on pMDF measurements

Table A-5: Comparison of the model parameters and components obtained for the non-IFE corrected and normalized polarized EEM/t-EEM$_\parallel$, EEM/t-EEM$_\perp$, and unpolarized EEM/t-EEM$_T$ PARAFAC models (three repetition models) of the rIgG native state. The t-EEM datasets were interpolated.

	rIgG native state non-IFE corrected					
	EEM$_\parallel$	**EEM$_\perp$**	**EEM$_T$**	**t-EEM$_\parallel$**	**t-EEM$_\perp$**	**t-EEM$_T$**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/336	278/336	280/336	280/334	278/334	278/336
C1 Fit model (%)	99.86	99.72	99.85	93.02	79.36	93.52
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	300/342	296/346	298/346	296/338	290/342	294/340
C2 Fit model (%)	0.14	0.28	0.15	6.98	20.64	6.48
Variance explained (%)	99.98	99.98	99.98	99.99	99.99	99.99
CONCORDIA (%)	99.02	96.51	96.88	81.93	83.90	70.67
Split-half analysis (%)	99.69	99.82	99.86	99.50	99.74	99.88

Table A-6: Comparison of the model parameters and components obtained for the IFE corrected and normalized polarized EEM/t-EEM$_\parallel$, EEM/t-EEM$_\perp$, and unpolarized EEM/t-EEM$_T$ PARAFAC models (three repetition models) of the rIgG native state. The t-EEM datasets were interpolated.

	rIgG native state IFE corrected					
	EEM$_\parallel$	**EEM$_\perp$**	**EEM$_T$**	**t-EEM$_\parallel$**	**t-EEM$_\perp$**	**t-EEM$_T$**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	278/336	278/336	278/336	280/338	280/342	280/340
C1 Fit model (%)	99.38	99.65	99.64	96.99	92.62	95.03
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	298/340	296/342	296/342	278/298	278/314	278/312
C2 Fit model (%)	0.62	0.35	0.36	3.01	7.38	4.97
Variance explained (%)	99.99	99.99	99.99	99.99	99.99	99.99
CONCORDIA (%)	98.23	96.77	96.91	29.14	30.12	24.03
Split-half analysis (%)	99.91	99.37	99.89	99.97	99.95	99.96

7.8 EEM rIgG thermally stressed

Visible aggregates were formed at 75°C, followed by the formation of a gel-like solution, increasing optical density (Figure A-8), limiting IFE correction up to 70°C (Figure A-2), and restricting PARAFAC analysis to this temperature point.

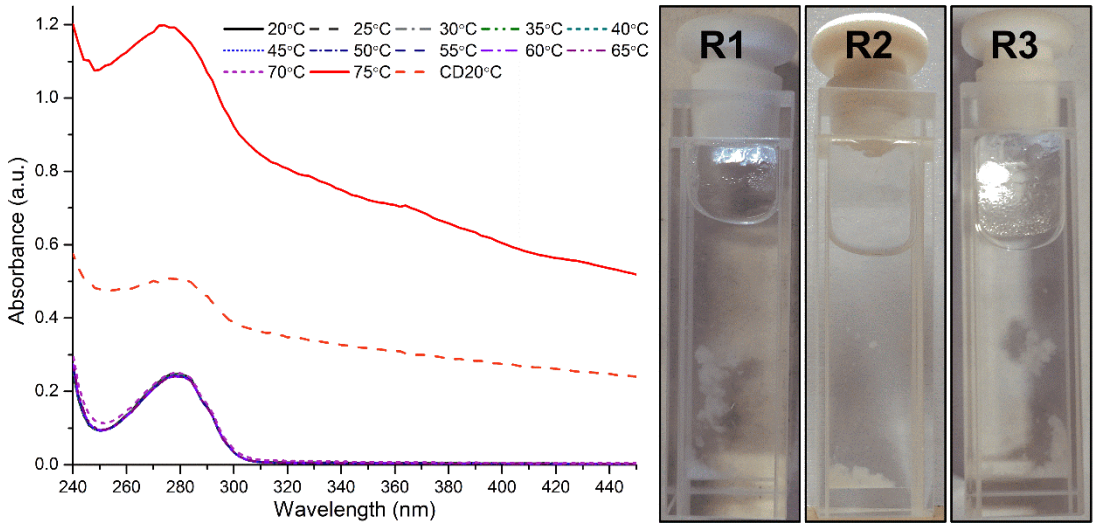

Figure A-8: (Left) UV-Vis absorbance spectra for the rIgG R2 samples (PBS buffer spectrum subtracted) from 20° to 75°C and after cooling down to 20°C (CD20°C). (Right) Visible aggregates were formed at 75°C, followed by a gel formation at CD20°C. The UV-vis absorbance spectra collected at 75°C and CD20°C cannot be used for IFE correction purposes due to the high optical density. The UV-vis absorbance spectra collected at the temperatures used for IFE correction showed no significant difference in rIgG spectra up to 70°C.

Table A-7: Comparison of the model parameters and components obtained for the normalized polarized $EEM_{\parallel}$, $EEM_{\perp}$, and unpolarized EEM_T two-, three-, and four-component PARAFAC models of the thermally stressed rIgG not corrected for instrumental factor over the 20−70°C temperature range.

	$EEM_{\parallel}$	$EEM_{\perp}$	EEM_T	$EEM_{\parallel}$	$EEM_{\perp}$	EEM_T	$EEM_{\parallel}$	$EEM_{\perp}$	EEM_T
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 328	278/ 334	278/ 334	280/ 338	280/ 342	280/ 342	278/ 340	280/ 350	278/ 328
C1 Fit model (%)	64.54	87.69	86.04	96.03	64.97	93.31	67.35	58.86	48.07
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 350	282/ 356	280/ 354	280/ 300	280/ 350	278/ 298	278/ 310	280/ 352	280/ 350
C2 Fit model (%)	35.46	12.31	13.96	3.34	21.77	4.79	29.52	24.31	33.45
C3 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	282/ 362	278/ 320	282/ 354	296/ 336	280/ 322	280/ 356
C3 Fit model (%)	–	–	–	0.63	13.25	1.90	2.69	11.74	16.22
C4 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	–	–	–	282/ 366	276/ 320	280/ 298
C4 Fit model (%)	–	–	–	–	–	–	0.44	5.10	2.26
Variance explained (%)	99.97	99.98	99.98	99.99	99.99	99.99	99.99	99.99	99.99
CONCORDIA (%)	98.96	99.83	99.99	−75.49	-1.1×10^5	-4.6×10^3	−5092	-1.3×10^5	-1.3×10^6
Split-half analysis (%)	70.39	58.87	67.48	3.76	14.19	8.24	0.29	40.08	2.95

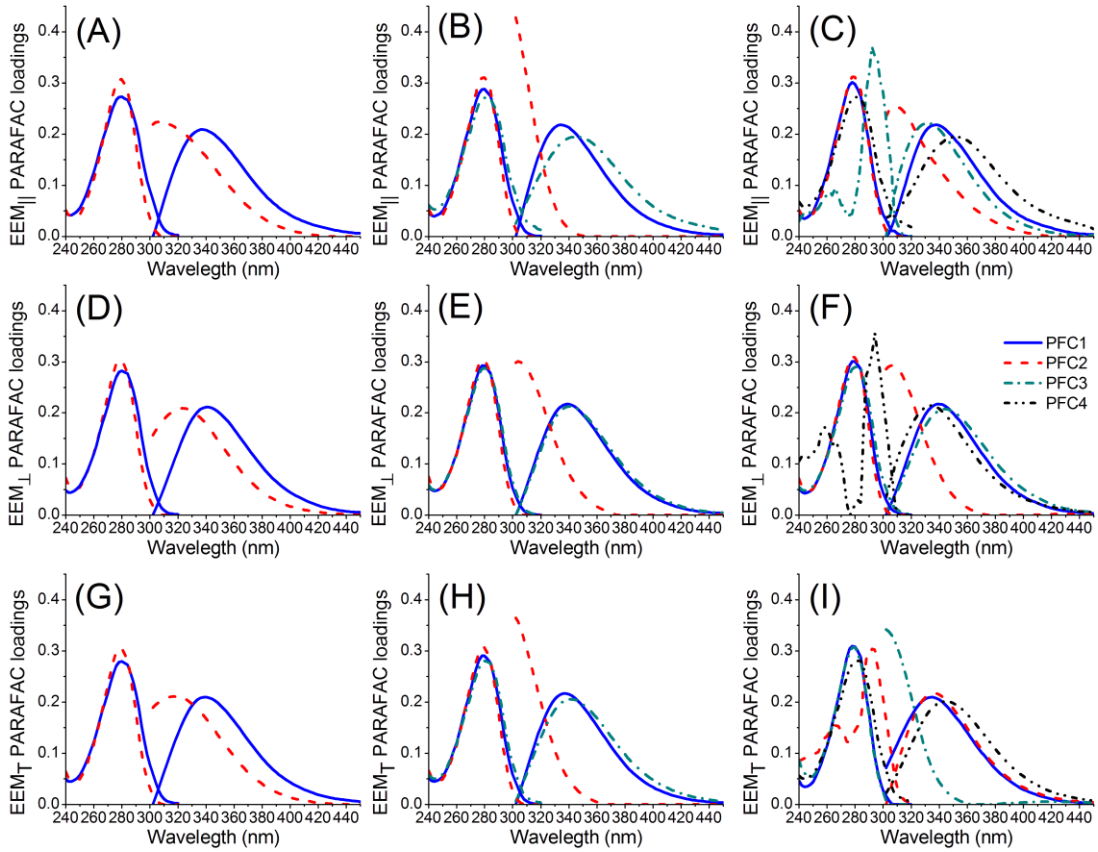

Figure A-9: Comparison of the two-, three-, and four-component PARAFAC models for the pre-processed polarized EEM‖, EEM⊥, and unpolarized EEM_T thermally stressed (20−70°C) rIgG.

For the three-component models, PFC1 resolved an average of all Trp fluorophores in the native state, PFC2 represented the pure Tyr contribution, and PFC3 indicated a change in the average Trp populations with the increase in temperature. A visual analysis of PFC1 and PFC3 loadings would indicate that these two components were the same for pEEM⊥ (Figure A-9E), however, scores plots for these same components (Figure A-10E) showed a change in Trp population. This minimal red shift in Trp maximum emission (2 nm) between PFC1 and PFC3 resolved from the pEEM⊥ data could be attributed to the fact that PARAFAC extracts an average of all Trp fluorophores within the rIgG structure. The change in Trp populations was more obvious for the pEEM‖ data (Figure A-9B), with a 10 nm red shift from PFC1 to PFC3. The minimal contribution of PFC3 (<1%) for the pEEM‖ indicated that this polarization was more sensitive to changes in Trp environment.

The four-component models (Figure A-9/A-10C/F/I) resolution was slightly different between the two polarizations. PFC1 and PFC2 resolved a pure Trp and Tyr contribution, respectively, in the protein native state. While PFC1 and PFC2 resolved the same fluorophores for pEEM‖ and pEEM⊥, the variance explained for each

component was different between the two polarizations (Table 5.1). A similar contribution of Trp (51.49%) and Tyr (44.45%) for the pEEM∥ confirmed that this polarization was more sensitive to changes in Tyr-to-Trp FRET in comparison with pEEM⊥, which showed a major contribution from Trp (83.05%) fluorophores. Two different Trp populations were recovered for each pEEM, a partially exposed to the solved (pEEM⊥) and a completely exposed to the solvent Trp population (pEEM∥). Scores plot for these two different Trp populations (Figure A-10C/F) resolved for the pEEM data showed the same trend, with a change in Trp population at 65/70°C. The last component resolved represented the directly excited Trp fluorophores. This component could also be assigned to differences between the replicate samples, particularly at 70°C (Figure A-9), as each sample unfolded in a different way.

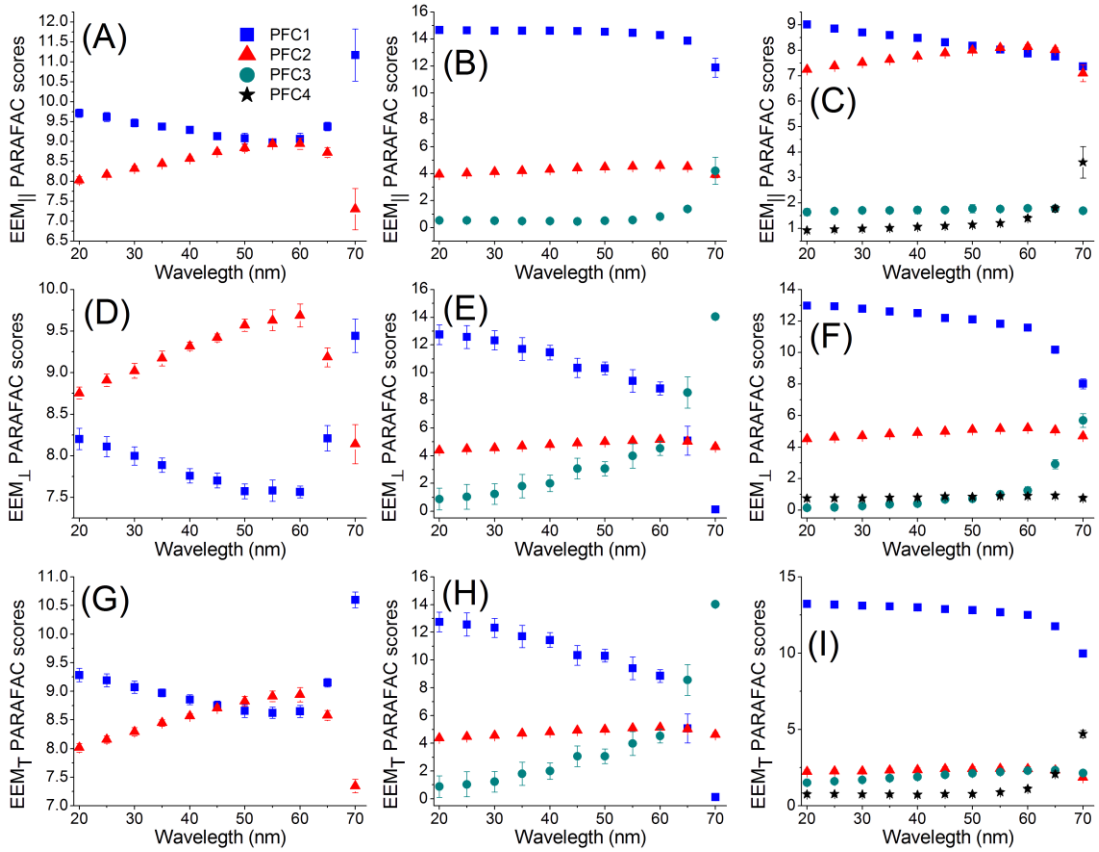

Figure A-10: PARAFAC scores of the two-, three-, and four-component PARAFAC models for the pre-processed polarized EEM∥, EEM⊥, and unpolarized EEMT thermally stressed (20−70°C) rIgG.

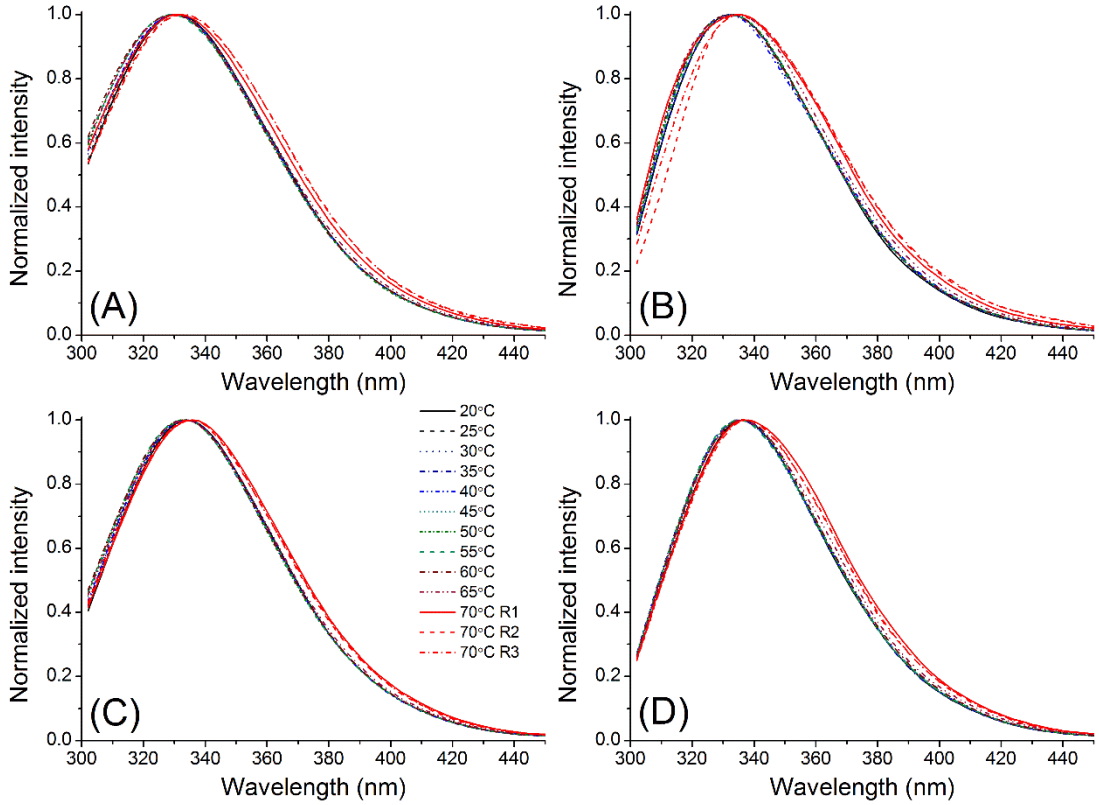

Figure A-11: Comparison between the (top) pEEM∥, and (bottom) pEEM⊥ normalized emission spectra at λ_{ex} (A/C) 280 and (B/D) 296 nm for the thermally stressed (20°−70°C) rIgG. The emission spectrum is different for each replicate sample measured at 70°C.

Table A-8: Comparison of the two-, three-, and four-component PARAFAC models obtained for the normalized polarized EEM∥, EEM⊥, and unpolarized EEM_T of the thermally stressed rIgG corrected for instrumental factor over the 40−70°C temperature range.

	EEM∥	**EEM⊥**	**EEM_T**	**EEM∥**	**EEM⊥**	**EEM_T**	**EEM∥**	**EEM⊥**	**EEM_T**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	278/ 316	280/ 330	278/ 324	280/ 334	280/ 330	280/ 336	278/ 316	280/ 342	278/ 334
C1 Fit model (%)	61.46	84.21	68.85	89.82	94.49	90.21	79.16	60.58	50.59
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 340	288/ 342	280/ 342	278/ 302	294/ 334	278/ 302	278/ 338	280/ 312	280/ 336
C2 Fit model (%)	38.54	15.79	31.315	7.77	3.01	8.44	16.28	29.99	37.68
C3 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	280/ 346	286/ 356	284/ 346	292/ 332	292/ 334	280/ 302
C3 Fit model (%)	–	–	–	2.40	2.50	1.36	2.91	6.02	6.29
C4 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	–	–	–	280/ 348	284/ 346	280/ 344
C4 Fit model (%)	–	–	–	–	–	–	1.65	3.42	5.43
Variance explained (%)	99.97	99.98	99.98	99.99	99.99	99.99	99.99	99.98	99.99
CONCORDIA (%)	99.48	99.41	99.35	−704.85	−64.70	−1285	−7137	−8010	−15701
Split-half analysis (%)	69.98	88.71	59.24	10.44	37.84	9.90	12.43	10.06	0.52

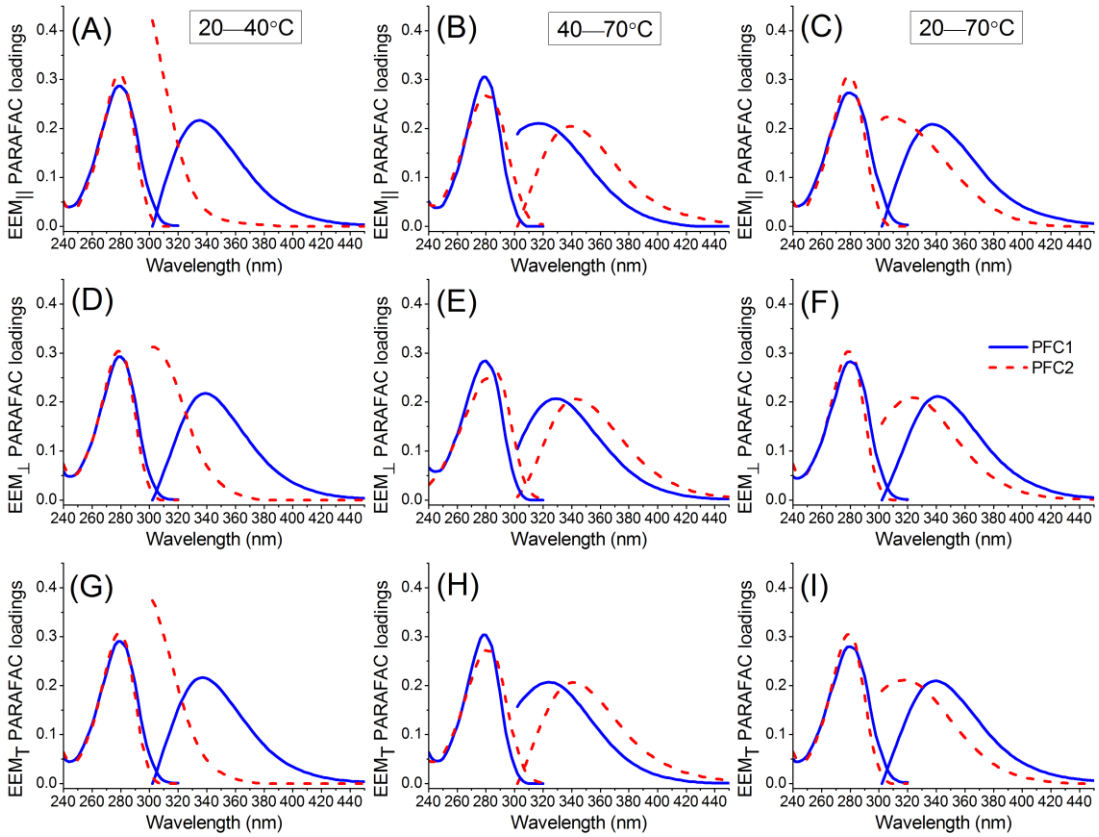

Figure A-12: Comparison of the two-component PARAFAC models for the pre-processed polarized EEM‖, EEM⊥, and unpolarized EEM$_T$ thermally stressed rIgG over the (A/D/G) 20−40°C, (B/E/H) 40−70°C, and (C/F/I) 20−70°C temperature ranges.

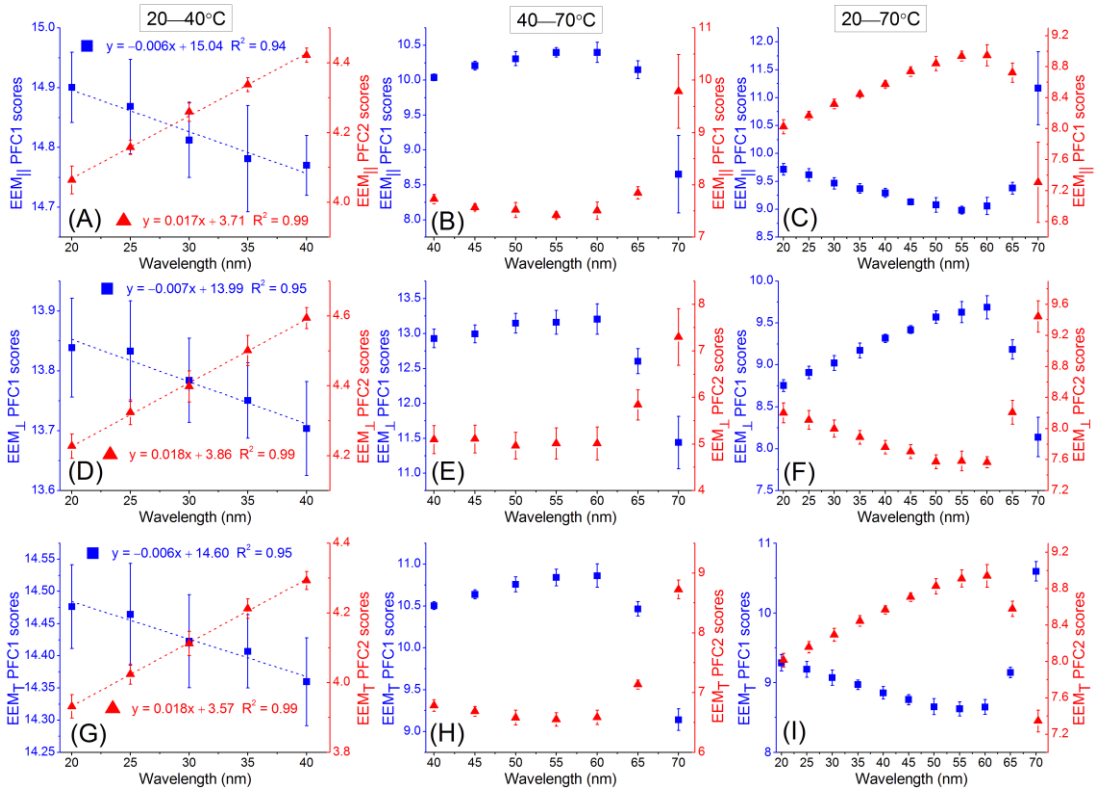

Figure A-13: Comparison of the PARAFAC scores of the two-component PARAFAC models for the pre-processed polarized EEM‖, EEM⊥, and unpolarized EEM$_T$ thermally stressed rIgG over the (A/D/G) 20−40°C, (B/E/H) 40−70°C, and (C/F/I) 20−70°C temperature ranges.

The same information about changes in Tyr and Trp (mainly due to changes in Tyr-to-Trp FRET) was extract from PARAFAC analyses from the full EEM and two selected excitations (Tables 5.1, Tables A-7/A-9), pointing out that the use of a wide range of excitation and emission wavelength does not necessary allows the extraction of more information. Same scores trend for the four-component PARAFAC models resolved for the full EEM and when only two excitation wavelengths were selected (Figure A-14). The fact that collecting only two λ_{ex} extracted the same information from the full pEEM makes this method more suitable for industrial applications. But PARAFAC models from 2 excitations are not stable (very low split-half analysis) – not really a good option.

Table A-9: Comparison of the two-, three-, and four-component PARAFC models parameters and components obtained for the normalized polarized EEM$_{\parallel}$, EEM$_{\perp}$, and unpolarized EEM$_{T}$ of the thermally stressed rIgG collected at λ_{ex} 280 and 296 nm over the 20−70°C temperature range.

	EEM$_{\parallel}$	**EEM$_{\perp}$**	**EEM$_{T}$**	**EEM$_{\parallel}$**	**EEM$_{\perp}$**	**EEM$_{T}$**	**EEM$_{\parallel}$**	**EEM$_{\perp}$**	**EEM$_{T}$**
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 336	280/ 334	280/ 338	280/ 334	280/ 338	280/ 336	280/ 336	280/ 336	280/ 336
C1 Fit model (%)	69.74	52.55	63.80	91.75	65.83	87.69	58.21	91.25	92.71
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	280/ 302	280/ 340	280/ 308	280/ 302	280/ 340	280/ 302	280/ 308	280/ 312	280/ 304
C2 Fit model (%)	30.26	47.45	36.20	7.65	21.46	9.18	33.75	7.59	5.78
C3 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	280/ 348	280/ 302	280/ 342	296/ 322	296/ 334	296/ 334
C3 Fit model (%)	–	–	–	0.60	12.71	3.13	4.68	1.00	1.14
C4 $\lambda_{ex}/\lambda_{em}$ (nm)	–	–	–	–	–	–	280/ 342	280/ 358	280/ 352
C4 Fit model (%)	–	–	–	–	–	–	3.37	0.16	0.36
Variance explained (%)	99.97	99.98	99.98	99.99	99.99	99.99	99.99	99.99	99.99
CONCORDIA (%)	65.29	99.99	95.42	−77.45	−5364	−94.18	50.79	83.27	50.14
Split-half analysis (%)	61.43	15.15	14.46	27.13	67.65	28.53	5.62	2.78	4.48

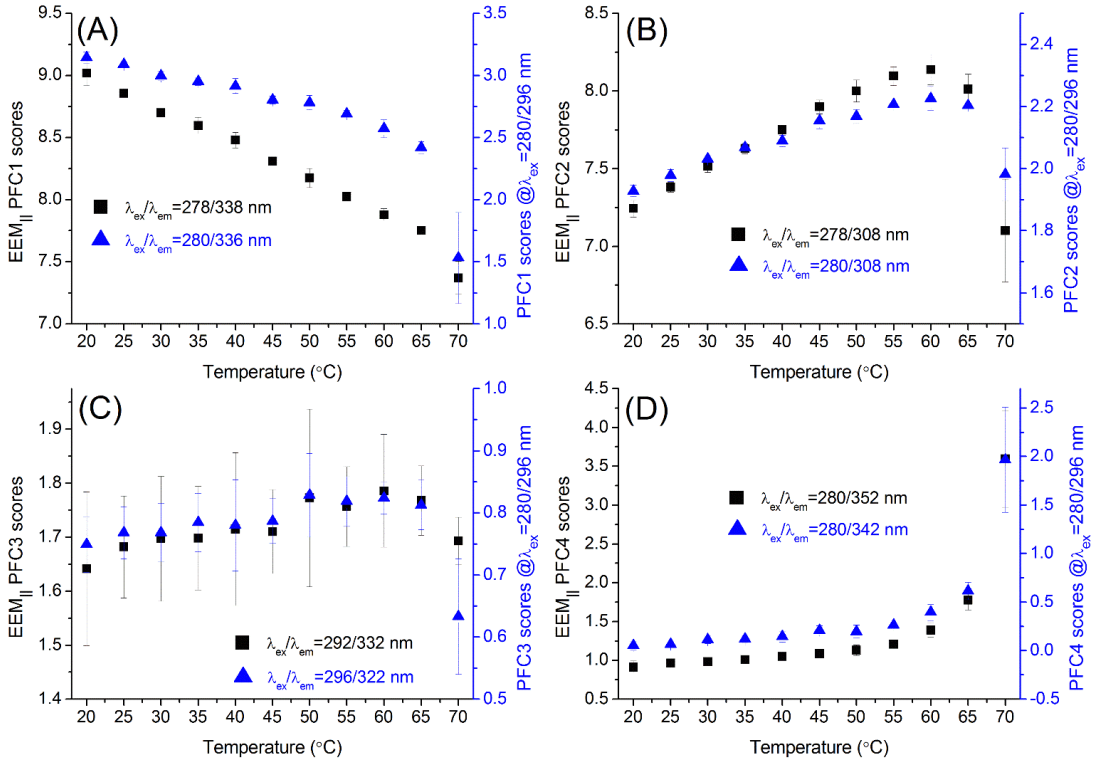

Figure A-14: Scores plots for the four-component PARAFAC models for the (black squares) full pEEM$_{\|}$, and (blue triangles) selected λ_{ex} (280/296 nm) thermally stressed rIgG.

8. **References**

1. Groza, R. C.; Li, B. Y.; Ryder, A. G., Anisotropy resolved multidimensional emission spectroscopy (ARMES): A new tool for protein analysis. *Analytica Chimica Acta* **2015,** *886*, 133-142.
2. Groza, R. C. Anisotropy resolved multi-dimensional emmission spectroscopy (ARMES): A new tool for the quantitative and structural analysis of proteins. Thesis (Ph.D.) : NUI Galway., Galway, 2015.
3. Groza, R. C.; Calvet, A.; Ryder, A. G., A fluorescence anisotropy method for measuring protein concentration in complex cell culture media. *Analytica Chimica Acta* **2014,** *821*, 54-61.
4. Reichert, J. M., Marketed therapeutic antibodies compendium. *Mabs* **2012,** *4* (3), 413-415.
5. Ecker, D. M.; Jones, S. D.; Levine, H. L., The therapeutic monoclonal antibody market. *Mabs* **2015,** *7* (1), 9-14.
6. Grilo, A. L.; Mantalaris, A., The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends in Biotechnology* **2019,** *37* (1), 9-16.
7. Murphy, K. a., *Janeway's immunobiology*. 9th edition. ed.; New York, NY : Garland Science/Taylor & Francis Group, LLC: 2016.
8. Salfeld, J. G., Isotype selection in antibody engineering. *Nature Biotechnology* **2007,** *25* (12), 1369-1372.
9. Pastoret, P.-P., *Handbook of vertebrate immunology*. San Diego ; London : Academic Press: San Diego ; London, 1998.
10. Pinheiro, A.; Neves, F.; de Matos, A. L.; Abrantes, J.; van der Loo, W.; Mage, R.; Esteves, P. J., An overview of the lagomorph immune system and its genetic diversity. *Immunogenetics* **2016,** *68* (2), 83-107.
11. Knight, K. L.; Burnett, R. C.; McNicholas, J. M., Organization and Polymorphism of Rabbit Immunoglobulin Heavy-Chain Genes. *Journal of Immunology* **1985,** *134* (2), 1245-1250.
12. Ionescu, R. M.; Vlasak, J.; Price, C.; Kirchmeier, M., Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies. *Journal of Pharmaceutical Sciences* **2008,** *97* (4), 1414-1426.
13. Arosio, P.; Rima, S.; Morbidelli, M., Aggregation Mechanism of an IgG2 and two IgG1 Monoclonal Antibodies at low pH: From Oligomers to Larger Aggregates. *Pharmaceutical Research* **2013,** *30* (3), 641-654.
14. Sharma, B., Immunogenicity of therapeutic proteins. Part 1: Impact of product handling. *Biotechnology Advances* **2007,** *25* (3), 310-317.
15. Girardi, E.; Holdom, M. D.; Davies, A. M.; Sutton, B. J.; Beavil, A. J., The crystal structure of rabbit IgG-Fc. *Biochemical Journal* **2009,** *417*, 77-83.

16. Arai, H.; Glabe, C.; Luecke, H., Crystal structure of a conformation-dependent rabbit IgG Fab specific for amyloid prefibrillar oligomers. *Biochimica Et Biophysica Acta-General Subjects* **2012,** *1820* (12), 1908-1914.
17. Cerasoli, E.; Ravi, J.; Garfagnini, T.; Gnaniah, S.; le Pevelen, D.; Tranter, G. E., Temperature denaturation and aggregation of a multi-domain protein (IgG1) investigated with an array of complementary biophysical methods. *Analytical and Bioanalytical Chemistry* **2014,** *406* (26), 6577-6586.
18. Li, C. H.; Nguyen, X.; Narhi, L.; Chemmalil, L.; Towers, E.; Muzammil, S.; Gabrielson, J.; Jiang, Y., Applications of circular dichroism (CD) for structural analysis of proteins: qualification of near- and far-UV CD for protein higher order structural analysis. *Journal of Pharmaceutical Sciences* **2011,** *100* (11), 4642-54.
19. Ranjbar, B.; Gill, P., Circular Dichroism Techniques: Biomolecular and Nanostructural Analyses- A Review. *Chemical Biology & Drug Design* **2009,** *74* (2), 101-120.
20. Zolls, S.; Tantipolphan, R.; Wiggenhorn, M.; Winter, G.; Jiskoot, W.; Friess, W.; Hawe, A., Particles in therapeutic protein formulations, Part 1: Overview of analytical methods. *Journal of Pharmaceutical Sciences* **2012,** *101* (3), 914-935.
21. Knäbelin, J.; Müller, R. H., *Modern Biopharmaceuticals : design, development and optimization*. Weinheim : Wiley-VCH: Weinheim, 2005.
22. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. 3rd Edition ed.; Springer: New York, 2006.
23. Eftink, M. R., Fluorescence techniques for studying protein-structure. *Methods of Biochemical Analysis* **1991,** *35*, 127-205.
24. Valeur, B., *Molecular fluorescence: principles and applications*. Wiley-VCH: New York, 2002.
25. Togashi, D. M.; Ryder, A. G.; O'Shaughnessy, D., Monitoring Local Unfolding of Bovine Serum Albumin During Denaturation Using Steady-State and Time-Resolved Fluorescence Spectroscopy. *Journal of Fluorescence* **2010,** *20* (2), 441-452.
26. Togashi, D. M.; Ryder, A. G., Assessing protein-surface interactions with a series of multi-labeled BSA using fluorescence lifetime microscopy and Forster Energy Resonance Transfer. *Biophysical Chemistry* **2010,** *152* (1-3), 55-64.
27. Chen, R. F., Measurements of absolute values in biochemical fluorescence spectroscopy. *Journal of Research of the National Bureau of Standards Section C-Engineering and Instrumentation* **1972,** *A 76* (6), 593-606.

28. Eaton, D. F., Reference materials for fluorescence measurement. *Pure and Applied Chemistry* **1988,** *60* (7), 1107-1114.
29. Taniguchi, M.; Lindsey, J. S., Database of Absorption and Fluorescence Spectra of > 300 Common Compounds for use in PhotochemCAD. *Photochemistry and Photobiology* **2018,** *94* (2), 290-327.
30. Augustijn, D.; Mahapatra, S.; Streicher, W.; Svilenov, H.; Kulakova, A.; Pohl, C.; Rinnan, A., Novel non-linear curve fitting to resolve protein unfolding transitions in intrinsic fluorescence differential scanning fluorimetry. *European Journal of Pharmaceutics and Biopharmaceutics* **2019,** *142*, 506-517.
31. Robbins, R. J.; Fleming, G. R.; Beddard, G. S.; Robinson, G. W.; Thistlethwaite, P. J.; Woolfe, G. J., Photophysics of Aqueous Tryptophan - pH and Temperature Effects. *Journal of the American Chemical Society* **1980,** *102* (20), 6271-6279.
32. Lakowicz, J. R., On spectral relaxation in proteins. *Photochemistry and Photobiology* **2000,** *72* (4), 421-437.
33. Vivian, J. T.; Callis, P. R., Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal* **2001,** *80* (5), 2093-2109.
34. Reshetnyak, Y. K.; Burstein, E. A., Decomposition of protein tryptophan fluorescence spectra into log-normal components. II. The statistical proof of discreteness of tryptophan classes in proteins. *Biophysical Journal* **2001,** *81* (3), 1710-1734.
35. Burstein, E. A.; Vedenkina, N. S.; Ivkova, M. N., Fluorescence and Location of Tryptophan Residues in Protein Molecules. *Photochemistry and Photobiology* **1973,** *18* (4), 263-279.
36. Valeur, B.; Weber, G., Resolution of Fluorescence Excitation Spectrum of Indole into 1La and 1Lb Excitation Bands. *Photochemistry and Photobiology* **1977,** *25* (5), 441-444.
37. Callis, P. R., L-1(a) and L-1(b) transitions of tryptophan: Applications of theory and experimental observations to fluorescence of proteins. *Fluorescence Spectroscopy* **1997,** *278*, 113-150.
38. Eftink, M. R.; Selvidge, L. A.; Callis, P. R.; Rehms, A. A., Photophysics of indole-derivatives - experimental resolution of La and Lb transition and comparison with theory. *Journal of Physical Chemistry* **1990,** *94* (9), 3469-3479.
39. Ghisaidoobe, A. B. T.; Chung, S. J., Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Forster Resonance Energy Transfer Techniques. *International Journal of Molecular Sciences* **2014,** *15* (12), 22518-22538.
40. Wilke, J.; Wilke, M.; Brand, C.; Spiegel, J. D.; Marian, C. M.; Schmitt, M., Modulation of the L-a/L-b Mixing in an Indole Derivative:

A Position Dependent Study Using 4-, 5-, and 6-Fluoroindole. *Journal of Physical Chemistry A* **2017,** *121* (8), 1598-1607.
41. Petrovic, D. M.; Hesp, B.; Broos, J., Emitting State of 5-Hydroxyindole, 5-Hydroxytryptophan, and 5-Hydroxytryptophan Incorporated in Proteins. *Journal of Physical Chemistry B* **2013,** *117* (37), 10792-10797.
42. Garidel, P.; Hegyi, M.; Bassarab, S.; Weichel, M., A rapid, sensitive and economical assessment of monoclonal antibody conformational stability by intrinsic tryptophan fluorescence spectroscopy. *Biotechnology Journal* **2008,** *3* (9-10), 1201-1211.
43. Liu, L.; Braun, L. J.; Wang, W.; Randolph, T. W.; Carpenter, J. F., Freezing-Induced Perturbation of Tertiary Structure of a Monoclonal Antibody. *Journal of Pharmaceutical Sciences* **2014,** *103* (7), 1979-1986.
44. Lakowicz, J. R.; Maliwal, B. P., Oxygen quenching and fluorescence depolarization of tyrosine residues in proteins. *Journal of Biological Chemistry* **1983,** *258* (8), 4794-4801.
45. Weber, G., Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds. *Biochemical Journal* **1960,** *75*, 335-345.
46. Wiberg, K.; Sterner-Molin, A.; Jacobsson, S. P., Simultaneous determination of albumin and immunoglobulin G with fluorescence spectroscopy and multivariate calibration. *Talanta* **2004,** *62* (3), 567-574.
47. Burstein, E. A.; Vedenkina, N. S.; Ivkova, M. N., Fluorescence and the location of tryptophan residues in protein molecules. Photochemistry and Photobiology, 1973; Vol. 18, pp 263–279.
48. Eisinger, J., Intramolecular Energy Transfer in Adrenocorticotropin. *Biochemistry* **1969,** *8* (10), 3902-&.
49. Boteva, R.; Zlateva, T.; DorovskaTaran, V.; Visser, A.; Tsanev, R.; Salvato, B., Dissociation equilibrium of human recombinant interferon gamma. *Biochemistry* **1996,** *35* (47), 14825-14830.
50. Kothawala, D. N.; Murphy, K. R.; Stedmon, C. A.; Weyhenmeyer, G. A.; Tranvik, L. J., Inner filter correction of dissolved organic matter fluorescence. *Limnology and Oceanography-Methods* **2013,** *11*, 616-630.
51. Tarai, M.; Mishra, A. K., Inner filter effect and the onset of concentration dependent red shift of synchronous fluorescence spectra. *Analytica Chimica Acta* **2016,** *940*, 113-119.
52. Divya, O.; Mishra, A. K., Understanding the concept of concentration-dependent red-shift in synchronous fluorescence spectra: Prediction of lambda(max)(SFS) and optimization of Delta lambda for synchronous fluorescence scan. *Analytica Chimica Acta* **2008,** *630* (1), 47-56.

53. Ryder, A. G.; Stedmon, C. A.; Harrit, N.; Bro, R., Calibration, standardization, and quantitative analysis of multidimensional fluorescence (MDF) measurements on complex mixtures (IUPAC Technical Report). *Pure and Applied Chemistry* **2017,** *89* (12), 1849-1870.
54. Xu, J. X.; Vithanage, B. C. N.; Athukorale, S. A.; Zhang, D. M., Scattering and absorption differ drastically in their inner filter effects on fluorescence, resonance synchronous, and polarized resonance synchronous spectroscopic measurements. *Analyst* **2018,** *143* (14), 3382-3389.
55. Nettles, C. B.; Hu, J.; Zhang, D. M., Using Water Raman Intensities To Determine the Effective Excitation and Emission Path Lengths of Fluorophotometers for Correcting Fluorescence Inner Filter Effect. *Analytical Chemistry* **2015,** *87* (9), 4917-4924.
56. Silva, A.; Elcoroaristizabal, S.; Ryder, A. G., Multi-attribute quality screening of immunoglobulin G using polarized Excitation Emission Matrix spectroscopy. *Analytica Chimica Acta* **2020,** *1101*, 99-110.
57. Casamayou-Boucau, Y.; Ryder, A. G., Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters, validation standards, and Rayleigh scatter removal methods. *Methods and Applications in Fluorescence* **2017,** *5* (3), 9.
58. Steiner, R. F., Fluorescence Anisotropy: Theory and Applications. In *Topics in Fluorescence Spectroscopy*, Lakowicz, J. R., Ed. Springer: Boston, MA, 2002; Vol. 2, pp 1-52.
59. Jameson, D. M.; Ross, J. A., Fluorescence Polarization/Anisotropy in Diagnostics and Imaging. *Chemical Reviews* **2010,** *110* (5), 2685-2708.
60. Ameloot, M.; vandeVen, M.; Acuna, A. U.; Valeur, B., Fluorescence anisotropy measurements in solution: Methods and reference materials (IUPAC Technical Report). *Pure and Applied Chemistry* **2013,** *85* (3), 589-608.
61. Goletz, C.; Wagner, M.; Grubel, A.; Schmidt, W.; Korf, N.; Werner, P., Standardization of fluorescence excitation-emission-matrices in aquatic milieu. *Talanta* **2011,** *85* (1), 650-656.
62. Casamayou-Boucau, Y. Anisotropy resolved multidimensional emission spectroscopy (ARMES) for the analysis and resolution of insulin oligomer emission. Thesis (Ph.D.) : NUI Galway, Galway, 2019.
63. Zhang, S. Z.; Zhao, F. L.; Li, K. A.; Tong, S. Y., Determination of glycogen by Rayleigh light scattering. *Analytica Chimica Acta* **2001,** *431* (1), 133-139.
64. Siriwardana, K.; Vithanage, B. C. N.; Zou, S. L.; Zhang, D. M., Quantification of the Depolarization and Anisotropy of Fluorophore

Stokes-Shifted Fluorescence, On-Resonance Fluorescence, and Rayleigh Scattering. *Analytical Chemistry* **2017,** *89* (12), 6686-6694.
65. Weber, G., Fluorescence-polarization spectrum and electronic-energy transfer in proteins. *Biochemical Journal* **1960,** *75*, 345-352.
66. Jameson, D. M.; Croney, J. C., Fluorescence polarization: Past, present and future. *Combinatorial Chemistry & High Throughput Screening* **2003,** *6* (3), 167-176.
67. Li, B.; Shanahan, M.; Calvet, A.; Leister, K. J.; Ryder, A. G., Comprehensive, quantitative bioprocess productivity monitoring using fluorescence EEM spectroscopy and chemometrics. *Analyst* **2014,** *139* (7), 1661-1671.
68. Warner, I. M.; Christian, G. D.; Davidson, E. R.; Callis, J. B., Analysis of Multicomponent Fluorescence Data. *Analytical Chemistry* **1977,** *49* (4), 564-573.
69. Patra, D.; Mishra, A. K., Recent developments in multi-component synchronous fluorescence scan analysis. *Trac-Trends in Analytical Chemistry* **2002,** *21* (12), 787-798.
70. Li, B.; Ryan, P. W.; Shanahan, M.; Leister, K. J.; Ryder, A. G., Fluorescence Excitation-Emission Matrix (EEM) Spectroscopy for Rapid Identification and Quality Evaluation of Cell Culture Media Components. *Applied Spectroscopy* **2011,** *65* (11), 1240-1249.
71. Kumar, K.; Mishra, A. K., Analysis of dilute aqueous multifluorophoric mixtures using excitation-emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy: A comparative evaluation. *Talanta* **2013,** *117*, 209-220.
72. Ohadi, K.; Legge, R. L.; Budman, H. M., Intrinsic fluorescence-based at situ soft sensor for monitoring monoclonal antibody aggregation. *Biotechnology Progress* **2015,** *31* (5), 1423-1432.
73. Kumar, K.; Mishra, A. K., Application of parallel factor analysis to total synchronous fluorescence spectrum of dilute multifluorophoric solutions: Addressing the issue of lack of trilinearity in total synchronous fluorescence data set. *Analytica Chimica Acta* **2012,** *755*, 37-45.
74. Kumar, K.; Mishra, A. K., Parallel factor (PARAFAC) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets in excitation-emission matrix fluorescence (EEMF) layout: Certain practical aspects. *Chemometrics and Intelligent Laboratory Systems* **2015,** *147*, 121-130.
75. Kumar, K.; Tarai, M.; Mishra, A. K., Unconventional steady-state fluorescence spectroscopy as an analytical technique for analyses of complex-multifluorophoric mixtures. *Trac-Trends in Analytical Chemistry* **2017,** *97*, 216-243.

76. Wold, S., Chemometrics; What do we mean with it, and what do we want from it? *Chemometrics and Intelligent Laboratory Systems* **1995,** *30* (1), 109-115.
77. Brereton, R. G., The evolution of chemometrics. *Analytical Methods* **2013,** *5* (16), 3785-3789.
78. Bakeev, K. A., *Process analytical technology spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. Blackwell Pub.: Oxford, UK; Ames, Iowa, 2005.
79. Brown, S. D.; Sarabia, L. A.; Trygg, J., *Comprehensive chemometrics chemical and biochemical data analysis*. Amsterdam : Elsevier: Amsterdam, 2009.
80. Amigo, J. M.; Marini, F., Multiway Methods. In *Chemometrics in Food Chemistry*, Marini, F., Ed. Elsevier Science Bv: Amsterdam, 2013; Vol. 28, pp 265-313.
81. Ruckebusch, C.; Blanchet, L., Multivariate curve resolution: A review of advanced and tailored applications and challenges. *Analytica Chimica Acta* **2013,** *765*, 28-36.
82. Abdollahi, H.; Tauler, R., Uniqueness and rotation ambiguities in Multivariate Curve Resolution methods. *Chemometrics and Intelligent Laboratory Systems* **2011,** *108* (2), 100-111.
83. Bro, R., PARAFAC. Tutorial and applications. *Chemometrics And Intelligent Laboratory Systems* **1997,** *38* (2), 149-171.
84. Andersen, C. M.; Bro, R., Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics* **2003,** *17* (4), 200-215.
85. Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R., Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Analytical Methods* **2013,** *5* (23), 6557-6566.
86. Groza, R. C.; Li, B.; Ryder, A. G., Anisotropy resolved multidimensional emission spectroscopy (ARMES): A new tool for protein analysis. *Analytica Chimica Acta* **2015,** *886*, 133-42.
87. Casamayou-Boucau, Y.; Ryder, A. G., Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters, validation standards, and Rayleigh scatter removal methods. *Methods Appl Fluoresc* **2017,** *5* (3), 037001.
88. Casamayou-Boucau, Y.; Ryder, A. G., Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR). *Analytica Chimica Acta* **2018,** *1000*, 132-143.
89. Groza, R. C. Anisotropy Resolved Multi-dimensional Emission Spectroscopy (ARMES): A new tool for the quantitative and structural analysis of proteins. PhD thesis. National University of Ireland Galway, Galway, Ireland, 2016.

90. Tauler, R.; Marques, I.; Casassas, E., Multivariate curve resolution applied to three-way trilinear data: Study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths. *Journal of Chemometrics* **1998,** *12* (1), 55-75.
91. Steiner-Browne, M.; Elcoroaristizabal, S.; Casamayou-Boucau, Y.; Ryder, A. G., Investigating native state fluorescence emission of Immunoglobulin G using polarized Excitation Emission Matrix (pEEM) spectroscopy and PARAFAC. *Chemometrics and Intelligent Laboratory Systems* **2019,** *185*, 1-11.
92. Steiner-Browne, M.; Elcoroaristizabal, S.; Ryder, A. G., Using Polarized Total Synchronous Fluorescence Spectroscopy (pTSFS) with PARAFAC analysis for Characterizing Intrinsic Protein Emission. *Chemometrics and Intelligent Laboratory Systems* **2019,** *194*.
93. Reshetnyak, Y. K.; Koshevnik, Y.; Burstein, E. A., Decomposition of protein tryptophan fluorescence spectra into log-normal components. III. Correlation between fluorescence and microenvironment parameters of individual tryptophan residues. *Biophysical Journal* **2001,** *81* (3), 1735-1758.
94. Shanker, N.; Bane, S. L., Basic aspects of absorption and fluorescence spectroscopy and resonance energy transfer methods. In *Biophysical Tools for Biologists: Vol 1 in Vitro Techniques*, Correia, J. J.; Detrich, H. W., Eds. Elsevier Academic Press Inc: San Diego, 2008; Vol. 84, pp 213-242.
95. Ahmed, H., *Principles and Reactions of Protein Extraction, Purification, and Characterization*. United States of America, 2004; p 480.
96. Sigma-Aldrich IgG from rabbit serum. https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/2/i5006pis.pdf.
97. Hawe, A.; Kasper, J. C.; Friess, W.; Jiskoot, W., Structural properties of monoclonal antibody aggregates induced by freeze-thawing and thermal stress. *European Journal of Pharmaceutical Sciences* **2009,** *38* (2), 79-87.
98. Hauptmann, A.; Podgorsek, K.; Kuzman, D.; Srcic, S.; Hoelzl, G.; Loerting, T., Impact of Buffer, Protein Concentration and Sucrose Addition on the Aggregation and Particle Formation during Freezing and Thawing. *Pharmaceutical Research* **2018,** *35* (5), 16.
99. Wang, W.; Roberts, C. J., *Aggregation of Therapeutic Proteins*. John Wiley & Sons, Inc.: United States of America, 2010.
100. Wang, W.; Wang, Y. J.; Wang, D. Q., Dual effects of Tween 80 on protein stability. *International Journal of Pharmaceutics* **2008,** *347* (1-2), 31-38.

101. Kamerzell, T. J.; Ramsey, J. D.; Middaugh, C. R., Immunoglobulin dynamics, conformational fluctuations, and nonlinear elasticity and their effects on stability. *Journal of Physical Chemistry B* **2008,** *112* (10), 3240-3250.
102. Lim, K.; Jameson, D. M.; Gentry, C. A.; Herron, J. N., Molecular-dynamics of the antifluorescein-4-4-20 antigen-binding fragment. 2. Time-resolved fluorescence spectroscopy. *Biochemistry* **1995,** *34* (21), 6975-6984.
103. Pfeifer, D.; Hoffmann, K.; Hoffmann, A.; Monte, C.; Resch-Genger, U., The calibration kit spectral fluorescence standards - A simple and certified tool for the standardization of the spectral characteristics of fluorescence instruments. *Journal of Fluorescence* **2006,** *16* (4), 581-587.
104. Kumar, K.; Mishra, A. K., Multivariate curve resolution alternating least square (MCR-ALS) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets: Comparing certain ways of arranging TSFS-based three-way array. *Chemometrics and Intelligent Laboratory Systems* **2015,** *147*, 66-74.
105. Li, B.; Shanahan, M.; Calvet, A.; Leister, K. J.; Ryder, A. G., Comprehensive, quantitative bioprocess productivity monitoring using fluorescence EEM spectroscopy and chemometrics. *Analyst* **2014,** *139* (7), 1661-71.
106. Kumar, K.; Mishra, A. K., Understanding the effect of calibration set design for the application of MCR-ALS analysis on excitation-emission matrix fluorescence (EEMF) data sets under commonly used non-negativity constraints. *Chemometrics and Intelligent Laboratory Systems* **2015,** *149*, 70-77.
107. Elcoroaristizabal, S.; de Juan, A.; Garcia, J. A.; Durana, N.; Alonso, L., Comparison of second-order multivariate methods for screening and determination of PAHs by total fluorescence spectroscopy. *Chemometrics and Intelligent Laboratory Systems* **2014,** *132*, 63-74.
108. Patra, D.; Mishra, A. K., Total synchronous fluorescence scan spectra of petroleum products. *Analytical and Bioanalytical Chemistry* **2002,** *373* (4-5), 304-309.
109. Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A., Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. *Journal of Chemometrics* **2006,** *20* (3-4), 99-105.
110. Elcoroaristizabal, S.; Bro, R.; Garcia, J. A.; Alonso, L., PARAFAC models of fluorescence data with scattering: A comparative study. *Chemometrics and Intelligent Laboratory Systems* **2015,** *142*, 124-130.
111. Rinnan, A.; Booksh, K. S.; Bro, R., First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. *Analytica Chimica Acta* **2005,** *537* (1-2), 349-358.

112. Savitzky, A.; Golay, M. J. E., Smoothing + Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* **1964,** *36* (8), 1627-&.
113. Bro, R.; Kiers, H. A. L., A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics* **2003,** *17* (5), 274-286.
114. Gerretzen, J.; Szymanska, E.; Jansen, J. J.; Bart, J.; van Manen, H. J.; van den Heuvel, E. R.; Buydens, L. M. C., Simple and Effective Way for Data Preprocessing Selection Based on Design of Experiments. *Analytical Chemistry* **2015,** *87* (24), 12096-12103.
115. Xu, J.; Liu, X. F.; Wang, Y. T., A symmetrical subtraction combined with interpolated values for eliminating scattering from fluorescence EEM data. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2016,** *165*, 1-14.
116. Eisinger, J.; Feuer, B.; Lamola, A. A., Intramolecular Singlet Excitation Transfer. Applications to Polypeptides. *Biochemistry* **1969,** *8* (10), 3908-+.
117. Chiu, H. C.; Bersohn, R., Electronic-Energy Transfer Between Tyrosine and Tryptophan in Peptides Trp-(Pro)N-Tyr. *Biopolymers* **1977,** *16* (2), 277-288.
118. Davis, K. B.; Zhang, Z. H.; Karpova, E. A.; Zhang, J., Application of tyrosine-tryptophan fluorescence resonance energy transfer in monitoring protein size changes. *Analytical Biochemistry* **2018,** *557*, 142-150.
119. Yguerabide, J.; Epstein, H. F.; Stryer, L., Segmental Flexibility in an Antibody Molecule. *Journal of Molecular Biology* **1970,** *51* (3), 573-590.
120. Hanson, D. C.; Yguerabide, J.; Schumaker, V. N., Segmental flexibility of immunoglobulin-G antibody molecules in solution - a new interpretation. *Biochemistry* **1981,** *20* (24), 6842-6852.
121. Demchenko, A. P., The red-edge effects: 30 years of exploration. *Luminescence* **2002,** *17* (1), 19-42.
122. Li, B.; Ryder, A. G., Similarity index: a rapid classification method for multivariate data arrays. Google Patents: 2011.
123. Runnels, L. W.; Scarlata, S. F., Theory and Applications of Fluorescence Homotransfer to Melittin Oligomerization. *Biophysical Journal* **1995,** *69* (4), 1569-1583.
124. Weber, G., Fluorescence-polarization spectrum and electronic-energy transfer in proteins. *Biochemical Journal* **1960,** *75* (2), 345-52.
125. Visser, N. V.; Westphal, A. H.; van Hoek, A.; van Mierlo, C. P. M.; Visser, A.; van Amerongen, H., Tryptophan-tryptophan energy migration as a tool to follow apoflavodoxin folding. *Biophysical Journal* **2008,** *95* (5), 2462-2469.

126. Kayser, V.; Chennamsetty, N.; Voynov, V.; Helk, B.; Trout, B. L., Tryptophan-Tryptophan Energy Transfer and Classification of Tryptophan Residues in Proteins Using a Therapeutic Monoclonal Antibody as a Model. *Journal of Fluorescence* **2011,** *21* (1), 275-288.
127. Kumar, K.; Mishra, A. K., Application of 'multivariate curve resolution alternating least square (MCR-ALS)' analysis to extract pure component synchronous fluorescence spectra at various wavelength offsets from total synchronous fluorescence spectroscopy (TSFS) data set of dilute aqueous solutions of fluorophores. *Chemometrics and Intelligent Laboratory Systems* **2012,** *116*, 78-86.
128. Kumar, K., Processing Excitation-Emission Matrix Fluorescence and Total Synchronous Fluorescence Spectroscopy Data Sets with Constraint Randomised Non-negative Factor Analysis: a Novel Fluorescence Based Analytical Procedure to Analyse the Multifluorophoric Mixtures. *Journal of Fluorescence* **2018,** *28* (5), 1075-1092.
129. Tomasi, G.; Bro, R., PARAFAC and missing values. *Chemometrics and Intelligent Laboratory Systems* **2005,** *75* (2), 163-180.
130. Thygesen, L. G.; Rinnan, A.; Barsberg, S.; Moller, J. K. S., Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area. *Chemometrics and Intelligent Laboratory Systems* **2004,** *71* (2), 97-106.
131. Timofeev, V. P.; Dudich, I. V.; Sykulev, Y. K.; Nezlin, R. S., Rotational Correlation Times of IgG and its Fragments Spin-Labeled at Carbohydrate or Protein Moieties - Spatially Fixed Position of Fc Carbohydrate. *FEBS Letters* **1978,** *89* (2), 191-195.
132. Bobone, S.; van de Weert, M.; Stella, L., A reassessment of synchronous fluorescence in the separation of Trp and Tyr contributions in protein emission and in the determination of conformational changes. *Journal of Molecular Structure* **2014,** *1077*, 68-76.
133. Vermeer, A. W. P.; Norde, W., The thermal stability of immunoglobulin: Unfolding and aggregation of a multi-domain protein. *Biophysical Journal* **2000,** *78* (1), 394-404.
134. Brader, M. L.; Estey, T.; Bai, S. J.; Alston, R. W.; Lucas, K. K.; Lantz, S.; Landsman, P.; Maloney, K. M., Examination of Thermal Unfolding and Aggregation Profiles of a Series of Developable Therapeutic Monoclonal Antibodies. *Molecular Pharmaceutics* **2015,** *12* (4), 1005-1017.
135. Sousa, F.; Sarrnento, B.; Neves-Petersen, M. T., Biophysical study of bevacizumab structure and bioactivity under thermal and pH-stresses. *European Journal of Pharmaceutical Sciences* **2017,** *105*, 127-136.

136. Menzen, T.; Friess, W., Temperature-Ramped Studies on the Aggregation, Unfolding, and Interaction of a Therapeutic Monoclonal Antibody. *Journal of Pharmaceutical Sciences* **2014,** *103* (2), 445-455.
137. Liu, H. C.; Chumsae, C.; Gaza-Bulseco, G.; Goedken, E. R., Domain-level stability of an antibody monitored by reduction, differential alkylation, and mass spectrometry analysis. *Analytical Biochemistry* **2010,** *400* (2), 244-250.
138. Harn, N.; Allan, C.; Oliver, C.; Middaugh, C. R., Highly concentrated monoclonal antibody solutions: Direct analysis of physical structure and thermal stability. *Journal of Pharmaceutical Sciences* **2007,** *96* (3), 532-546.
139. Vermeer, A. W. P.; Norde, W.; van Amerongen, A., The unfolding/denaturation of immunogammaglobulin of isotype 2b and its F-ab and F-c fragments. *Biophysical Journal* **2000,** *79* (4), 2150-2154.
140. Li, B.; Ryder, A. G. Similarity Index: a Rapid Classification Method for Multivariate Data Arrays. 2011.
141. Bekard, I. B.; Dunstan, D. E., Tyrosine Autofluorescence as a Measure of Bovine Insulin Fibrillation. *Biophysical Journal* **2009,** *97* (9), 2521-2531.
142. Chen, H.; Kenny, J. E., Application of PARAFAC to a two-component system exhibiting Fluorescence Resonance Energy Transfer: from theoretical prediction to experimental validation. *Analyst* **2012,** *137* (1), 153-162.
143. Lenhardt, L.; Bro, R.; Zekovic, I.; Dramicanin, T.; Dramicanin, M. D., Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey. *Food Chemistry* **2015,** *175*, 284-291.
144. Nomine, Y.; Ristriani, T.; Laurent, C.; Lefevre, J. F.; Weiss, E.; Trave, G., A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. *Protein Engineering* **2001,** *14* (4), 297-305.
145. Raynal, B.; Lenormand, P.; Baron, B.; Hoos, S.; England, P., Quality assessment and optimization of purified protein samples: why and how? *Microbial Cell Factories* **2014,** *13*, 10.
146. Drenski, M. F.; Brader, M. L.; Alston, R. W.; Reed, W. F., Monitoring protein aggregation kinetics with simultaneous multiple sample light scattering. *Analytical Biochemistry* **2013,** *437* (2), 185-197.
147. Svilenov, H.; Markoja, U.; Winter, G., Isothermal chemical denaturation as a complementary tool to overcome limitations of thermal differential scanning fluorimetry in predicting physical stability of protein formulations. *European Journal of Pharmaceutics and Biopharmaceutics* **2018,** *125*, 106-113.

9. Presentations & Publications

9.1 Peer-reviewed publications

- Steiner-Browne, M.; Elcoroaristizabal, S.; Casamayou-Boucau, Y.; Ryder, A. G., Investigating native state fluorescence emission of Immunoglobulin G using polarized Excitation Emission Matrix (pEEM) spectroscopy and PARAFAC. Chemometrics and Intelligent Laboratory Systems **2019**, 185, 1-11.
- Steiner-Browne, M.; Elcoroaristizabal, S.; Ryder, A. G., Using Polarized Total Synchronous Fluorescence Spectroscopy (pTSFS) with PARAFAC analysis for Characterizing Intrinsic Protein Emission. Chemometrics and Intelligent Laboratory Systems **2019**, 194.

9.2 International presentations

- Poster presentation: ‘‘Investigating Native State Fluorescence Emission of Immunoglobulin G (IgG) using polarized Excitation Emission Matrix (pEEM) Spectroscopy and PARAFAC’’, M. Steiner-Browne, S. Elcoroaristizabal, Y. Casamayou-Boucau, and A.G. Ryder. *Joint 12th EBSA, 10th ICBP-IUPAP Biophysics Congress*, Madrid, **Spain**, 20-24 July **2019**.
- Poster presentation: ‘‘Monitoring Structural Stability and Aggregation of Immunoglobulin G (IgG) Under Thermal Stress using Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES)’’, M. Steiner, S. Elcoroaristizabal, Y. Casamayou-Boucau, and A.G. Ryder. *15th Conference on Methods and Applications of Fluorescence Spectroscopy, Imaging and Probes, MAF15*, Bruges, **Belgium**, 10-13 September **2017**.

9.3 National presentations

- Oral presentation (**1st prize**): ‘‘Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES) for the Analysis of IgG Type Proteins’’, M. Steiner-Browne, and A.G. Ryder. *2020 BOC Gases Bursary*, Galway, **Ireland**, 23rd June **2020**.

- Oral presentation (**2nd prize**): ‘‘Monitoring rabbit IgG Structure by Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES)’’, M. Steiner-Browne, and A.G. Ryder. *Eli Lilly Postgraduate Prize 2019*, Galway, **Ireland**, 22nd February **2019**.
- Oral presentation: ‘‘Monitoring rabbit IgG structure by Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES)’’, M. Steiner-Browne, and A.G. Ryder. *Chemistry Research Day 2019*, NUI Galway, Galway, **Ireland**, 9th January **2019**.
- Poster presentation (**1st prize**): ‘‘Assessing freeze-thaw effects in IgG using Anisotropy Resolved Multi-Dimensional Emission Spectroscopy (ARMES)’’, M. Steiner-Browne, S. Elcoroaristizabal, Y. Casamayou-Boucau, and A.G. Ryder. *BioPharma Ambition 2018*, Dublin, **Ireland**, 21-22 February **2018**.
- Poster presentation: ‘‘Spectroscopic Studies of Tertiary Structure Stability and Aggregation of Immunoglobulin G (IgG) Solutions Under Different Storage Conditions’’, M. Steiner, S. Elcoroaristizabal, and A.G. Ryder. *7th Annual NUI Galway/UL Alliance Research Day,* Galway, **Ireland**, 19th April **2017**.

Chemometrics and Intelligent Laboratory Systems 185 (2019) 1–11

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Investigating native state fluorescence emission of Immunoglobulin G using polarized Excitation Emission Matrix (pEEM) spectroscopy and PARAFAC

Marina Steiner-Browne, Saioa Elcoroaristizabal, Yannick Casamayou-Boucau, Alan G. Ryder *

Nanoscale BioPhotonics Laboratory, School of Chemistry, National University of Ireland Galway, Galway, H91CF50, Ireland

ARTICLE INFO

Keywords:
Protein
Immunoglobulin G
Fluorescence
Multidimensional
Spectroscopy
Anisotropy
PARAFAC

ABSTRACT

Intrinsic fluorescence spectroscopy (IFS) measurements for protein structural analysis can be enhanced by the use of anisotropy resolved multidimensional emission spectroscopy (ARMES). ARMES attempts to overcome the tryptophan (Trp) and tyrosine (Tyr) spectral overlap problem and resolve emitting components by combining anisotropy measurements with chemometric analysis. Here we investigate for the first time the application of polarized excitation-emission matrix (pEEM) measurements and Parallel Factor (PARAFAC) analysis to study IFS from an Immunoglobulin G (IgG) type protein, rabbit IgG (rIgG), in its native state. Protein IFS is a non-trilinear system primarily because of Förster resonance energy transfer (FRET). Non-trilinearity is also caused by inner filter effects, and Rayleigh/Raman scattering, both of which can be corrected by data pre-processing. However, IFS FRET cannot be corrected for, and thus here we carefully evaluated the impact of various different data pre-processing methods on IFS data which used for PARAFAC. Care must be taken with data pre-processing and interpolation, as both had an impact on PARAFAC modelling and the recovered anisotropy values, with residual shot noise from the Rayleigh scatter which overlapped the emission blue edge being the root cause.

pEEM spectra from thawed rIgG solutions (15–35 °C temperature range) were collected with an expectation being that this temperature range should cause sufficient emission variation to facilitate component resolution but without major structural changes. However, the only significant changes observed were of the overall intensity due to thermal motion induced quenching and this was confirmed by the PARAFAC scores. PARAFAC resolved one major component (>99%) for the emission data (polarized and unpolarized) which mostly represented the large Tyr-to-Trp hetero-FRET process, with a second, very weak component (<1%) apparently a contribution from directly excited Trp emission. PARAFAC scores recovered from normalized pEEM data showed minimal change which was further proof for negligible structural change. The results of this study serves as the starting point for the use of PARAFAC analysis of IFS from IgG type proteins and important processes such as denaturation and aggregation.

1. Introduction

Immunoglobulin G (IgG) is a common class of antibody which is often used for therapeutic purposes [1,2]. Monoclonal antibodies (mAb) have become a very significant class of active pharmaceutical ingredients (API) because of their high specificity, and the leveraging of the immense resources applied to studying the human genome. IgG type proteins in general have a flexible 'Y'-shaped structure, composed of two regions: a constant (Fc), and a variable region (Fab), connected by di-sulphide bonds and a hinge region (~150 kDa). Changes in the Fab portion amino acid sequence are responsible for IgG diversity (Fig. S1, supplemental information, SI) while variations in the connection between Fab and Fc portion produces different IgG isotypes, which are species dependent. Humans have four isotypes [3,4], while rabbits have only one [5–7]. These Fab portion differences between IgG isotypes, as well as changes in amino acid sequences, and protein structure regulate protein stability which is an important consideration for their use as APIs [8,9].

Protein structure changes with physical and chemical stresses and the most serious involves tertiary structure unfolding which can lead to refolding into new structures and/or aggregation. These changes can cause a loss in function and potentially adverse immunogenicity issues [10]. Stresses can be induced by improper manipulation during manufacturing, storage or use, and thus understanding and measuring the stability of proteins in solution is therefore, a key factor in ensuring

* Corresponding author. Nanoscale BioPhotonics Laboratory School of Chemistry, National University of Ireland – Galway, Galway, H91 CF50, Ireland.
E-mail address: alan.ryder@nuigalway.ie (A.G. Ryder).

https://doi.org/10.1016/j.chemolab.2018.12.007
Received 7 November 2018; Received in revised form 14 December 2018; Accepted 17 December 2018
Available online 20 December 2018

efficacy and safety. The various quality attributes of therapeutic proteins can be monitored and measured using a variety of techniques. For example, high-performance liquid chromatography-size exclusion (HPLC-SEC) and dynamic light scattering (DLS) are most commonly used to assess protein aggregation [11]. Spectroscopic methods, such as circular dichroism (CD) and Fourier transform infrared (FTIR) can also be used to characterize the native structure and monitor changes in protein structure [12,13]. While far-UV CD is sensitive to changes in secondary structure, near-UV CD detects changes in tertiary structure, protein concentration has to be carefully selected, as sedimentation and light scattering can negatively impact CD measurements. FTIR can be used with liquid and solid samples in high concentrations and with large aggregates, but proteins can bind/adhere to the crystal and the water signal has to be carefully removed from the spectra. Therefore, a combination of different analytical tools, some of which are expensive and time consuming, are required to characterize protein structure and stability in solution [14].

Intrinsic fluorescence spectroscopy (IFS), which is non-destructive, can also be used to study protein structure in solution, due to the presence of phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp) fluorophores in proteins [15]. Protein emission is mostly comprised of overlapping emission from Tyr and Trp because Phe has a very low quantum yield [15,16]. Förster resonance energy transfer (FRET) is critical in IFS and determines how much emission from each fluorophore can be detected. For example, Phe with its small quantum yield, also undergoes FRET to Tyr or Trp [15], thus, Phe fluorescence cannot usually be discriminated in the presence of the much stronger Tyr or Trp emission. Similarly, the strong spectral overlap between Tyr emission and Trp absorbance also leads to high FRET rates; therefore, protein emission is dominated by Trp if present. In IgG type proteins with multiple Trp (~20/30) and Tyr (~50/60) present (Fig. S1, SI), many located within a Förster radius of each other, the rates of homo- and hetero-FRET will be high, leading to Trp dominated emission. At longer excitation wavelengths, $\lambda_{ex} \geq 295$ nm, Trp is preferentially excited and there is minimal Tyr excitation, therefore it is common to use $\lambda_{ex} = 295$ nm for monitoring tertiary structural changes in proteins caused by physical or chemical stresses [11,17,18].

For the analysis of multi-fluorophore proteins or mixtures, multidimensional fluorescence spectroscopy (MDF) provides more information. The most common measurement methods are excitation-emission matrix (EEM) [19] and total synchronous fluorescence spectroscopy (TSFS) [20]. Three-dimensional EEM and TSFS spectra provide a spectral signature of the multiple fluorophores presents in a protein or a complex multi-fluorophore sample [21–23]. The primary advantages of TSFS over EEM measurements are to avoid the collection of Rayleigh scatter and that it is a somewhat quicker measurement. However, Rayleigh scatter does contain information about the physical nature of the sample and can be used to monitor particle formation such as aggregation [24].

Anisotropy resolved multidimensional emission spectroscopy (ARMES) combines anisotropy, MDF measurements, and chemometrics [25]. ARMES generates multidimensional datasets (excitation wavelength (λ_{ex}), emission wavelength (λ_{em}) or wavelength offset ($\Delta\lambda$), intensity (I), anisotropy (r)) and provides additional information with which one can characterize proteins and differentiate the emission of each fluorophore or a family of fluorophores. For multi-fluorophore proteins, anisotropy varies across the fluorescence emission space (plotted as an *aniso*-MDF map) because of the different fluorophores and the interactions (e.g. FRET) between them [22,26]. The *aniso*-MDF maps are thus sensitive to structural change and this can be used as a different diagnostic measurement for assessing unfolding, refolding, and denaturation processes [25]. MDF data analysis requires the use of chemometric modelling techniques to try and resolve contributions from different fluorophores which can be used for analytical applications [27,28]. For MDF/ARMES, multi-way decomposition methods, like multivariate curve resolution (MCR) [29,30] and parallel factor analysis (PARAFAC) [31–33] are used to identify the spectral contribution of individual constituents. Care is needed with applying these methods to IFS because of the inherent non-linearity of the data, however, they are the most widely available tools and can produce useful outputs under specific conditions. This combination of anisotropy and MDF has previously enabled the differentiation of fluorophores with similar emission properties in complex mixtures. Anisotropy provides for fluorophore differentiation based on factors such as: FRET, rotational speed, hydrodynamic volume/molecular size, and for macromolecules the mobility/flexibility of the constituent fluorophores [25,26,34]. In proteins, all these factors are present, making the analysis of anisotropy changes a complicated process.

Previously we showed that ARMES of Human Serum Albumin (HSA) using MCR modelling yielded multiple components [25], however these studies used thin film polarizers with no transmission below 300 nm which affected the TSFS data structure. Further developments of ARMES to enable short (<300 nm) wavelength excitation [26] and validation measurements have shown that one can accurately recover individual fluorophore emission in the absence of inner filter effects (IFE) and FRET [35]. These early studies also undertook MCR analysis on all four polarization measurements, mainly in an effort to confirm that the chemometric modelling was robust [26,36].

Here we investigate the use of ARMES to better characterize the fluorescence emission of a much more complex IgG protein in its native state, over a small temperature interval. It was hoped that the thermally induced variations over the 15–35 °C range, might be small enough to avoid issues with non-linear changes in the spectral measurements (caused by large structural/FRET changes) yet enough to induce sufficient fluorescence fluctuations for PARAFAC analysis. As the temperature changes, one might expect the different fluorophore populations to quench with varyvaring rates, this might be sufficient to discriminate, the internal versus the externally located Trp emission. This modelling of the native state emission is important because it serves as a baseline with which we can investigate processes such as unfolding and aggregation using ARMES and curve resolution methods in general.

2. Materials and methods

2.1. Materials

IgG from rabbit serum (≥95% essentially salt-free, lyophilized powder) was obtained from Sigma-Aldrich and used as received without further purification. Size Exclusion Chromatography (SEC) analysis of the three lots used for these studies showed that the monomer content was ~80 ± 1% (data not shown). Sodium phosphate monobasic, sodium phosphate dibasic hepta-hydrate, and sodium chloride were used to prepare a 0.01 M Phosphate 0.15 M saline buffer (PBS) at pH 6.5 ± 0.1 in high purity water (HPW) from Honeywell (HPLC grade). Except for the HPW, all the other materials were purchased from Sigma-Aldrich and used without further purification. Rabbit IgG (rIgG) solutions in PBS buffer were prepared in triplicate ($1.1 \pm 0.2\ mg\ mL^{-1}$) using different vials of protein (but with the same lot number, SLBM2617V) and then membrane filtered (0.2 μm) using polyethersulfone (PES) Captiva Premium Syringe filters (Agilent). The three freshly prepared solutions had concentrations of: $R1 = 1.3\ mg\ mL^{-1}$, $R2 = 0.9\ mg\ mL^{-1}$, and $R3 = 1.2\ mg\ mL^{-1}$, and these did not change after defrosting. Aliquots (sufficient for a single measurement) of each replicate solution were then dispensed into 1.5 mL Lobind tubes (Eppendorf) and stored at −70 °C prior to use (over 13 weeks). All sample solution preparation was carried out in a laminar flow hood using aseptic techniques to minimize contamination. Prior to making spectroscopic measurements the samples were slowly defrosted overnight at 4–8 °C, taking care to ensure that there were no ice crystals remaining before transfer into the cuvette.

2.2. Instrumentation and data collection

UV–Vis absorbance spectra (200–800 nm) were collected using a Cary

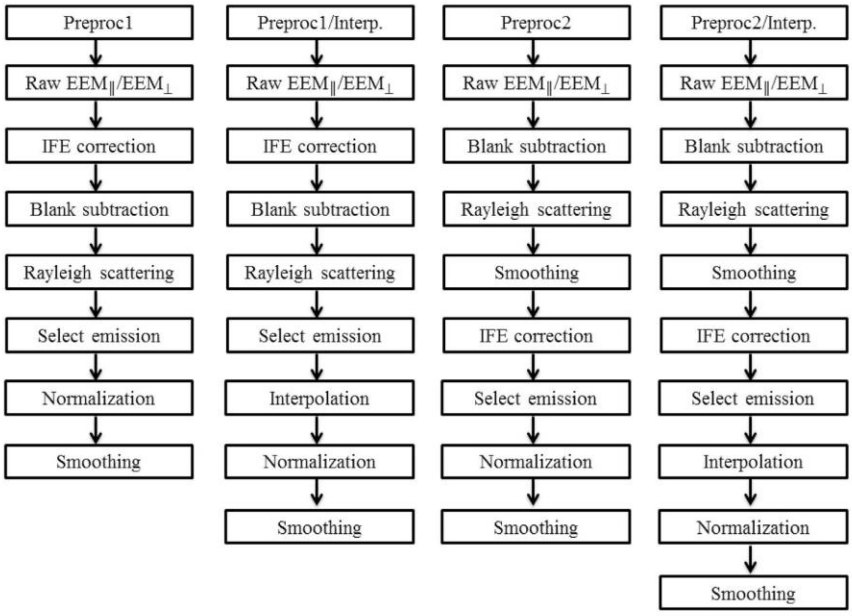

Fig. 1. Schematic of the different sequences of pre-processing steps applied to pEEM data collected from the rIgG native state emission.

60 (Agilent) at a scan rate of 1200 nm min^{-1}. EEM, and ARMES, were performed using a Cary Eclipse fluorimeter (Agilent) fitted with bespoke dual wire grid polarizers [26] and a temperature-regulated multi-cell holder. All spectroscopic measurements were made in triplicate, using 10 × 2 mm pathlength quartz cuvettes (Lightpath Optical, UK). UV–Vis absorbance spectra were collected with a 2 mm pathlength; for the EEM measurements, samples were excited along the short axis (2 mm) and emission collected from the long axis (10 mm). Polarized EEM (pEEM) data were collected over an excitation and emission range of λ_{ex} = 240–320 nm and λ_{em} = 260–450 nm (2 nm step increments for both axis). Excitation and emission monochromators slit widths were 10 nm, the scan rate was 1200 nm min^{-1} and the photomultiplier tube (PMT) detector voltage was set to 650 V. All samples were measured using four polarizer settings: VH (vertical-horizontal), VV (vertical-vertical), HH (horizontal-horizontal), and HV (horizontal-vertical). The anisotropy (r) was calculated using the standard anisotropy formula [15], which was then used to construct the corresponding *aniso*-EEM maps. For the native state characterization, rIgG solutions were measured at 9 different temperatures (15°, 17°, 20°, 23°, 25°, 27°, 30°, 33°, and 35 °C) with 5 min thermal equilibration between measurements and it took ~8 min to collect each pEEM spectrum.

2.3. Data analysis and chemometric methods

Data analysis was performed using the PLS_Toolbox ver. 8.2.1 (Eigenvector Research Inc.), MATLAB ver. 9.1.0 (The Mathworks Inc.),

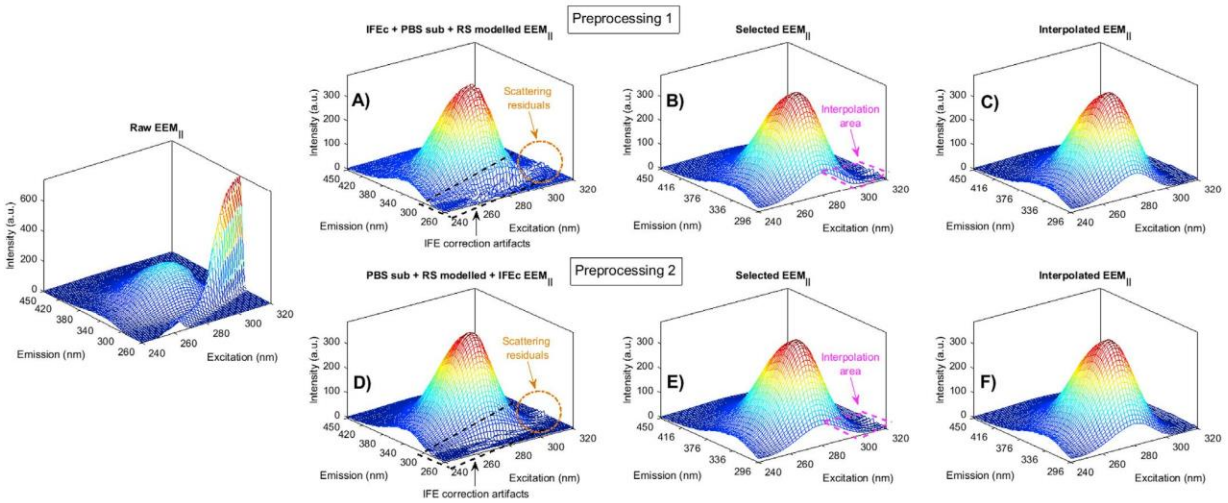

Fig. 2. (Left) Raw EEM‖ spectra of rIgG native state measured at 20 °C. The two different schemes for pre-processing: Preproc1 started with: (A) IFE correction (IFEc), followed by blank subtraction (PBS sub) and then Rayleigh scattering removal. Preproc2 started with: (D) blank subtracted, followed by RS removal, and the IFE correction. Both sets of corrected data were the cut at λ_{em} 296 nm to remove residual scattering and noise (B/E) before being interpolated to remove most of the remaining residual light scatter and finally smoothed (C/F).

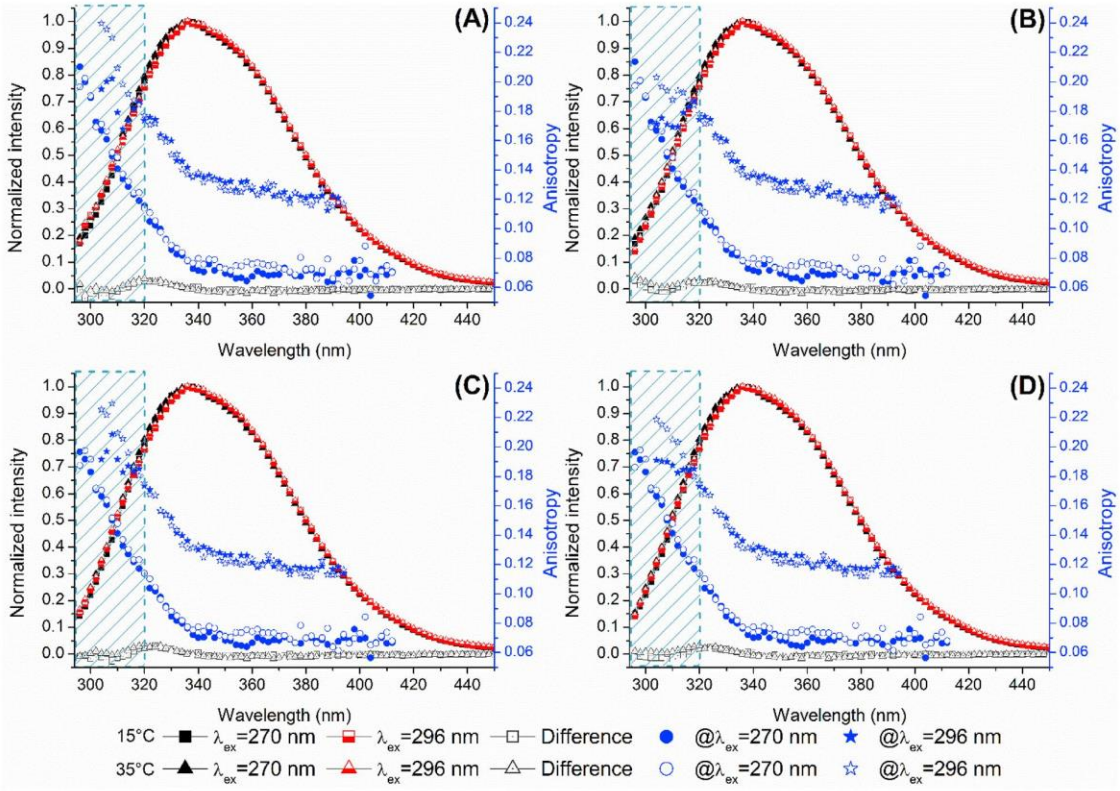

Fig. 3. Normalized EEM_T emission spectra of rIgG excited at λ_{ex} 270 and 296 nm and the difference spectra (= λ_{ex} 270 − λ_{ex} 296) of the rIgG native state corrected with: (A) Preproc1, (B) Preproc1/Interpolated, (C) Preproc2, and (D) Preproc2/Interpolated, at two different temperatures (15° and 35 °C), overlaid with the emission anisotropy at λ_{ex} 270 and 296 nm (blue circles and stars). The shaded boxes represent the spectral region affected by residual noise. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

and in-house written program (FluorS). The EEM_{HV} and EEM_{HH} measurements were only used here to calculate the G-factor ($G = I_{HV}/I_{HH}$, Fig. S2, SI) and thus correct the VH measurement to produce the corrected perpendicular emission [37]. The detector (and emission optics/monochromator) can have a different wavelength dependent sensitivity for vertically and horizontally polarized light and this must be corrected (Fig. S3, SI) over the complete EEM spectrum [15,37]. G factor corrected EEM_{VH} spectra are designated perpendicular polarization ($EEM_{\perp}$) and the EEM_{VV} spectra as parallel polarization ($EEM_{\parallel} = EEM_{VV}$). The total unpolarized EEM spectra (EEM_T) were calculated using $EEM_T = EEM_{\parallel} + 2 \times EEM_{\perp}$, and these served as controls for chemometric modelling [15,37,38].

EEM three-way arrays can be decomposed using bilinear methods such as MCR-ALS [39] or trilinear models like PARAFAC [33]. MCR-ALS offers the advantage to switch between bilinear, partial trilinear, and fully trilinear models [35], however, rotational ambiguities can be associated with the bilinear solutions [22,23,29,30,39,40]. Thus, different constraints and augmentation modes can yield different solutions from the correct solution [30]. PARAFAC was preferred here due to the uniqueness of the solutions, which can be achieved with EEM data following a trilinear model [33] or when deviations from trilinearity are very small and not linked to the signal of the analyte to be determined [41]. Prior to PARAFAC, EEM data were pre-processed to minimize IFE, Rayleigh, and Raman scattering, all of which have very large adverse effects on anisotropy measurement accuracy [37]. The type of procedures used for pre-processing are very data/sample dependent and the sequence of steps could also be critical [33,42,43]. Furthermore, the use of interpolation of IFS spectra also needed to be carefully evaluated as there could be unwanted, unexpected spectral distortions induced. Here, multiple methods and combinations were evaluated for the specific issues associated with IgG emission, and Fig. 1 shows the four best pre-processing sequences implemented.

Raman and Rayleigh first order bands were present in the EEM data (Fig. 2) and as scattered light [15] is non-trilinear, it should be removed prior to PARAFAC analysis. Raman scatter was removed by subtracting a blank (PBS buffer) spectrum from the pEEM measurements (Figure S4C/S5B, SI). The first order Rayleigh scatter was modelled and corrected (Fig. 2 A/D) as previously described [26,44].

This EEM data required IFE correction because of the high optical density ($A_{280} = 0.32 \pm 0.05$, 2 mm pathlength, Fig. S6, SI) of the protein solutions [45,46]. An absorbance-based correction (Equation (1)) [15] was used: the measured absorbance (A_λ) at each excitation (λ_{ex}) and emission (λ_{em}) wavelength combination was used to convert the observed fluorescence intensity (F^{obs}) to a corrected fluorescence intensity (F^{corr}):

$$F^{corr}_{\lambda ex,\lambda em} = F^{obs}_{\lambda ex,\lambda em} \times 10^{\left(\frac{A_{\lambda ex}+A_{\lambda em}}{2}\right)}, \text{ if } F^{obs}_{\lambda ex,\lambda em} > LOR\, F_{\lambda em,\lambda ex} \quad (1)$$

Only the spectral co-ordinates which had fluorescence intensities that were above the limit of reporting ($LOR\, F_{\lambda em,\lambda ex} = F_{blank(\lambda em,\lambda ex)} + 10 \times SD(F_{blank(\lambda em,\lambda ex)})$) [45] were corrected. The LOR at each $\lambda_{em}/\lambda_{ex}$ coordinate was calculated for each polarized measurement from the standard deviation of 10 blank replicate measurements (Fig. S7, SI). IFE correction was critical here for PARAFAC resolution and anisotropy calculations because the polarized measurements were much weaker than normal, unpolarized EEM measurements [25].

IFE correction introduced some artefacts related to amplified noise in the $\lambda_{em} < 292$ nm region where emission was weak. To solve this, the spectra were edited to eliminate this region, yielding new data, with

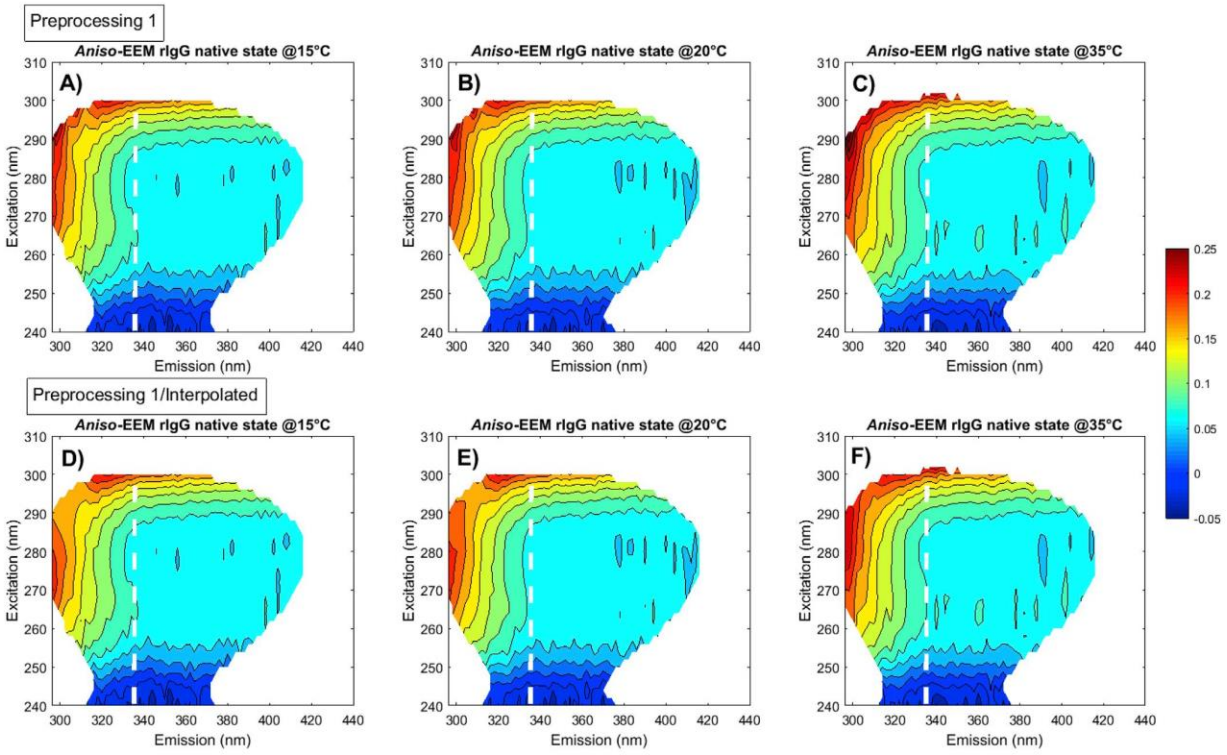

Fig. 4. *Aniso*-EEM maps for rIgG measured at (A/D) 15 °C, (B/E) 20 °C, and (C/F) 35 °C corrected with (Top) Preproc1 and (Bottom) Preproc1/Interp. The dashed lines show the $\lambda_{em} = 336$ nm excitation line. The colour bar on the right represents the anisotropy scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

$\lambda_{em} \geq 296$ nm (Fig. 2B/E). This removed area comprised mostly of weak Tyr emission and the Trp emission blue edge (at short wavelengths) [15,34]. The other pre-processing steps were to remove the residual noise in the Rayleigh scatter region from the EEM by interpolation [43] and to reduce unwanted noise using Savitzky-Golay smoothing using a second-order polynomial with a 15-point window size (Fig. 2C/F) [48]. These pre-processed pEEM spectra were used to calculate the corresponding anisotropy (r) at each $\lambda_{ex}/\lambda_{em}$ coordinate (Equation (2)) [37]. This was then used to generate a multidimensional data matrix ($\lambda_{ex} \times \lambda_{em} \times r$) over the full emission space (*aniso*-EEM map):

$$\overline{r}_{(\lambda ex,\lambda em)} = \frac{EEM_{\parallel\ (\lambda ex,\lambda em)} - EEM_{\perp(\lambda ex,\lambda em)}}{EEM_{\parallel\ (\lambda ex,\lambda em)} + 2 \times EEM_{\perp(\lambda ex,\lambda em)}} \quad (2)$$

For PARAFAC, the pre-processed EEM_T, and $EEM_{\parallel}/EEM_{\perp}$ were arranged in a three-dimensional structure (X) of size, 27 samples × 78 λ_{em} × 41 λ_{ex} (Fig. 2B/E), and normalized by peak maximum [31]. The datasets were normalized to remove intensity differences due to small concentration differences during solution preparation and day-to-day fluctuations in excitation light intensity. Normalization focuses the model on purely structural variations of the protein rather than the signal, magnitude. This should increase the probability that the weaker fluorophore contributions might be resolved. In any case, the PARAFAC models for the non-normalized and normalized datasets were the same, but with a slightly improvement in model validation (especially RSD of the score values) for the normalized datasets. Several methods were used for determining the correct number of PARAFAC factors: CORe CONsistency DIAgnostic test (CONCORDIA) [49], the percentage of variance explained by the model (Table S1, SI), and visual inspection of the recovered spectral profile and residuals. Non-negative constraints for all modes (samples and both emission and excitation profiles) were applied and PARAFAC initialization was made by selecting the best-fitting model from several trial models fitted using a low number of iterations. Validation of spectral deconvolution results were performed using split-half analysis [33].

First, the $EEM_{\parallel}$, $EEM_{\perp}$, and EEM_T datasets were analysed by PARAFAC to see if: (a) the different pre-processing schemes had an impact; (b) the use of interpolation distorted the data; (c) different fluorophores could be resolved; and (d) if there were differences in the recovered loadings (spectra) for the differently polarized emission. This was a long and complex iterative process to carefully select the best pre-processing method for PARAFAC analysis of proteins which unlike the small molecule case [35], there is no *a-priori* knowledge about the precise component profile to be recovered other than empirical knowledge of the peak positions for Tyr and Trp emission.

3. Results and discussion

3.1. 2D spectral analysis

2D emission wavelength plots were used to quickly assess the degree of spectral variation generated over the temperature range and determine if there were significant differences in Trp and Tyr emission, and Tyr-to-Trp FRET [15,50]. The difference between the emission spectra excited at 270 (Trp and Tyr excited) and 296 nm (mainly Trp excited) should represent the directly excited Tyr emission (i.e. the fraction which does not undergo FRET with the numerous Trp acceptors) [51–53]. The gross difference plots showed some evidence for weak Tyr emission overlapped with Trp emission (λ_{em} ~320 nm), in the unpolarized EEM_T datasets (Fig. 3).

There was minimal change in spectral profile over the 15−35 °C temperature range, with a relative standard deviation (RSD) of <6.1% and <5.7% for Preproc1 and Preproc2, respectively. This confirmed that

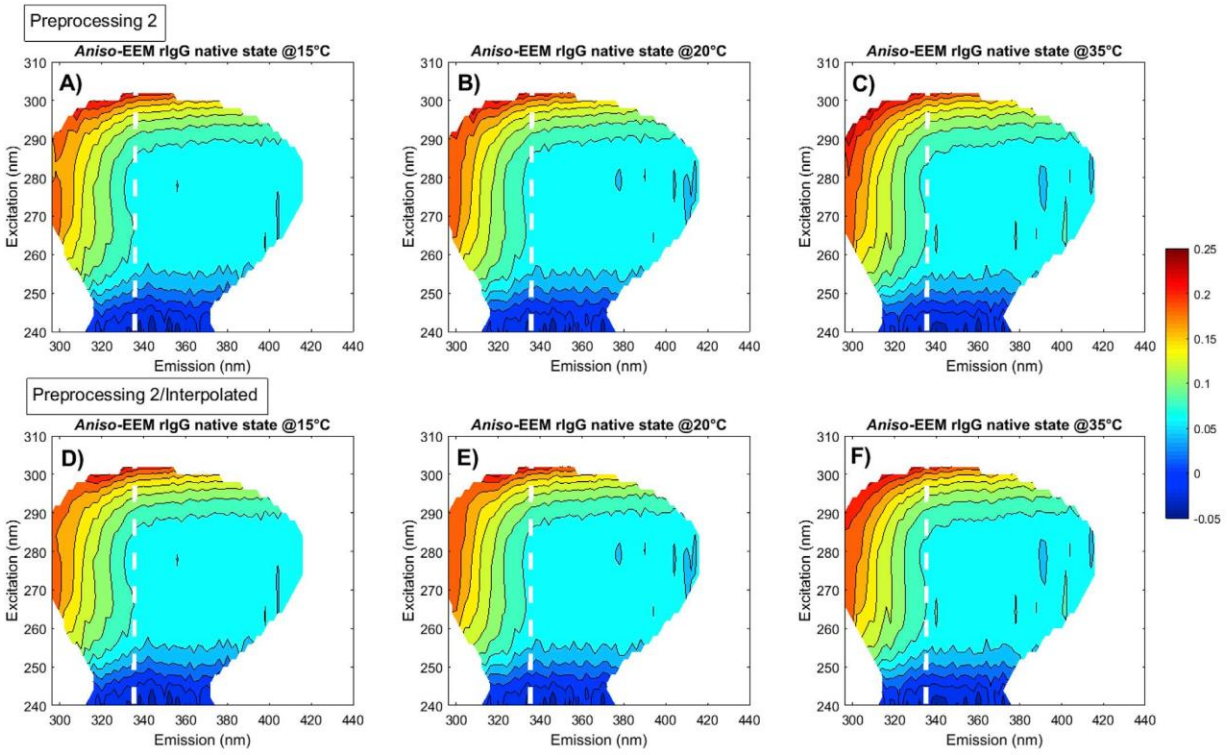

Fig. 5. *Aniso*-EEM maps for rIgG measured at (A/D) 15 °C, (B/E) 20 °C, and (C/F) 35 °C corrected with (top) Preproc2 and (bottom) Preproc2/Interp. The dashed lines show the $\lambda_{em} = 336$ nm excitation line. The colour bar on the right represents the anisotropy scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the rIgG structure did not change significantly and thus could be considered to be in its native form throughout this temperature range. The anisotropy plots did not change much with the different pre-processing methods (Fig. 3), however there was a small change between the anisotropy values at the blue edge, with an RSD <3.8% and <7.4% for the anisotropy values with λ_{ex} 270 and 296 nm, respectively. This variance in anisotropy at the blue edge ($\lambda_{em} < 320$ nm) was related to residual noise from the Rayleigh light scatter. Preproc2 (Fig. 3C) better removed the artefact created by IFE correction compared to Preproc1 (Fig. 3A), but residual noise was still present as shown by the high anisotropy values. The use of interpolation removed this residual noise without affecting the anisotropy outside of the scattering affected zones (Fig. 3B/D). However, there were noticeable differences with the 270 nm excitation, which means that the values in this region were suspect (*vide infra*).

3.2. Aniso-EEM maps

However, the use of single excitation wavelength spectra omits a lot of potentially useful information and while the selection of excitation wavelengths at 270 and 296 nm showed changes at the blue edge, it might miss subtle red edge (at longer wavelengths) effects for example [54]. This is the rationale for examining the full emission space. The *aniso*-EEM maps for rIgG were heterogeneous with the variation across the emission space being generated by multiple factors such as: the type and number of fluorophores, fluorophore location within the protein, differences in fluorophore motility, variations in intra-molecular FRET, and differences in the physicochemical environments of the fluorophores. A visual inspection of the *aniso*-EEM maps for the different pre-processing methods (Fig. 4/5) at various temperatures showed a small change for $\lambda_{em} > 336$ nm. Neither pre-processing method affected the Trp emission region (λ_{em}~330–350 nm) however, visual inspection of the *aniso*-EEM maps in the shorter emission wavelengths ($\lambda_{em} < 336$ nm) showed a clear difference between the pre-processing methods. The Preproc1 corrected *aniso*-EEM maps (Fig. 4A−C) showed a more irregular contour compared to those generated using Preproc2 (Fig. 5A−C). This was due to the residual shot noise from Rayleigh scatter being amplified in the data corrected using Preproc1. The maximum anisotropy values were higher and more variable for Preproc1 than Preproc2 (Fig. S8, SI), with an RSD of 8.90% and 3.88% (for triplicate measurements at all temperatures). We can conclude therefore that Preproc2 (Fig. 2E) was more efficient removing residual scattered light and did not amplify noise and IFE correction artefacts as much as Preproc1 (Fig. 2B). Therefore one should implement the subtractive elements (Raman and Rayleigh corrections) prior to multiplicative elements (IFE correction) otherwise one gets increased noise contributions. This issue is particularly important in the UV ($\lambda < 250$ nm) where the fluorescence emission signal is weak.

Incomplete light scatter removal produces larger anisotropy values (higher $EEM_{\parallel}$ intensities). To try and completely remove the residual scatter, interpolation was used. The inspection of the *aniso*-EEM maps for Preproc1/Interp. showed a reduction in the contour irregularity and lower anisotropy values in the scatter area (RSD = 4.90%, for triplicate measurements at all temperatures) in comparison with the non-interpolated Preproc1 (Fig. 4). The differences between the *aniso*-EEM maps for the Preproc2/Interp. (RSD = 3.15%, triplicate measurements at all temperatures) and non-interpolated Preproc2 (Fig. 5) were less significant. The maximum anisotropy values (measured over the whole map) decreased between the non-interpolated and the interpolated pre-processed cases (Fig. S8, SI). These differences in maximum anisotropy values between the Preproc1 and Preproc1/Interp. indicated that interpolation was efficient at removing the residual noise. The small

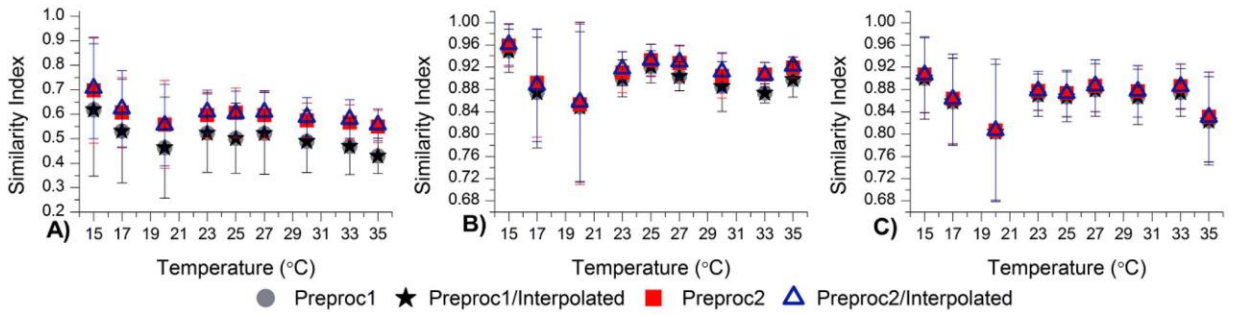

Fig. 6. *Sim*I values calculated between the *aniso*-EEM for each pre-processing method for: (A) the full emission space, (B) the area related to Tyr emission ($\lambda_{ex}/\lambda_{em}$ = 270–290/300−320 nm); and (C) the area related to Trp ($\lambda_{ex}/\lambda_{em}$ = 280–300/320−360 nm) emission. A *Sim*I value of one indicates identical maps. Error bars generated from the standard deviation from triplicate measurements of the independent samples.

differences between the maximum anisotropy values of Preproc2 and Preproc2/Interp. showed that this pre-processing method was more efficiently removing Rayleigh scatter and noise from the data.

However, care must be exercised here because proving that interpolation does not distort the true emission is not easy in FRET dominated systems like proteins. In the ideal small molecule case where no FRET occurs then it is possible to verify that the corrected spectra matches the individual components or mixtures thereof. Therefore, one has to ensure that whichever pre-processing/interpolation method is selected that this is fixed throughout the analysis, and that the anisotropy values in these regions be understood to be estimates rather than accurate values.

Interpolation improved scattering/noise removal as confirmed by the reduction in maximum anisotropy values (Fig. S8, SI). The mean anisotropy values over the full emission space did not change with the different pre-processing methods: an RSD of 7.67% and 7.04% for Preproc1 and Preproc2, and 7.79% and 7.03% for Preproc1/Interp. and Preproc2/Interp. (for triplicate measurements at all temperatures). This indicated that overall, the degree of structural change was small and as there were no differences in the mean anisotropy between pre-processing with and without interpolation, one might assume that the interpolation did not affect the data. However, the use of mean anisotropy values for this type of assessment is not correct as the anisotropy intrinsically varies over the full emission space and thus alternative assessment methods which can quantify the degree of change were required.

A similarity index (*Sim*I) [55] was calculated (using the *aniso*-EEM map from the first replicate measurement at 15 °C as the reference spectrum) for all the samples (Fig. 6). This showed that the *aniso*-EEM maps using Preproc2 (RSD = 22.06%), Preproc2/Interp. (RSD = 21.50%) were more similar to each other than the *aniso*-EEM maps for Preproc1 and Preproc1/Interp. (RSD = 35.71%) which can be largely ascribed to the noise contamination arising fro Rayleigh light scatter.

The small dip at 20 °C (Fig. 6) is a real effect since the data were collected on three different days, at approximately the same time using identical measurement settings, and each stock solution was prepared from a different source vial (but with the same lot number). This dip (and increase in the error bars) was possibly due to a change in local motion at 20 °C, but at present we have no clear evidence to support this [56,57]. A reduction in flexibility could cause measurable changes in polarization, but probably not much in terms of overall emission profile. As the temperature increases above 23 °C up to 35 °C, *Sim*I values remained constant, indicating a stable structure. This requires further investigation as the protein was polyclonal in origin with only ~80% monomer and rabbit IgG composition is known to vary [58].

To better assess the changes in Tyr/Trp emission in the *aniso*-EEM maps (Fig. 4/5) two areas were selected (Fig. S10, SI): Tyr ($\lambda_{ex}/\lambda_{em}$ = 270–290/300−320 nm) and Trp ($\lambda_{ex}/\lambda_{em}$ = 280–300/320−360 nm). Here the variation was considerably smaller (Fig. 6B/C) with very small differences between the *aniso*-EEM maps (RSD~8.60%) for the Trp region. Likewise, the Tyr region did not show any significant differences in the *aniso*-EEM maps (RSD~7.65%). *Sim*I analysis of specific Tyr and Trp emission regions showed that there was no difference in the *aniso*-maps due to the pre-processing methods outside of the Rayleigh scattering region. Overall, the poor discrimination of *aniso*-EEM map analysis can be attributed to the greater errors associated with anisotropy measurements. This coupled with the small structural changes and the relatively high noise spectral measurements implemented here make it a less useful method for assessing small changes in IgG structure.

3.3. PARAFAC

Visual inspection and *Sim*I analysis of *aniso*-EEM maps (Fig. 4/5) suggested that we had observed some Trp emission changes caused by temperature induced variations in the rIgG structure. It was hoped that sufficient spectral variance had been induced so as to enable PARAFAC to resolve individual fluorophore emissions from the rIgG native state in a similar fashion to previously reported for HSA [25]. We know that protein emission does not vary linearly because of FRET and when this occurs then the assumptions of variability, trilinearity, and additivity required [33] for successful and chemically meaningful PARAFAC analysis do not hold. However, for small structural variations it might be reasonable to expect that the structural changes were sufficient to induce significant emission intensity fluctuations (via varying quenching rates for different fluorophores) without significant changes in spectral shape. PARAFAC models were generated from pre-processed $EEM_{\parallel}$, $EEM_{\perp}$, and EEM_T data of all samples (9 temperatures × triplicate measurements). From an rIgG structure assessment (Fig. S1, SI), one might expect to recover at least three components, one from Tyr, a second from Trp located in more hydrophobic environments (λ_{em}~330 nm), and a third one from more solvent exposed Trp (λ_{em} > 340 nm). In addition to environmental factors, Trp emission in IgG is generated via three photochemical processes: *(i)* direct excitation of, and emission from the excited state, should produce emission with a high anisotropy, *(ii)* homo-FRET from other Trp fluorophores, should result in lower anisotropy [60], and *(iii)* hetero-FRET from Tyr, again producing low anisotropy emission. Therefore, one might expect to see this reflected in the presence of one or two additional components with an appreciable score.

Even though rIgG contains ~50 Tyr and ~24 Trp fluorophores, PARAFAC only recovered two components (PFC1 and PFC2) from the EEM data (Table 1, Fig. 7) and the contribution of the second component was very small, and very sensitive to pre-processing. This was largely based more on differences in the excitation spectra. Increasing the component number up to four (Table S1, SI) and analysis of the residuals produced no better fit models. For all the PARAFAC models the first component resolved was virtually identical for all polarizations (*Sim*I = 0.9494, RSD = 1.78%). This coupled with the facts that the emission was relatively depolarized (~0.1 for most of the emission), and that the excitation anisotropy spectra (see below) were distinctive indicated that it originated mostly from FRET rather than direct excitation/

Table 1
Comparison of the model parameters and components obtained for the normalized polarized $EEM_{||}$, $EEM_{\perp}$, and unpolarized EEM_T PARAFAC models of the rIgG native state corrected with the different pre-processing methods.

	Preproc1			Preproc1/Interp.			Preproc2			Preproc2/Interp.		
	$EEM_{||}$	$EEM_{\perp}$	EEM_T	$EEM_{||}$	$EEM_{\perp}$	EEM_T	$EEM_{||}$	$EEM_{\perp}$	EEM_T	$EEM_{||}$	$EEM_{\perp}$	EEM_T
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/336	280/336	278/336	280/336	280/336	278/336	280/336	278/336	280/336	278/336	278/336	278/336
C1 Fit model (%)	99.73	99.70	99.85	98.74	99.71	99.65	99.75	99.74	99.85	99.34	99.75	99.84
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	296/342	296/346	296/346	298/336	296/346	298/342	296/342	296/346	298/346	298/338	296/346	298/346
C2 Fit model (%)	**0.27**	**0.30**	**0.15**	**1.26**	**0.29**	**0.35**	**0.12**	**0.26**	**0.15**	**0.66**	**0.25**	**0.16**
Variance explained (%)	99.97	99.99	99.99	99.98	99.99	99.98	99.98	99.99	99.98	99.98	99.99	99.98
CONCOR DIA	95.22	88.71	97.02	96.92	86.80	69.21	96.84	91.53	96.61	57.96	88.43	95.76
Split-half analysis (%)	99.13	99.67	99.74	98.39	99.74	99.82	99.14	99.64	99.82	99.57	99.67	99.86

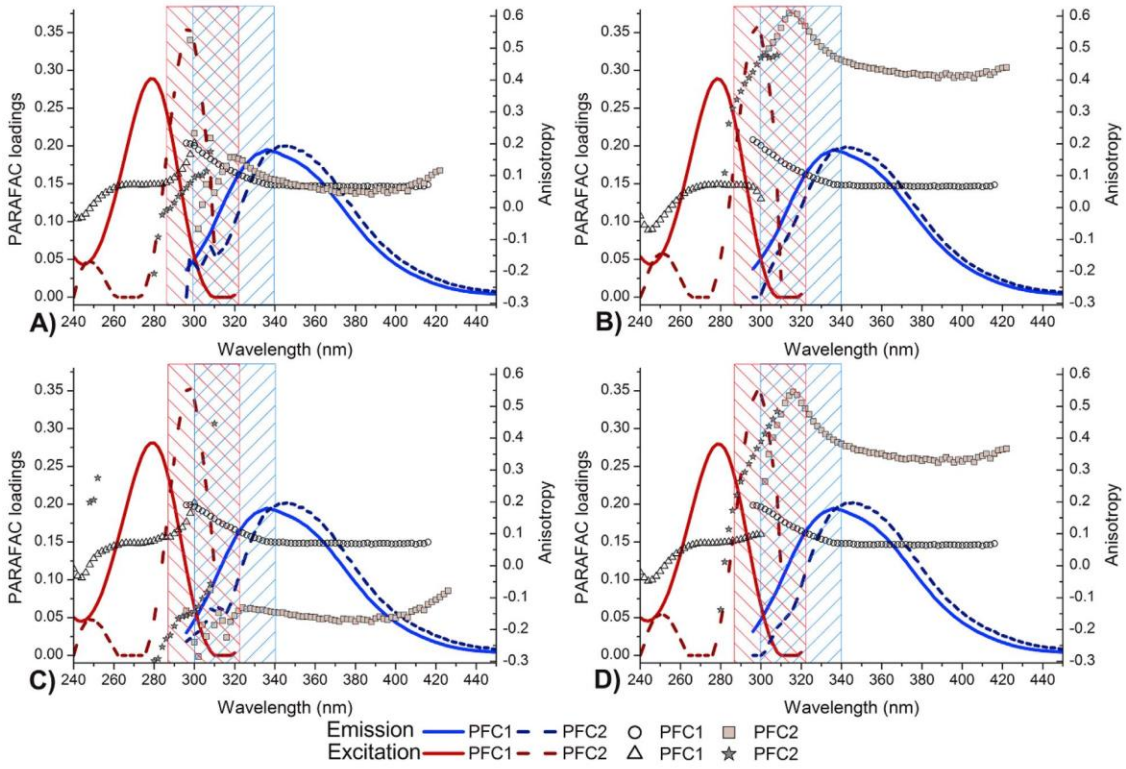

Fig. 7. Comparison of PARAFAC modelling of the rIgG native structure for the EEM_T datasets with the different pre-processing methods. PFC1 and PFC2 (A) Preproc1, (B) Preproc1/Interp., (C) Preproc2, and (D) Preproc2/Interp. emission and excitation profiles recovered for unpolarized EEM_T, overlaid with the component anisotropy recovered at 20 °C. The anisotropy values were calculated for each component at $\lambda_{ex/em} = 280/336$ and 298/346 for PFC1 and PFC2 respectively. The shaded boxes represent the spectral regions most affected by scatter in the EEM.

emission of the fluorophores. There were several reasons why only two components were recovered: *(i)* the structural changes and thus emission changes over the temperature range sampled were too small; *(ii)* the Tyr-to-Trp FRET was very strong and effectively constant over the 15–35 °C temperature range; *(iii)* the emission spectra of the different Trp populations overlapped and the observed emission spectrum represented an average of buried and partially exposed Trp fluorophores (λ_{em}~336 nm).

For the freely diffusing small molecule case, the excitation spectrum of Tyr anisotropy should be positive from $\lambda_{ex} = 260$ nm, with an increase up to 290 nm [61,62]. For Trp, the excitation anisotropy spectra should not be constant, due to the overlapping excitation of the two transition states. A minimum anisotropy value at ~λ_{ex} 290 nm should be observed when the maximum excitation is due to the ^1L_b excited state, and a maximum at ~λ_{ex} 300 nm indicates excitation of the ^1L_a excited state of Trp [15,37,38,61–63]. From an analysis of the loadings it was clear that PFC1 was the same for all pre-processing methods (Fig. 7) and represented a composite signal from both Trp and Tyr fluorophores. In all cases, the excitation spectrum was very similar to that of Tyr [61,62], whereas the emission was that of Trp. Therefore, we can conclude that this component largely represents the hetero-FRET Tyr-to-Trp process, with probably a minor contribution from directly excited Tyr emission at the blue edge of the emission (in the region where the anisotropy increases). The fact that PFC1 is largely due to hetero-FRET would also explain the low calculated anisotropy recovered (~0.05 for $\lambda_{em} > 340$ nm). The slightly higher anisotropy values for the non-interpolated pre-processing (Fig. 7 A/C) EEM PFC1 might be attributed to the presence of noise (i.e. the Rayleigh shot noise) in the spectral data which also impacted on component recovery [15].

The very weak second PARAFAC component, while definitely present, was not clearly resolved, and its properties (e.g. anisotropy) was very sensitive to pre-processing. This makes it hard to unambiguously assign this to any specific emission or use for quantitative assessments.

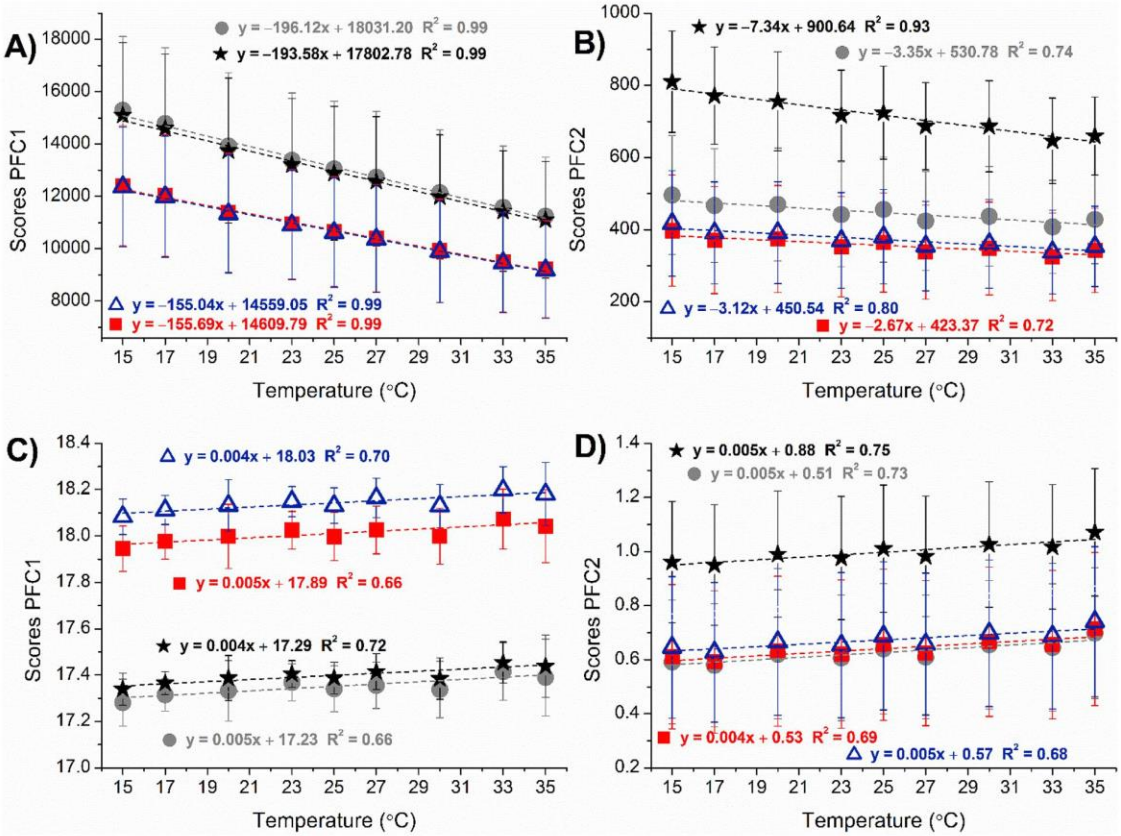

Fig. 8. Scores plots of the two-component PARAFAC models for the non-normalized (top) and normalized (bottom) unpolarized rIgG EEM_T datasets. (Left) PFC1 and (Right) PFC2 scores for the various pre-processing methods used to correct the data. The grey circles and black stars Preproc1/Preproc1-Interpolated respectively, while the red squares and blue triangles are Preproc2/Preproc2-Interpolated respectively. P values for the regression fits in A & B were <0.05 indicating that the changes were statistically significant. For C& D the regression fits had *P*-values of <0.001. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

However, it does look like that PFC2 was related to Trp emission that originated from direct excitation because of the high anisotropy for PFC2 at longer wavelengths and the excitation and emission wavelengths recovered (Table 1). This would suggest that the Trp residues that give rise to this component were most likely located in the hyper variable Fab region. We do have to note however that interpolation had a big impact on the recovered anisotropy at short emission wavelengths (Fig. 7) because of the effects on $EEM_{\parallel}$ (Table 1). Interpolation generated a ~5-fold increase in the PFC2 scores, which can be ascribed to the fact that the scatter signal in $EEM_{\parallel}$ was much greater than $EEM_{\perp}$ or EEM_T. This means that the remaining shot noise was very significant and was indistinguishable from emission and was thus included in the data for PARAFAC. This then also contributed to the increased anisotropy in these regions and a lot of the variability in the *aniso*-EEM maps. This meant that the component anisotropy recovered for <340 nm was unreliable, but at the longer emission wavelengths it should be a sensible value.

Because the differences between the different polarized datasets were all small and with low scores (Fig. 8, and Fig. S12, SI), and since the data was relatively noisy, this limited our ability to identify any real photophysical or structural changes. However, if the PARAFAC modelling process was robust one might expect that the results using the different pre-processing methods should yield the same result within reason. Indeed, all four pre-processing methods yielded virtually the same PARAFAC solutions (Fig. S11, SI), with differences in the model validation (Table 1). Most importantly, all the significant differences in PARAFAC loadings were within the interpolated area (Fig. 7). For the normalized unpolarized EEM_T data of both PARAFAC models, the component scores showed no change with temperature (Fig. 8, and Fig. S12, SI) indicating that there were no significant change in spectral shape (and thus structure). For the non-normalized EEM_T data, the scores decreased linearly for both components indicating that the major change in emission properties was due to simple fluorescence quenching due to increasing thermal motion [11,17,64].

4. Conclusions

For rIgG the fluorescence changes over the 15–35 °C temperature range were mostly attributable to simple emission quenching with very little change in spectral shape. This indicated that the native state structure was preserved and thus there were minimal changes in FRET. In the rIgG native state, most Trp fluorophores are located within the structure (higher quantum yield, more blue emitting) with only a few being partially solvent exposed (lower quantum yield, more red emitting) [15]. This would result in the majority of Trp emission coming from these internal Trp residues and therefore we might not be able to discriminate the different Trp populations. Likewise, with Tyr, the lower quantum yields and extensive Tyr-to-Trp FRET will reduce emission intensity to such an extent that one might not be able to resolve the Tyr emission. This was the case here and PARAFAC only recovered one major and one minor component. We suggest that the observed component discrimination was largely based on the difference in Trp excitation mechanisms that is, between direct excitation/emission (including homo-FRET), and

that originating via energy transfer from Tyr [15,65,66]. However, the fit of the second component were low (<1.3%) and very sensitive to the precise sequence of data pre-processing steps.

The inability to extract any more components was due to a combination of factors. First, there were too many Trp fluorophores (>20) in different positions within the protein, and this generated a distribution of overlapping emission spectra rather than discrete spectra which is a problem for PARAFAC (or MCR). Second, because the protein structure did not change significantly, there was no major changes in Trp environment which kept the Tyr-to-Trp FRET rates constant and thus there were minimal changes in component ratios. Third, despite the larger number of Tyr residues present, its lower quantum yield, and the large Tyr-to-Trp energy transfer rates reduced Tyr emission intensity to very low levels [15,50]. Fourth, the use of UV transparent wire grid polarizers [26] ensured that the emission data recorded was different to the red-edge excited TSFS data collected previously for HSA [25], and this may also have some effect on component resolution. Finally, as the only significant measured changes were thermal quenching related and were apparently the same for all fluorophores, this produced a high degree of co-linearity, thus no mechanism to break the linear dependencies, and so no additional components.

For the rIgG EEM data, for all the correction methods implemented, residual Rayleigh scattering (probably the shot noise), and artefacts induced by IFE correction caused significant issues, and in cases made the PARAFAC results less reliable. It was better to implement subtraction-based corrections (blank subtraction and scattering correction) and smoothing before any multiplication-based steps (IFE correction). The use of interpolation in a small spectral region had minimal impact on the components recovered and was shown to apparently improve some aspects of PARAFAC modelling. However, it was obvious that overlap between the emission blue edge with residual noise (shot noise from the Rayleigh scatter) was a critical factor (Figs. S11 and SI). This can be easily explained once we take into consideration the 10 nm slit widths of used here, and the fact that the scatter became much stronger for $\lambda_{ex} > 280$ nm, yielding a heavily scatter contaminated $\lambda_{ex}/\lambda_{em} = 280–320/290–330$ nm region.

The presence of relatively high shot noise signal associated with this strong Rayleigh scatter in the parallel polarization measurements caused problems because after correction there was a relatively large (compared to other regions in the EEM) shot noise signal that had an average signal greater than the baseline and thus was indistinguishable from fluorescence. This distorted the $EEM_{\|}$ spectra which in turn affected anisotropy values in this region and PARAFAC modelling (Table 1). Thus, the relatively high noise signal intensities, combined with Tyr and Trp emission overlap made it difficult to accurately recover the emission blue edge. It was clear that if we want to more accurately resolve emission components from EEM data then we need either to collect data with much less Rayleigh scatter or use a denoising method that can handle the heterogenous noise pattern in the EEM space. The first can be achieved using more expensive spectrometers with better stray light rejection (double emission monochromators) and/or the use of smaller emission slits, neither of which may not always be practical or feasible with the standard fluorimeters found in most laboratories.

In conclusion, this study has provided a starting point from which we can explore the use of PARAFAC for the analysis of larger structural changes in IgG type proteins such as unfolding and aggregation. In particular the fact that PARAFAC potentially discriminated Trp emission on the basis of whether or not it was directly excited or excited via FRET from Tyr is potentially very useful for explaining changes in IgG emission spectral profiles.

5. Supplemental information available

Supporting information is available which provides additional spectral data further details on the chemometric analysis.

Acknowledgements

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland and is co-funded under the European Regional Development Fund under Grand number (14/IA/2282, *Advanced Analytics for Biological Therapeutic Manufacture*, to AGR). We also thank Agilent Technologies (Mulgrave Victoria, Australia) for the loan of a fluorescence spectrometer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2018.12.007.

References

[1] J.M. Reichert, Marketed therapeutic antibodies compendium, mAbs 4 (2012) 413–415.
[2] D.M. Ecker, S.D. Jones, H.L. Levine, The therapeutic monoclonal antibody market, mAbs 7 (2015) 9–14.
[3] K.M. Murphy, C. Weaver, Janeway's Immunobiology, ninth ed., Garland Science/Taylor & Francis Group, LLC., New York, NY, 2016.
[4] J.G. Salfeld, Isotype selection in antibody engineering, Nat. Biotechnol. 25 (2007) 1369–1372.
[5] P.-P. Pastoret, Handbook of Vertebrate Immunology, San Diego; London, Academic Press, San Diego ; London, 1998.
[6] A. Pinheiro, F. Neves, A.L. de Matos, J. Abrantes, W. van der Loo, R. Mage, P.J. Esteves, An overview of the lagomorph immune system and its genetic diversity, Immunogenetics 68 (2016) 83–107.
[7] K.L. Knight, R.C. Burnett, J.M. McNicholas, Organization and polymorphism of rabbit immunoglobulin heavy-chain genes, J. Immunol. 134 (1985) 1245–1250.
[8] R.M. Ionescu, J. Vlasak, C. Price, M. Kirchmeier, Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies, J. Pharmacol. Sci. 97 (2008) 1414–1426.
[9] P. Arosio, S. Rima, M. Morbidelli, Aggregation mechanism of an IgG2 and two IgG1 monoclonal antibodies at low pH: from oligomers to larger aggregates, Pharm. Res. 30 (2013) 641–654.
[10] B. Sharma, Immunogenicity of therapeutic proteins. Part 1: impact of product handling, Biotechnol. Adv. 25 (2007) 310–317.
[11] E. Cerasoli, J. Ravi, T. Garfagnini, S. Gnaniah, D. le Pevelen, G.E. Tranter, Temperature denaturation and aggregation of a multi-domain protein (IgG1) investigated with an array of complementary biophysical methods, Anal. Bioanal. Chem. 406 (2014) 6577–6586.
[12] C.H. Li, X. Nguyen, L. Narhi, L. Chemmalil, E. Towers, S. Muzammil, J. Gabrielson, Y. Jiang, Applications of circular dichroism (CD) for structural analysis of proteins: qualification of near- and far-UV CD for protein higher order structural analysis, J. Pharmacol. Sci. 100 (2011) 4642–4654.
[13] B. Ranjbar, P. Gill, Circular dichroism techniques: biomolecular and nanostructural analyses- a review, Chem. Biol. Drug Des. 74 (2009) 101–120.
[14] S. Zolls, R. Tantipolphan, M. Wiggenhorn, G. Winter, W. Jiskoot, W. Friess, A. Hawe, Particles in therapeutic protein formulations, Part 1: overview of analytical methods, J. Pharmacol. Sci. 101 (2012) 914–935.
[15] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, third ed. ed., Springer, New York, 2006.
[16] K. Wiberg, A. Sterner-Molin, S.P. Jacobsson, Simultaneous determination of albumin and immunoglobulin G with fluorescence spectroscopy and multivariate calibration, Talanta 62 (2004) 567–574.
[17] P. Garidel, M. Hegyi, S. Bassarab, M. Weichel, A rapid, sensitive and economical assessment of monoclonal antibody conformational stability by intrinsic tryptophan fluorescence spectroscopy, Biotechnol. J. 3 (2008) 1201–1211.
[18] L. Liu, L.J. Braun, W. Wang, T.W. Randolph, J.F. Carpenter, Freezing-induced perturbation of tertiary structure of a monoclonal antibody, J. Pharmacol. Sci. 103 (2014) 1979–1986.
[19] I.M. Warner, G.D. Christian, E.R. Davidson, J.B. Callis, Analysis of multicomponent fluorescence data, Anal. Chem. 49 (1977) 564–573.
[20] D. Patra, A.K. Mishra, Recent developments in multi-component synchronous fluorescence scan analysis, Trac. Trends Anal. Chem. 21 (2002) 787–798.
[21] B. Li, P.W. Ryan, M. Shanahan, K.J. Leister, A.G. Ryder, Fluorescence excitation-emission matrix (EEM) spectroscopy for rapid identification and quality evaluation of cell culture media components, Appl. Spectrosc. 65 (2011) 1240–1249.
[22] B. Li, M. Shanahan, A. Calvet, K.J. Leister, A.G. Ryder, Comprehensive, quantitative bioprocess productivity monitoring using fluorescence EEM spectroscopy and chemometrics, Analyst 139 (2014) 1661–1671.
[23] K. Kumar, A.K. Mishra, Analysis of dilute aqueous multifluorophoric mixtures using excitation-emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy: a comparative evaluation, Talanta 117 (2013) 209–220.
[24] K. Ohadi, R.L. Legge, H.M. Budman, Intrinsic fluorescence-based at situ soft sensor for monitoring monoclonal antibody aggregation, Biotechnol. Prog. 31 (2015) 1423–1432.

[25] R.C. Groza, B. Li, A.G. Ryder, Anisotropy resolved multidimensional emission spectroscopy (ARMES): a new tool for protein analysis, Anal. Chim. Acta 886 (2015) 133–142.
[26] Y. Casamayou-Boucau, A.G. Ryder, Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters, validation standards, and Rayleigh scatter removal methods, Methods Appl. Fluoresc. 5 (2017) 037001.
[27] S.D. Brown, L.A. Sarabia, J. Trygg, Comprehensive Chemometrics Chemical and Biochemical Data Analysis, Elsevier, Amsterdam, Amsterdam, 2009.
[28] J.M. Amigo, F. Marini, Multiway methods, in: F. Marini (Ed.), Chemometrics in Food Chemistry, Elsevier Science Bv, Amsterdam, 2013, pp. 265–313.
[29] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: a review of advanced and tailored applications and challenges, Anal. Chim. Acta 765 (2013) 28–36.
[30] H. Abdollahi, R. Tauler, Uniqueness and rotation ambiguities in multivariate curve resolution methods, Chemometr. Intell. Lab. Syst. 108 (2011) 100–111.
[31] R. Bro, PARAFAC. Tutorial and applications, chemometrics intellig, Lab. Syst. 38 (1997) 149–171.
[32] C.M. Andersen, R. Bro, Practical aspects of PARAFAC modeling of fluorescence excitation-emission data, J. Chemom. 17 (2003) 200–215.
[33] K.R. Murphy, C.A. Stedmon, D. Graeber, R. Bro, Fluorescence spectroscopy and multi-way techniques. PARAFAC, Anal. Methods 5 (2013) 6557–6566.
[34] R.C. Groza, A. Calvet, A.G. Ryder, A fluorescence anisotropy method for measuring protein concentration in complex cell culture media, Anal. Chim. Acta 821 (2014) 54–61.
[35] Y. Casamayou-Boucau, A.G. Ryder, Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR), Anal. Chim. Acta 1000 (2018) 132–143.
[36] R.C. Groza, Anisotropy Resolved Multi-dimensional Emission Spectroscopy (ARMES): a New Tool for the Quantitative and Structural Analysis of Proteins, PhD Thesis, School of Chemistry, National University of Ireland Galway, Galway, Ireland, 2016.
[37] M. Ameloot, M. vandeVen, A.U. Acuna, B. Valeur, Fluorescence anisotropy measurements in solution: methods and reference materials (IUPAC Technical Report), Pure Appl. Chem. 85 (2013) 589–608.
[38] M.R. Eftink, L.A. Selvidge, P.R. Callis, A.A. Rehms, Photophysics of indole-derivatives - experimental resolution of La and Lb transition and comparison with theory, J. Phys. Chem. 94 (1990) 3469–3479.
[39] K. Kumar, A.K. Mishra, Multivariate curve resolution alternating least square (MCR-ALS) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets: comparing certain ways of arranging TSFS-based three-way array, Chemometr. Intell. Lab. Syst. 147 (2015) 66–74.
[40] K. Kumar, A.K. Mishra, Parallel factor (PARAFAC) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets in excitation-emission matrix fluorescence (EEMF) layout: certain practical aspects, Chemometr. Intell. Lab. Syst. 147 (2015) 121–130.
[41] S. Elcoroaristizabal, A. de Juan, J.A. Garcia, N. Durana, L. Alonso, Comparison of second-order multivariate methods for screening and determination of PAHs by total fluorescence spectroscopy, Chemometr. Intell. Lab. Syst. 132 (2014) 63–74.
[42] M. Bahram, R. Bro, C. Stedmon, A. Afkhami, Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation, J. Chemom. 20 (2006) 99–105.
[43] S. Elcoroaristizabal, R. Bro, J.A. Garcia, L. Alonso, PARAFAC models of fluorescence data with scattering: a comparative study, Chemometr. Intell. Lab. Syst. 142 (2015) 124–130.
[44] A. Rinnan, K.S. Booksh, R. Bro, First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes, Anal. Chim. Acta 537 (2005) 349–358.
[45] D.N. Kothawala, K.R. Murphy, C.A. Stedmon, G.A. Weyhenmeyer, L.J. Tranvik, Inner filter correction of dissolved organic matter fluorescence, Limnol Oceanogr. Methods 11 (2013) 616–630.
[46] M. Tarai, A.K. Mishra, Inner filter effect and the onset of concentration dependent red shift of synchronous fluorescence spectra, Anal. Chim. Acta 940 (2016) 113–119.
[48] A. Savitzky, M.J.E. Golay, Smoothing + differentiation of data by simplified least squares procedures, Anal. Chem. 36 (1964) 1627–&.
[49] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemom. 17 (2003) 274–286.
[50] R. Boteva, T. Zlateva, V. DorovskaTaran, A. Visser, R. Tsanev, B. Salvato, Dissociation equilibrium of human recombinant interferon gamma, Biochemistry 35 (1996) 14825–14830.
[51] J. Eisinger, B. Feuer, A.A. Lamola, Intramolecular singlet excitation transfer. Applications to polypeptides, Biochemistry 8 (1969) 3908-+.
[52] J. Eisinger, Intramolecular energy transfer in adrenocorticotropin, Biochemistry 8 (1969) 3902–&.
[53] H.C. Chiu, R. Bersohn, Electronic-energy transfer between tyrosine and tryptophan in peptides trp-(pro)N-Tyr, Biopolymers 16 (1977) 277–288.
[54] A.P. Demchenko, The red-edge effects: 30 years of exploration, Luminescence 17 (2002) 19–42.
[55] B. Li, A.G. Ryder, Similarity Index: a Rapid Classification Method for Multivariate Data Arrays, 2011. US7983874B2.
[56] T.J. Kamerzell, J.D. Ramsey, C.R. Middaugh, Immunoglobulin dynamics, conformational fluctuations, and nonlinear elasticity and their effects on stability, J. Phys. Chem. B 112 (2008) 3240–3250.
[57] K. Lim, D.M. Jameson, C.A. Gentry, J.N. Herron, Molecular-dynamics of the antifluorescein-4-4-20 antigen-binding fragment. 2. Time-resolved fluorescence spectroscopy, Biochemistry 34 (1995) 6975–6984.
[58] D.C. Hanson, J. Yguerabide, V.N. Schumaker, Segmental flexibility of immunoglobulin-G antibody molecules in solution - a new interpretation, Biochemistry 20 (1981) 6842–6852.
[60] L.W. Runnels, S.F. Scarlata, Theory and applications of fluorescence homotransfer to melittin oligomerization, Biophys. J. 69 (1995) 1569–1583.
[61] G. Weber, Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds, Biochem. J. 75 (1960) 335–345.
[62] G. Weber, Fluorescence-polarization spectrum and electronic-energy transfer in proteins, Biochem. J. 75 (1960) 345–352.
[63] B. Valeur, G. Weber, Resolution of fluorescence excitation spectrum of indole into 1La and 1Lb excitation bands, Photochem. Photobiol. 25 (1977) 441–444.
[64] R.J. Robbins, G.R. Fleming, G.S. Beddard, G.W. Robinson, P.J. Thistlethwaite, G.J. Woolfe, Photophysics of aqueous tryptophan - pH and temperature effects, J. Am. Chem. Soc. 102 (1980) 6271–6279.
[65] E.A. Burstein, N.S. Vedenkina, M.N. Ivkova, Fluorescence and location of tryptophan residues in protein molecules, Photochem. Photobiol. 18 (1973) 263–279.
[66] Y.K. Reshetnyak, E.A. Burstein, Decomposition of protein tryptophan fluorescence spectra into log-normal components. II. The statistical proof of discreteness of tryptophan classes in proteins, Biophys. J. 81 (2001) 1710–1734.

Chemometrics and Intelligent Laboratory Systems 194 (2019) 103871

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Using polarized Total Synchronous Fluorescence Spectroscopy (pTSFS) with PARAFAC analysis for characterizing intrinsic protein emission

Marina Steiner-Browne, Saioa Elcoroaristizabal, Alan G. Ryder *

Nanoscale BioPhotonics Laboratory, School of Chemistry, National University of Ireland, Galway, H91CF50, Galway, Ireland

ARTICLE INFO

Keywords:
Immunoglobulin G
Fluorescence
Multidimensional
Spectroscopy
Anisotropy
PARAFAC

ABSTRACT

Using polarized Excitation Emission Matrix (pEEM) spectroscopy to measure the intrinsic emission of proteins offers a potentially useful methodology for a wide variety of potential applications. However, the presence of Rayleigh light scatter causes significant problems when attempting to use Parallel Factor (PARAFAC) and for anisotropy calculations. The use of polarized Total Synchronous Fluorescence Spectroscopy (pTSFS) can minimize Rayleigh scatter and avoid the use of complex data correction methods. Here, we investigated for the first time the use of pTSFS and PARAFAC to analyze the intrinsic emission of an Immunoglobulin (IgG) type protein in its native state. To enable PARAFAC analysis however, TSFS data (which is not trilinear) must first be transformed into an EEM like layout (t-EEM) and this generated a region with no experimentally acquired information (<8%). Here we critically evaluated several data handling methods and determined that interpolation was the best solution for dealing with the spectral regions with no experimentally acquired data at the blue edge of the emission.

There were only subtle structural changes measured over the temperature range (15–35 °C) analyzed and PARAFAC only resolved two emitting components. A Trp emission component (average signal from all Trp present) which represented >92% of the explained variance, and a much weaker, mostly Tyr related emission with ~3% of the explained variance. The recovery of this Tyr component was only possible because pTSFS measurements were less contaminated by Rayleigh scattering. Changes in Tyr-to-Trp energy transfer rates caused by thermal motion were detected as an increase in Tyr contribution, which could not be resolved with the equivalent pEEM measurements due to light scatter contamination. The increased selectivity, sensitivity, and reproducibility of pTSFS measurements shows that this is a better option than pEEM for fluorescence emission based monitoring of protein structural change or lot-to-lot variance of IgG type proteins.

1. Introduction

The use of intrinsic fluorescence spectroscopy (IFS) to characterize protein structural change in solution [1] is potentially very useful for the characterization of therapeutic proteins such as the Immunoglobulin G (IgG). IgG type proteins have many features which make them attractive therapeutic agents [2,3], however, efficacy is very dependent on both tertiary and quaternary structure [4]. Phenylalanine (Phe), tyrosine (Tyr), and/or tryptophan (Trp) fluorophores are present in most proteins, which allows the use of IFS to characterize protein structural change in solution [1]. Protein intrinsic fluorescence emission is dominated by Trp when present, with Tyr and Phe contributing much less due to their lower quantum yields and Förster resonance energy transfer (FRET) processes [1,5]. Furthermore, Trp emission is highly sensitive to environmental changes [6], making it a particularly useful tool for monitoring protein structural changes [7–10]. Tyr emission is insensitive to changes in solvent polarity, but is affected by changes in FRET, which potentially provides information about conformational changes and size [1,7,11]. Tyr emits at ~305 nm, which means that the emission is highly affected by Rayleigh light scatter contamination at the blue edge and overlap with Trp fluorescence at the red edge of its emission. This makes resolving the Tyr signal very challenging and spectral deconvolution methods are usually required [12–14]. For IgG which has twice as many Tyr fluorophores (~50/60) as Trp (~20/30) and with many being positioned within the Förster radius of each other, it means that homo- and hetero-FRET is extensive, resulting in Trp dominant emission. In many cases the application of IFS for analytical protein studies use simple 2D spectra or single excitation/emission wavelength pairs, which take no advantage of all the spectral information available in the full excitation and emission space. Therefore, the use of multidimensional fluorescence

* Corresponding author. Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland – Galway, Galway, H91 CF50, Ireland.
E-mail address: alan.ryder@nuigalway.ie (A.G. Ryder).

https://doi.org/10.1016/j.chemolab.2019.103871
Received 24 July 2019; Received in revised form 9 October 2019; Accepted 12 October 2019
Available online 15 October 2019

spectroscopy (MDF) offers a potentially better approach to retrieving more useful information from multi-fluorophore proteins.

The commonest MDF measurement techniques for intrinsic fluorescence are Excitation-Emission Matrix (EEM) [15] and Total Synchronous Fluorescence Spectroscopy (TSFS) [16]. These provide a 3D spectral signature of the multiple fluorophores present in a protein or a complex mixture [17–19]. TSFS is potentially more useful than EEM for blue edge emission analysis because it is less affected by the first order Rayleigh scattering (RS) interference in the raw data and thus require much less or no data pre-processing for scattering removal compared to EEM. Thus, TSFS measurements may preserve more spectral information from the blue edge of protein emission and in particular the Tyr component. Furthermore, the presence of light scatter has a very large and adverse effect on anisotropy measurements which makes its elimination a priority.

Anisotropy resolved multidimensional emission spectroscopy (ARMES) is a combination of anisotropy, MDF measurements, and chemometrics [9]. ARMES uses anisotropy (r) as an additional layer of photophysical information in combination with the intensity (I), excitation wavelength (λ_{ex}), and emission wavelength (λ_{em}). ARMES provides extra information that can help better characterize changes in intrinsic protein fluorescence by better resolving the emission of individual fluorophores. For multi-fluorophore proteins, anisotropy varies across the fluorescence emission space (*aniso*-MDF map) which is due to the presence of multiple fluorophores and the interactions (*e.g.* FRET) among these fluorophores [19,20]. Thus, *aniso*-MDF maps are very sensitive to protein structure changes, which make these maps a good diagnostic tool for evaluating protein changes (unfolding, refolding, and aggregation) [9]. Multi-way decomposition methods, such as multivariate curve resolution (MCR) [21,22] and parallel factor analysis (PARAFAC) [23–25] are required to resolve fluorophore contributions from MDF data [26,27] which is a critical part of ARMES.

ARMES using TSFS measurements was previously used for the resolution and identification of individual fluorophores in relatively simple proteins with one or two Trp fluorophores [9,28]. This data, however, was polarizer dependent with no excitation below ~300 nm. The use of wire grid polarizers permitted excitation below 300 nm and thus collection of a more complete excitation-emission space including Tyr excitation/emission [20]. A previous study of ours [29], showed that for IgG, polarized EEM (pEEM) measurements were adversely affected by residual Rayleigh scatter, and it affected PARAFAC resolution of Tyr and Trp at the emission blue edge, $\lambda_{em} < 320$ nm, particularly for the parallel polarized emission. Solving this issue, requires the use of more costly spectrometers with improved stray light rejection (double emission monochromators) and/or the use of reduced emission slits to increase spectral resolution (but at the cost of signal-to-noise ratio), which is not always practical or feasible with standard benchtop instruments. TSFS therefore becomes the most practical option to overcome scattering issues and to enable better spectral resolution at the emission blue edge. As TSFS lacks trilinearity required for PARAFAC analysis a simple transformation into an EEM like layout (t-EEM) was required [30]. However, this t-EEM transformation generates regions with no experimentally acquired spectral data, that is the absent spectral data region (ASDR), and this could have a negative impact on PARAFAC modelling.

Here we explore the use of polarized TSFS (pTSFS) measurements to try to better characterize the fluorescence emission of native state rabbitt Immunoglobulin (rIgG) over a small temperature range compared to the previous pEEM/PARAFAC analysis where the Rayleigh scatter issue was a significant problem [29]. A critical factor was to deal with the ASDR, and here we evaluated the impact of using three different methods: missing data, insertion of zeros [31], and interpolation [32] on the PARAFAC model output (component profiles and scores). Of particular importance is the recovery of Tyr fluorescence from the blue edge of the emission which was badly affected by Rayleigh scatter contamination in EEM measurements.

2. Materials and methods

2.1. Materials

A 0.01 M Phosphate 0.150 M saline buffer (PBS) at pH 6.5 ± 0.1 was prepared with sodium phosphate monobasic, sodium phosphate dibasic hepta-hydrate and sodium chloride (purchased from Sigma-Aldrich and used without further purification), which were dissolved in HPLC grade high purity water (HPW) purchased from Honeywell. PBS buffer was used to reconstitute IgG from rabbit serum (≥95% essentially salt-free, lyophilized powder) that was purchased from Sigma-Aldrich and used without further purification. Aseptic sample preparation techniques were used to minimize contamination. Three rabbit IgG (rIgG) solutions were prepared in PBS buffer (1.1 ± 0.2 mg mL^{-1}); from three vials that were acquired at the same time and with the same lot number (SLBM2617V). Polyethersulfone (PES) Captiva Premium Syringe 0.2 μm filters (Agilent) were used to filter the solutions. The three rIgG solutions had a final concentration of 1.3 mg mL^{-1}, 0.9 mg mL^{-1}, and 1.2 mg mL^{-1}, for R1, R2, and R3, respectively. Lobind Eppendorf tubes (1.5 mL) were used to separate aliquots (enough for one measurement) of each rIgG solution and stored at −70 °C before use (over 11 weeks). Freshly prepared samples were immediately stored at −70 °C, without making any measurements. Prior to making spectroscopic measurements, each aliquot was slowly defrosted overnight at 4–8 °C.

2.2. Instrumentation and data collection

UV–visible and fluorescence measurements were made using conditions as previously described [29] and details are provided in the supplemental information. Slit widths (10 nm), scan rate (1200 nm mn^{-1}) and photomultiplier tube (PMT) detector voltage of 650 V were used [29] for all measurements. Polarized TSFS (pTSFS) were collected using an excitation range of $\lambda_{ex} = 240$–320 nm (same as EEM), and $\Delta\lambda$ interval of 20–210 nm at 2 nm step increments in each case. pTSFS measurements were extended until the red edge of the Trp emission ($\lambda_{em} > 450$ nm) at each λ_{ex}, in order to ensure comparability with previous EEM based studies (Fig. 1). TSFS measurements took approximately 10 min per polarization setting, and thus 40 min in total per temperature. To characterize native state emission, rIgG solutions, in triplicate, were measured at nine different temperatures (15°, 17°, 20°, 23°, 25°, 27°, 30°, 33° and 35 °C), and solutions were thermally equilibrated for five minutes at each temperature. Spectra were uncorrected for instrument response, which was also dependent on polarizer transmittance.

2.3. Data analysis and chemometric methods

PLS_Toolbox ver. 8.2.1 (Eigenvector Research Inc.) and FluorS (in-house written program) were implemented in MATLAB ver. 9.1.0 (The Mathworks Inc.) for the data analysis. TSFS behaves as a non-bilinear matrix, meaning that the spectral shape and emission intensity of each fluorophore varies as $\Delta\lambda$ changes (*vide infra*). Thus, each fluorophore in TSFS does not have a unique synchronous fluorescence spectrum and this prevents direct PARAFAC analysis, unless data restructuring is performed to generate trilinearity. Three-way TSFS data could be directly decomposed by bilinear methods like MCR-ALS [17,19,21,22,31,33,34], resolving the spectral profile of each fluorophore if proper data augmentation is performed and suitable designed sample sets are used [35,36]. However, in the case of IgG intrinsic emission, the emission of individual fluorophores have very different intensities, degrees of spectral overlap, and are non-radiatively connected, and in general experimental design cannot be implemented which could break these dependencies [31,33]. Thus, the trilinearity required for PARAFAC might help when analyzing the highly overlapped emission of IgG. In addition, PARAFAC of this TSFS data can be directly compared with PARAFAC modelling of similar EEM data [29], to unambiguously determine which measurement method is superior in the context of protein analysis.

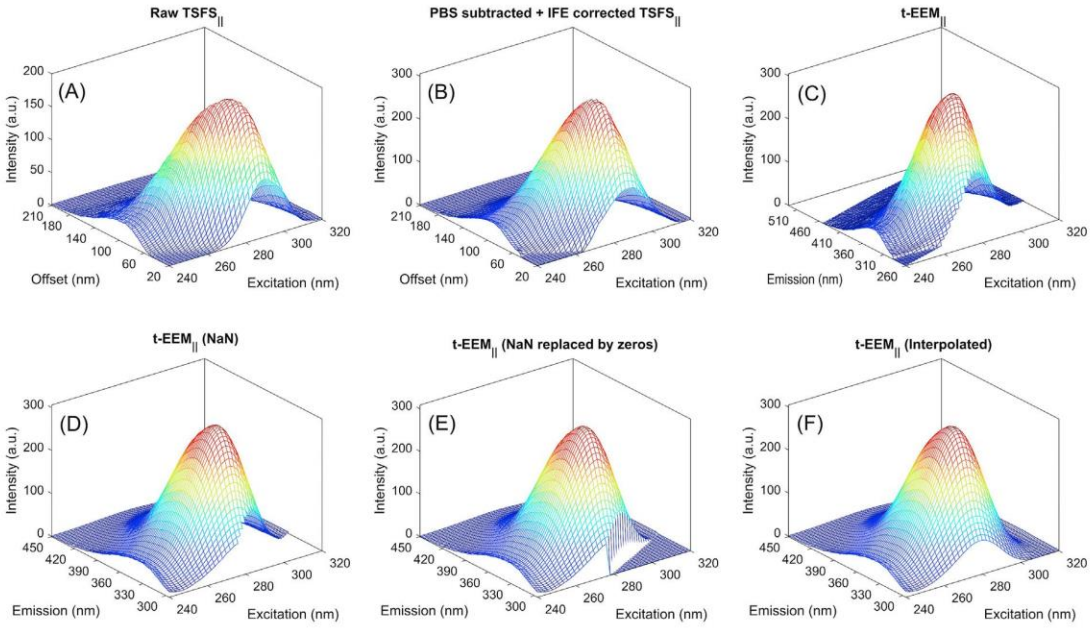

Fig. 1. (A) Raw $TSFS_{\|}$ spectra of rIgG native state measured at 20 °C. (B) $TSFS_{\|}$ PBS subtracted, and IFE corrected. (C) $TSFS_{\|}$ data was converted into t-$EEM_{\|}$. (D) t-$EEM_{\|}$ spectra were cut at λ_{em} 296−450 nm to remove most of the noisy area produced by IFE correction. (E) The areas outside the first Rayleigh scattering were replaced by zeros. (F) The t-$EEM_{\|}$ spectra were interpolated and smoothed.

$TSFS_{HV}$ and $TSFS_{HH}$ measurements were used to calculate the G-factor ($G = I_{HV}/I_{HH}$, Figs. S1/S3, SI) and this was used to correct the $TSFS_{VH}$ spectra, giving the corrected perpendicular $TSFS_{\perp}$ spectra (Fig. S2, SI) [29]. The $TSFS_{VV}$ spectra are the parallel polarized, $TSFS_{\|}$, spectra. The total unpolarized $TSFS_T$ spectra were calculated from the pTSFS spectra as follows: $TSFS_T = TSFS_{\|} + 2 \times TSFS_{\perp}$, and were used to compare the outputs of modelling polarized TSFS measurements compared to the simpler, and more common unpolarized TSFS measurements [1,29,37,38]. The polarized datasets, $TSFS_{\|}$ and $TSFS_{\perp}$, were then subjected to Raman scattering minimization and inner filter effect (IFE) correction [38].

The use of wavelength offsets, $\Delta\lambda \geq 20$ nm, minimizes the collection of Rayleigh scattered light (Fig. 1) and thus, Raman bands were the major source of scattered light [1]. However, the relatively wide slit widths (10 nm) coupled with the very short Stokes shifted emission for excitation wavelengths of <300 nm means that some light scatter will be present. If the sample solution contains particles, then this will increase further due to Mie and Tyndall scattering. Raman scattering was easily removed with a blank subtraction (PBS buffer spectrum) from the TSFS spectrum (Fig. 1B). IFE correction was necessary due to the high optical density (Abs = 0.31 ± 0.05, at 280 nm, 2 mm pathlength) of the protein solutions (Fig. S4, SI) [39,40]. The absorbance-based approach [1] using the limit of reporting (LOR) [39] as a guide was implemented as previously described [29]. The polarized MDF measurements yield much weaker fluorescence intensity than unpolarized MDF measurements [9], making the IFE correction a critical step for accurate PARAFAC resolution and anisotropy calculations.

The pTSFS emission spectral range of $\Delta\lambda = 20$–210 nm yielded matrices of 27 samples × 96 $\Delta\lambda$ × 41 λ_{ex}, which were then rearranged into t-pEEM (Fig. 1C), generating matrices of 27 samples × 136 λ_{em} × 41 λ_{ex} [17,31]. The weak emission bellow $\lambda_{em} \leq 292$ nm was mainly due to noise, which was amplified by IFE correction (Fig. 1C). To both eliminate IFE correction artifacts and to make the pTSFS data equivalent to the pEEM data, the t-pEEM data was reduced to $\lambda_{em} = 296$–450 nm, generating matrices of 27 samples × 78 λ_{em} × 41 λ_{ex} (Fig. 1D). The removed short emission wavelength was mostly related to weak Tyr emission as well as the blue edge of Trp emission [1,28]. The weak Tyr emission in the t-pEEM datasets was less affected than the same Tyr emission in the pEEM datasets [29], because of the much reduced scattered light contamination.

The t-pEEM data (Fig. 1C) as collected had a large number of coordinates (λ_{ex}, λ_{em}) with no intensity data (~30%) [31], the absent spectral data region (ASDR). This was reduced (Fig. 1D) to a much smaller region (~8%) when the spectra were edited down to the $\lambda_{em} = 296$–450 nm range used for data analysis. The use of missing data to deal with the ASDR makes PARAFAC free to estimate a continuous profile for the spectral components. Unfortunately, these solutions can easily be distorted, leading to discontinuities and sharp peaks [41], due to the presence of IFE correction induced artifacts, residual light scatter and noise, in the short wavelength emission regions which are close to these missing values. These factors are exacerbated here by the poor polarizer transmittance and instrument performance in the 250–300 nm region. The other solutions for facilitating PARAFAC modelling were either to replace the ASDR with zeros (Fig. 1E) [42] or use interpolated values (Fig. 1F). However, while using zeros may facilitate PARAFAC convergence [42], here the ASDR, contains the short Stokes shifted emission (both Tyr and Trp) region. This means that imposing a zero value, artificially distorts the true emission data being used for modelling, and this can lead to incorrect spectral profiles being recovered from the PARAFAC models [32]. The use of interpolation has been previously proven to obtain chemically meaningful solutions from EEM data [32], however, it has never been used to reconstruct t-pEEM data from protein emission. As with the other methods, areas with low SNR and residual scatter close to the ASDR will adversely affect PARAFAC. One issue with interpolation of protein emission is validating that the reconstructed emission is the true emission and that artifacts are not introduced. The use of interpolation for addressing the ASDR issue has been explored [17,30,31] but with a very different sample system.

Once the IFE and ASDR corrections had been implemented, the standard anisotropy [1] formula (Equation S1, SI) was used to calculate the anisotropy (r) at each emission wavelength, and then used to build *aniso*-MDF maps ($\lambda_{ex} \times \lambda_{em} \times r$) [38]. Finally, for PARAFAC analysis, the

t-pEEM data were then normalized (by peak maximum), to remove variances introduced by small concentration differences and in day-to-day measurements, to more clearly show the real spectral changes being induced by structural change [23].

For PARAFAC analysis, the number of components was selected based on several criteria: the CORe CONsistency DIAgnostic test (CONCORDIA) [23], on how much of variance was explained by the model (Table S1, SI), and visual inspection of the recovered spectral profiles and residuals. Validation of spectral deconvolution results were performed using split-half analysis. PARAFAC analysis was applied with non-negative constraints for all modes (sample, λ_{em}, λ_{ex}) using the best-fitting model as initialization method (from various test models fitted with a small number of iterations) [25,32]. First, the t-pEEM datasets containing missing data, filled with zeros and interpolated were analyzed by PARAFAC to see if the different emitters could be resolved and if there were differences in the recovered loadings (spectra) for the different emission polarization states. Once sensible models and components had been generated, anisotropy spectra for the recovered components were calculated. PARAFAC resolution of t-pEEM were then compared with that previously obtained from the same samples measured with pEEM [29].

3. Results and discussion

3.1. 2D spectral analysis of MDF methods

Two-dimensional spectra were used first to evaluate the spectral information obtained in the t-pEEM and EEM measurements, considering the differences between Trp and Tyr emission, Tyr-to-Trp FRET, and the amount of spectral variations produced over a 15–35 °C temperature range [29]. The difference in the recovered Tyr emission peak maxima (Fig. 2, gray spectra): t-EEM$_T$ (312 nm) and EEM$_T$ (316 nm) was caused by the EEM Rayleigh scatter correction procedure. This removed part of the weak Tyr signal which overlapped the Rayleigh scatter, leaving only the Tyr emission that overlapped with Trp emission. The shape of the 2D t-pEEM data was very similar to that extracted from pEEM measurements except that the Tyr emission was stronger in the t-pEEM data (Fig. 2). The difference spectra showed that Tyr fluorescence [12] corresponded to <5% of the EEM$_T$ (Fig. 2C/D) but ~13% of the t-EEM$_T$ (Fig. 2A/B) spectra. The better quality of the t-pEEM data is also revealed by the smaller anisotropy values obtained at short emission wavelengths (λ_{em} < 320 nm). The recovered anisotropy values (Fig. 2B) now agree better with what one might expect from a system where there is extensive FRET from Tyr to Trp, and where at least some of the Trp residues are directly excited with relatively high anisotropies [1]. These we expect to be the Trp residues located in the hydrophobic environments and thus at the blue side of the band.

Overall, conventional 2D spectral analysis of both t-pEEM/pEEM indicated that there were very little spectral shape changes which suggest that there was no substantial structural change, and thus little change in Tyr-to-Trp FRET as expected. This was to be expected as IgG are relatively stable proteins and the detailed pEEM study has verified this [29]. Nevertheless, there were some small, yet significant differences, with a relative standard deviation (RSD) of <5.4 (t-pEEM) and <5.0% (pEEM), at the blue edge (but not at the red edge) when the protein was excited at 270 and 296 nm. However, IFE correction, created an artifact that can be seen at λ_{em} < 310 nm in the t-pEEM and pEEM difference spectra (Fig. 2, and Fig. S6, SI), which may have an impact on PARAFAC modelling. This means that there is still some ambiguity with any measured data in this region.

The use of single excitation wavelengths, however, does not give the full picture and one needs to look at the full emission space, using chemometric methods to try and resolve fluorophore contributions. The first stage is to investigate how well the pTSFS measurements have been at removing any scattered light contamination compared to the conventional EEM based approach. The most sensitive way of doing this prior to PARAFAC analysis was to first look at the anisotropy maps with particular focus on the blue edge of the emission space.

3.2. Aniso-MDF maps of rIgG

The anisotropy variation across the rIgG emission space represented in the *aniso*-MDF maps was caused by a variety of factors such as: type and number of fluorophores present in the protein, fluorophore location in the protein, local fluorophore motion, changes in FRET, and variations in the physicochemical environment [9]. The presence of contamination from Rayleigh scattered light will also be evident as regions of abnormally high anisotropy at the blue edge of the emission. This is clearly shown in Fig. 3 where the *aniso*-EEM map is much more heavily distorted as evidenced by the much higher anisotropy at the emission blue edge which also means that the weak Tyr signal is masked (Fig. 4H). The

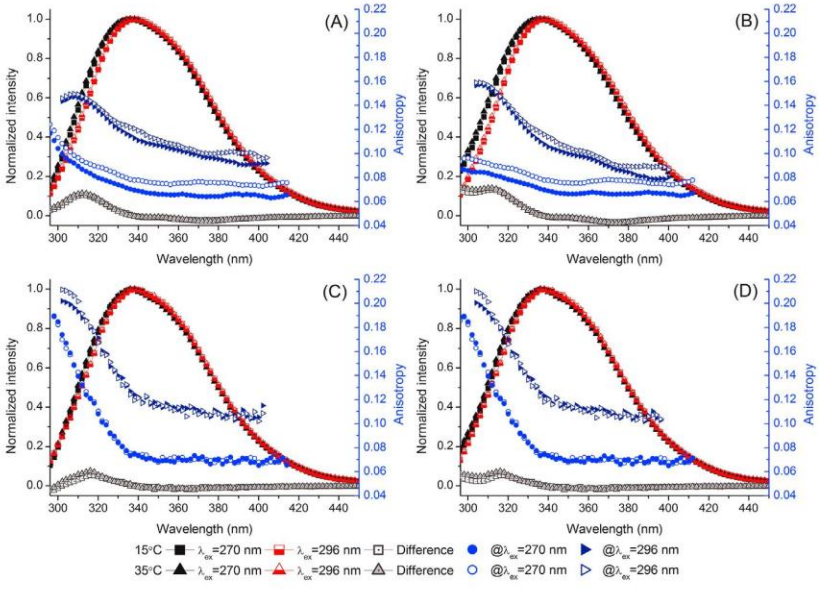

Fig. 2. Normalized (Top) t-EEM$_T$ and (Bottom) EEM$_T$ emission spectra of rIgG excited at λ_{ex} 270 and 296 nm and the difference spectra (=λ_{ex} 270 − λ_{ex} 296) of the rIgG native state at two different temperatures (15° and 35 °C), overlaid with the emission anisotropy at λ_{ex} 270 and 296 nm (blue circles and right pointing triangles). (A, C) Not IFE corrected. (B, D) IFE corrected using UV–Vis spectra collected at 20 °C before MDF measurements. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

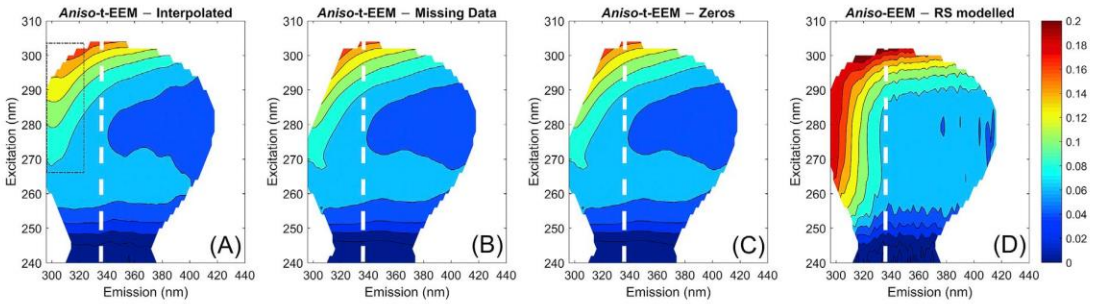

Fig. 3. (A/B/C) *Aniso*-t-EEM and (D) *aniso*-EEM maps (corrected for Rayleigh scatter) of rIgG at 20 °C. The white dashed lines mark the λ_{em} @336 nm excitation lines. The black box shows the Tyr region reconstructed by interpolation. The colour bar on the right represents the anisotropy scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

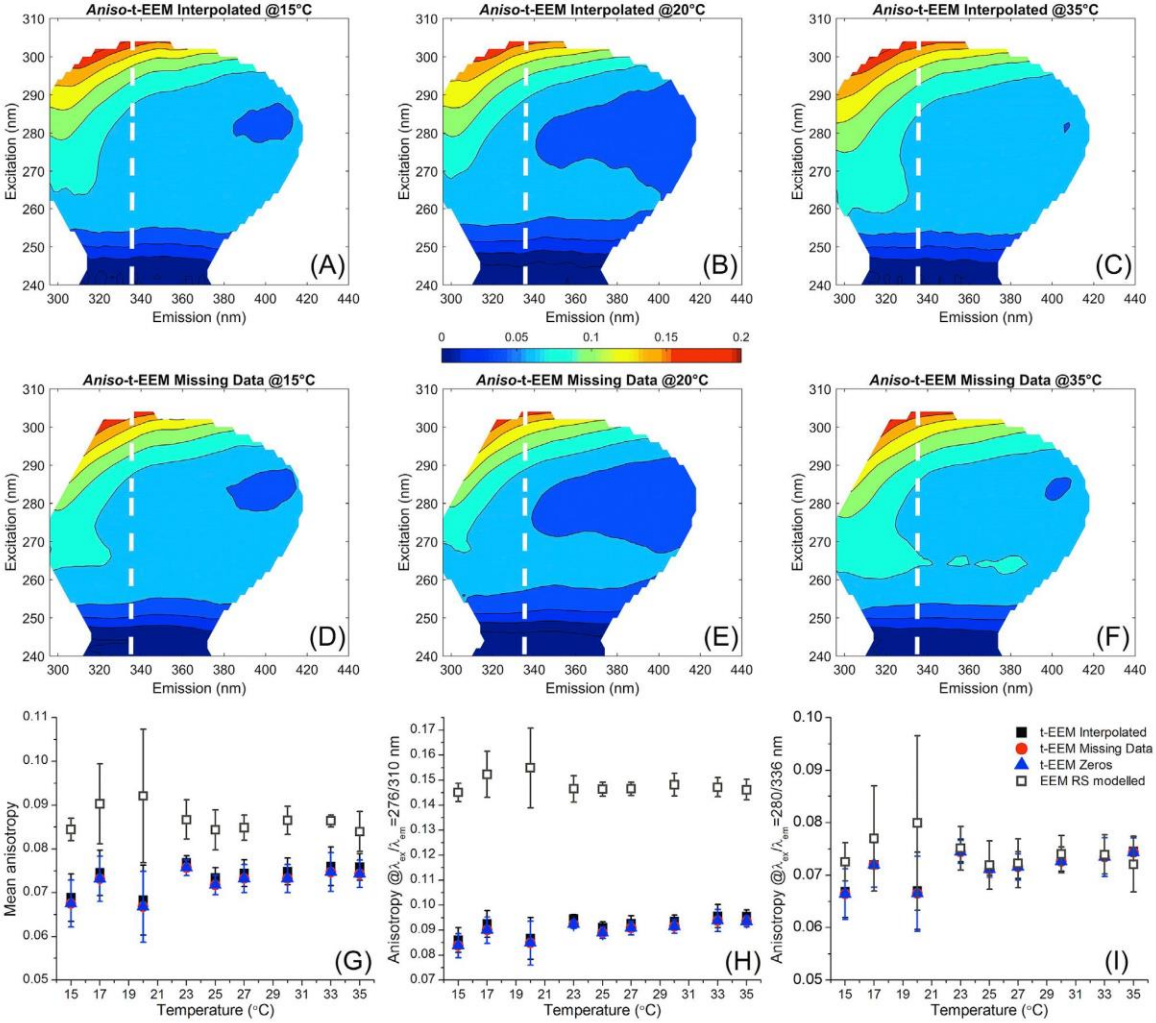

Fig. 4. *Aniso*-t-EEM maps for rIgG collected at (A/D) 15 °C, (B/E) 20 °C, and (C/F) 35 °C. *Aniso*-t-EEM maps for (top) interpolated datasets and (bottom) with missing data (<8% of data points). The white dashed lines mark the λ_{em} @336 nm. The colour bar on the middle represents the anisotropy scale. (G) Mean anisotropy ± Std Dev, (H) Tyr anisotropy at $\lambda_{ex}/\lambda_{em}$ = 276/310 nm ± Std Dev, and (I) Trp anisotropy at $\lambda_{ex}/\lambda_{em}$ = 280/336 nm ± Std Dev for the triplicate t-EEM and EEM measurements of rIgG native state. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

differences between the *aniso*-t-EEM maps that were generated from data which was interpolated (Fig. 3A), used the missing data approach, or filled with zeros (Fig. 3B/C), were mainly due to reconstruction of part of the Tyr emission region in the interpolated method (Fig. 3A, black box).

When we examine the change in the *aniso*-t-EEM maps over the temperature range (Fig. 4A–F), there was no significant differences (ANOVA, $p > 0.05$) between the *aniso*-EEM and *aniso*-t-EEM maps in the Trp emission region, $\lambda_{em} > 336$ nm, (Fig. 4I). Similar to our previous study, overall there were only small changes in anisotropy at these longer wavelengths [29]. However, when we look at the changes in the mean anisotropy (Fig. 4G) we see large differences (<38% difference) between the *aniso*-EEM and *aniso*-t-EEM maps which is due to residual scatter in the pEEM data. This is clear from Fig. 4H–I where the differences are much greater in the Tyr emission region.

To better assess the overall degree of change in *aniso*-MDF maps a similarity index (SimI) analysis [43] was undertaken [29], (Fig. 5). Maps with SimI values > 0.9 could be considered to have no significant differences from the reference spectrum, and from the 15 °C measurement it was clear that the inter-replicate variation was the most significant difference observed. This was due to a combination of concentration differences (up to ~30%) coupled with varying Rayleigh scatter (amplified by the IFE correction process). This residual scattered light is present in the $TSFS_{\|}$ measurements (*vide infra*). Another contributor to this variation was the fact that the experiments used multiple vials of a polyclonal rIgG and one freeze-thaw cycle [44]. SimI analysis over the full emission space clearly showed that *aniso*-t-EEM maps (RSD = 7.7 and 10.2% of the interpolated and missing data/zeros datasets, respectively) were more reproducible than *aniso*-EEM maps (RSD = 21.50%). The improved reproducibility of the interpolated *aniso*-t-EEM data compared to the missing data or added zeros data was due to the reconstructed ASDR. For the two specific emission regions (Fig. 5B/C) this variation was noticeably reduced, for the *aniso*-EEM (RSD = 7.6 and 8.6% for the Tyr and Trp regions, respectively [29]) and *aniso*-t-EEM maps (RSD~5.9 and 6.6% for the Tyr and Trp regions, respectively). This suggests that more of the temperature induced emission changes were concerned with Trp emission.

Also, it is important to note that the same temperature effect (small dip in SimI at 20 °C, followed by an increase between 23 °C and 35 °C, Figs. 4/5) was observed for both pEEM [29] and t-pEEM data. This indicated that this effect was real because the pTSFS and pEEM datasets were collected three weeks apart (three days of measurement each), with the samples having been stored at −70 °C for 11 and 14 weeks respectively. Each set of measurements used the same stock solution prepared from different vials of rIgG but with the same lot number. Changes in anisotropy were greater at lower temperatures (≤20 °C), with a significant difference between replicate measurements (Fig. 5). The same behavior was observed for both MDF measurements, increasing our suspicions that this was caused by a change in hydrodynamic behavior of rIgG at lower temperatures, differences in local motion, and/or flexibility of the protein structure, but we still need more evidence to support this [45,46].

The spectral and *aniso*-map assessments above, clearly showed that changes in emission arising from structural changes/fluctuations were very small over this temperature range. It was also clear that the t-pEEM measurements were much superior to pEEM because of the reduced scattered light contamination at the short emission wavelengths (<320 nm).

3.3. PARAFAC modelling of t-EEM

Meaningful PARAFAC deconvolution of intrinsic fluorophore emission from proteins is inherently challenging because of several factors: the non-linearity associated with Tyr-to-Trp FRET, emission and Rayleigh scatter overlap, particularly at short wavelengths, large differences in relative emission contributions between fluorophores, and the large numbers of fluorophores involved which only show small spectral variation. For EEM measurements, the scatter issue was most critical and limited the effective application of PARAFAC to resolve the Tyr emission contribution [29]. Here, the better quality of the t-pEEM data at short emission wavelengths has resulted in a stronger Tyr signal (Fig. 4H), which should facilitate the recovery of more than one component by PARAFAC. However, the combination of TSFS and PARAFAC analysis is not widely used [17,30,31] and requires first that the data is converted into an EEM like layout (t-pEEM). The issue with this is to determine what is the impact of the different methods for dealing with the ASDR (i.e. the blue edge of the emission) on PARAFAC modelling.

Results from PARAFAC analysis of t-pEEM data were compared (Table 1, Fig. 6) taking care that all the modelling was implemented under equivalent conditions. Table 1 shows that in all cases only two components were recovered (as suggested by CONCORDIA *and* split-half analysis) and that these were very similar. It was clear that all models resolved the same two fluorophores species, one associated with Trp emission (PFC1, >92–97%) and a second weaker one (PFC2, >3–7%) which is mainly associated with Tyr emission. Analysis of the scores (Fig. 7) for the un- and normalized data clearly shows that emission was quenched and that this was mainly involving the Trp emission. The increased thermally induced quenching of Trp, changed the ratio of the Tyr to Trp contributions facilitating the resolution of the two components. This was not possible with the pEEM data because of the light scatter contamination issue therefore justifying the use of pTSFS measurements.

However, there were significant differences in the profiles of the recovered fluorophore species, in particular with the weaker mostly Tyr emitting component (Fig. 6, and Fig. S7, SI), and this is an issue for the photophysical interpretation of the PARAFAC analysis. The differences were related to the ASDR pre-processing methods and their effect on the

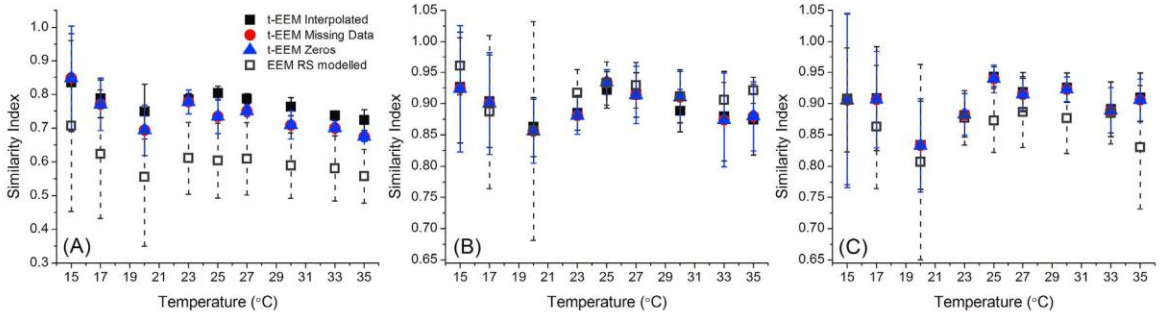

Fig. 5. Plots of SimI values calculated for the various *aniso*-MDF maps over the 15–35 °C temperature range. In each case, the reference spectrum was the MDF collected at 15 °C from the first replicate sample (R1). (A) the full emission space; (B) the Tyr emission region ($\lambda_{ex}/\lambda_{em}$ = 270–290/300−320 nm); and (C) the Trp emission region ($\lambda_{ex}/\lambda_{em}$ = 280–300/320−360 nm). Error bars were generated from the standard deviation from triplicate measurements of the independent samples.

Table 1
Comparison of the model parameters and components obtained for the normalized polarized t-EEM$_{\parallel}$, t-EEM$_{\perp}$, and unpolarized t-EEM$_T$ PARAFAC models (50 repetitions) of the rIgG native state. The spectral data were corrected using interpolation, missing data (7.9%), or filling with zeros. Samples were measured over the 15–35 °C temperature range (9 temperatures × triplicate measurements).

	Interpolated			Missing Data			Zeros		
	t-EEM$_{\parallel}$	t-EEM$_{\perp}$	t-EEM$_T$	t-EEM$_{\parallel}$	t-EEM$_{\perp}$	t-EEM$_T$	t-EEM$_{\parallel}$	t-EEM$_{\perp}$	t-EEM$_T$
C1 $\lambda_{ex}/\lambda_{em}$ (nm)	280/338	280/342	280/340	280/338	280/340	280/340	280/336	280/340	280/338
C1 Fit model (%)	96.99	92.60	95.03	96.55	95.17	96.46	96.35	96.07	96.23
C2 $\lambda_{ex}/\lambda_{em}$ (nm)	278/298 (278/310[a])	278/314	278/312	320/300 (278/300[a])	276/312	278/300 (278/310[a])	276/306	276/306	276/306
C2 Fit model (%)	3.01	7.40	4.97	3.45	4.83	3.54	3.65	3.93	3.77
Variance explained ±Std. Dev. (%)	99.99 ($\pm4\times10^{-9}$)	99.99 ($\pm3\times10^{-8}$)	99.99 ($\pm1\times10^{-8}$)	99.99 ($\pm2\times10^{-5}$)	99.99 ($\pm3\times10^{-9}$)	99.99 ($\pm2\times10^{-9}$)	99.58 ($\pm1\times10^{-5}$)	99.70[b] ($\pm9\times10^{-8}$)	99.66 ($\pm8\times10^{-8}$)
CONCORDIA ± Std. Dev (%)	29.25 (±0.12)	31.09 (±0.52)	23.87 (±0.38)	99.80 (±0.04)	63.27 (±0.004)	65.98 (±0.003)	85.80 (±0.46)	76.53 (±0.002)	78.59 (±0.004)
Split-half analysis (%)	99.97	99.96	99.95	99.80	99.03	99.61	56.84	99.72	85.67
Iterations	7 (±16)	22 (±32)	9 (±19)	3424 (±1871)	1242 (±1)	2119 (±2)	4 (±3)	4 (±0)	4 (±0)
Time (s)	7 (±1)	10 (±1)	8 (±1)	47 (±21)	22 (±2)	30 (±2)	5 (±1)	7 (±1)	6 (±1)

[a] PARAFAC components without artifacts.
[b] CONCORDIA values of local minima models were removed. One local minima model was found.

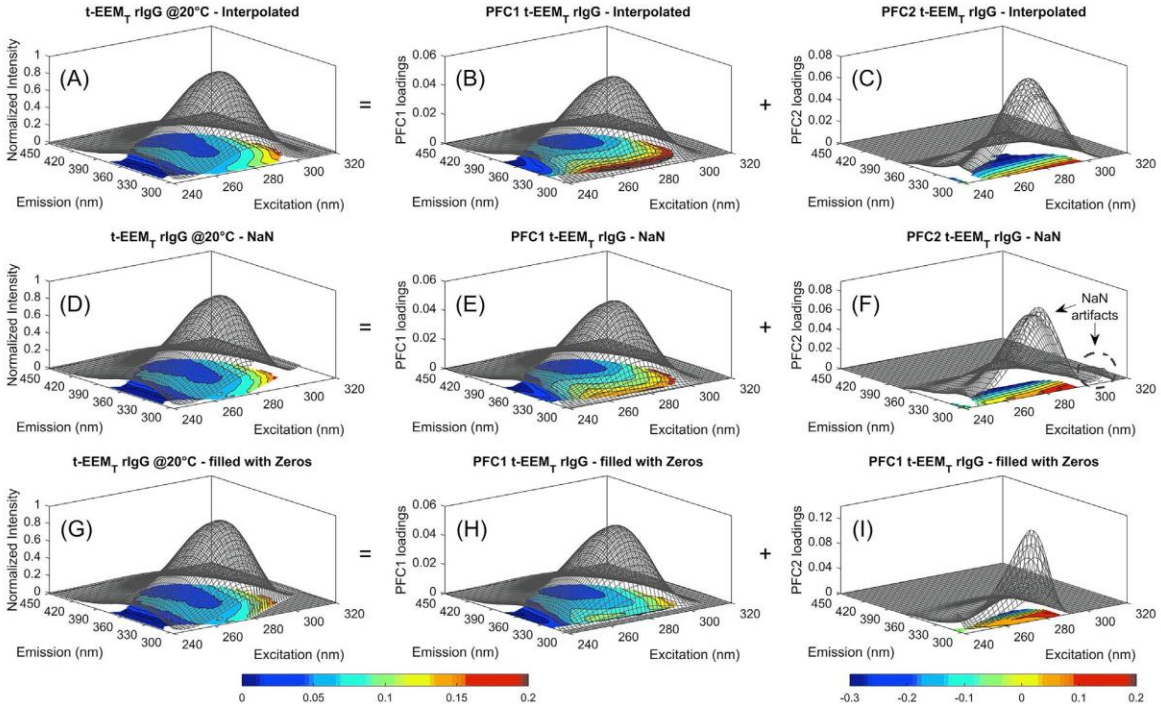

Fig. 6. Comparison of PARAFAC modelling of the t-EEM$_T$ rIgG native structure data (A) interpolated, (D) missing data (NaN), and (G) filled with zeros. PFC1 (Trp) was virtually identical for the datasets with (B) interpolation, (E) missing data and (H) filled with zeros. Even if PFC2 resolved the Tyr signal for the (D) interpolated, (F) missing data and (I) filled with zeros, there were slight differences caused by the presence of missing data and zeros. t-EEM$_T$ landscapes overlapped with *aniso*-t-EEM maps for rIgG measured at 20 °C. The colour bars on the bottom represent the anisotropy scale. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

short Stokes shifted emission region. PFC1 (Trp) was virtually identical in terms of the recovered spectral profiles, except for the models filled with zeros. The insertion of zeros method restricted PARAFAC to an abnormal solution which led to an underestimation of the blue edge emission of the Trp fluorophores [42]. This method facilitated PARAFAC convergence (lowest number of iterations, Table 1), but it could cause premature deconvolution, increasing the variability of the spectral profiles of the calculated models. Thus, insertion of zeros slightly underestimated both Trp and Tyr relative contributions (Fig. 6H/I) compared to the other methods.

When the missing data approach to ASDR was implemented, PARAFAC was free to better estimate the continuous shape of Trp and Tyr

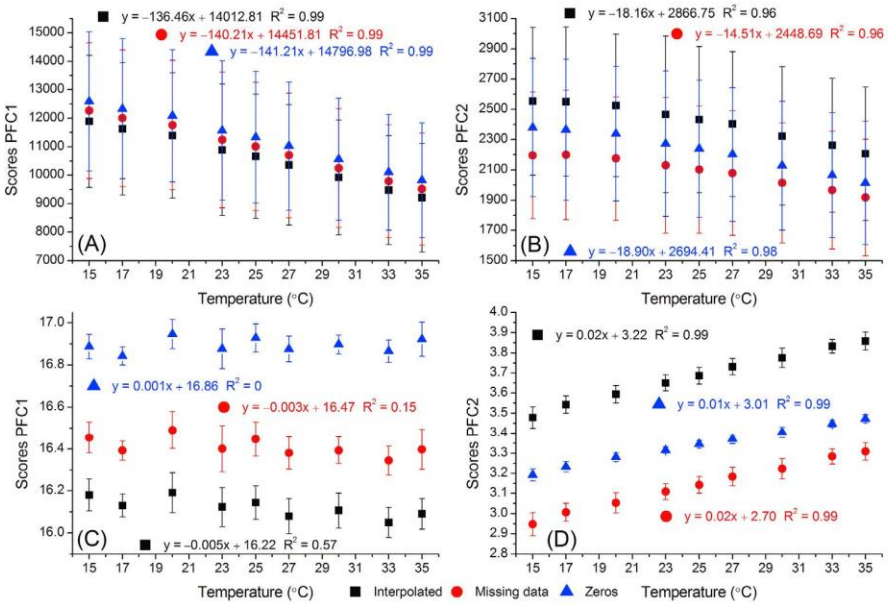

Fig. 7. Scores plots of two-component PARAFAC models for the: (top) non-normalized, and (bottom) normalized unpolarized rIgG t-EEM$_T$ datasets with the different ASDR treatment methods.

emission. This was because it did not restrict or modify the data, and thus better extracted spectral profiles which might be a truer representation of the actual emission. The missing data method did, however, create some artifacts at the emission blue edge (Fig. 6F) which was amplified by the IFE correction as shown in the excitation slab corresponding to PFC2 emission and this was particularly severe for the parallel polarized light data (Figs. S7B, SI). The use of interpolation significantly decreased the required computational time and led to a resolution that appears to be spectrally acceptable. However, CONCORDIA values were reduced for the interpolated data compared to the other methods, which was caused by extension of emission into the ASDR and thus increased emission overlap (Fig. 6, and Fig. S9, SI). One might expect that this suggests that three components should be recoverable by PARAFAC (Tyr, Trp from hydrophobic locations, and Trp externally located on the protein). However, the three component PARAFAC models (Table S1, SI) did not show any improvement in the model performance quality parameters.

Differences in quality and model parameters between the different polarizations (Table 1, and Fig. S7, SI) for the missing and interpolated t-pEEM methods were related to sample spectral characteristics, and specifically the intrinsic anisotropy of the emission from the large protein molecule, and the degree of FRET. This is clear to see when one compares the TSFS$_T$ results with the pTSFS models in the table above. The parallel polarization measurements are more directly sensitive to Tyr-to-Trp FRET (i.e. lower PFC2 scores) and to the presence of residual scatter. This is due to a combination of factors including the short fluorescence lifetime of Tyr (<4 ns) coupled with the long, 26 ns, rotational correlation time of IgG [47]. In the t-EEM$_{\parallel}$ models, it was clear that there was an element of the spectral profile that could be assigned to Rayleigh light scatter which was amplified by the IFE correction (PFC2 blue shifted). In the t-EEM$_{\perp}$ models there was a clear shoulder at ~350 nm (Fig. 6, and Fig. S7, SI) due to Trp emission which resulted in a higher score (~5–7%) for this component (Table 1, Fig. S8, SI). PFC2 scores for the t-EEM$_{\perp}$ were higher for the interpolated (~7%) than the missing data (~5%), suggesting that the interpolation method improved the resolution of Tyr emission. In contrast, there were no significant differences between the polarized and unpolarized t-EEM PARAFAC using the data filled with zeros, which is surprising, as there are real differences between the polarized and unpolarized spectra of rIgG. This therefore rules out the use of this ASDR correction procedure for intrinsic protein fluorescence analysis.

For better analysis of protein structural change, it would be useful to be able to fully resolve the Trp and Tyr emission [9,12,14], to provide both spectral profiles and relative contribution values. Analysis of the PARAFAC components, PFC1 (Trp) and PFC2 (Tyr) scores from the un-normalized t-EEM$_T$ data from each method showed a linear decrease with increasing temperature due to quenching (Fig. 7 A/B). The trends were the same for all three data handling methods with no significant differences (ANOVA, $p > 0.05$) in PFC1/2 scores trends, indicating that any of the PARAFAC scores could be used to monitor structural/concentration changes (r Pearson>0.99). The error bars in Fig. 7 A/B were due to differences in protein concentration between the replicate samples (~31%), which is why the normalized data is better for investigating structural changes.

Normalization removed variations caused by small protein concentration differences and instrument measurement error, and this was reflected in the smaller error bars (~1%) obtained for PARAFAC scores (Fig. 7C/D). Although the Trp (PFC1) trends showed that there were no significant structural changes in the rIgG over the 15° to 35 °C temperature range, while the PFC2 scores (Tyr emission) increased very slightly with increasing temperature (Fig. 7D) due to Trp quenching. The only significant differences in the scores for the different ASDR methods was that the interpolated data generally had higher PFC2 scores. This would suggest that this was the best option for producing protein fluorescence data suitable for PARAFAC modelling to monitor subtle changes in protein behavior and structure via changes in the ratio between Tyr and Trp emission. Here, the structural changes are very small, and the increased thermal motion generates a linear decrease in emission intensity due to quenching of Trp emission.

4. Conclusions

We have demonstrated that pTSFS is a better measurement method than pEEM for PARAFAC analysis of intrinsic protein fluorescence, and in particular at the blue edge of the emission related to Tyr fluorescence. This is because pTSFS minimized the collection of Rayleigh scattered light which caused a lot of problems at the emission blue edge in pEEM

spectra, largely due to the difficulty in removing the Rayleigh shot noise [29]. A simple transformation of TSFS data into an pEEM like layout (t-pEEM) provided the trilinear structure [30] required for PARAFAC analysis. However, the t-pEEM transformation, produced an ASDR (<8%) at the blue edge of the emission which is critical for resolving Tyr emission. Evaluation of several data handling methods, missing data, use of zeros, and interpolation, to deal with this issue revealed that PARAFAC results were very method dependent. From a spectroscopic perspective, in each case PARAFAC separated Tyr from Trp emission, thus, confirming the superior performance of TSFS measurements compared to EEM [29].

This study showed that significantly different PARAFAC outputs (spectral profiles and scores) were obtained when using different methods for dealing with ASDR. The use of zeros and missing data to fill the ASDR both produced PARAFAC models which potentially underestimated the relative contribution of the Tyr component (Fig. 7). Both methods were susceptible to producing spectral artifacts, particularly at the blue edge of the emission where the noise and light scatter contamination was an issue. One drawback for the use of missing data was the high computational time required for this method, making it prohibitive for online or inline applications. The most suitable option, t-pEEM in combination with interpolation generated the best quality PARAFAC models from the point of view of the spectral profiles and scores recovered. However, one must be cautious when interpreting this data and in particular the spectral profiles as the validation of PARAFAC output from intrinsic protein emission is still unproven (although known for small molecule examples). What has been proven here is that the general spectral trends observed in the scores were similar but the magnitudes were significantly different and thus when using scores plots to follow protein changes one might be advised to use a combination of two ASDR approaches (missing data and interpolation) to ensure a useable model outcome for analytical purposes (*e.g.* protein stability studies, IgG product variance, etc.).

The PARAFAC modelling also suggested that there was still some residual scattered light in the pTSFS data (which is unsurprising given the 10 nm slit widths used). This was seen in the different profiles recovered for the second component (mostly Tyr) in the $TSFS_{\parallel}$ and $TSFS_{\perp}$ data. Ultimately, this indicated that to extract a true uncontaminated Tyr emission component, and thus accurate scores, one needs to use narrower slits, which will decrease signal intensity that can only be recovered by increasing acquisition time. This is not practical currently for ARMES using conventional benchtop spectrometers where four measurements are required (and particularly for rapid or high throughput analytical applications). The $TSFS_{\perp}$ measurement offered the best quality data for modelling Tyr/Trp emission because it had the least amount of residual scattered light and thus generated the most reasonable estimate of the true contributions of each fluorophore to the total emission. However, the second $TSFS_{\perp}$ recovered component still does not represent Tyr only emission and there is some associated Trp emission present. Despite this, the ratio of PARAFAC scores generated by these pTSFS/PARAFAC models provides a unique and robust tool for assessing the change in emission caused by structural changes impacting on FRET.

In conclusion, this study, when considered with our previous pEEM study [29], has shown that the combination of pTSFS measurements, with precisely defined pre-processing, and PARAFAC modelling is the most suitable method for the analysis of intrinsic protein fluorescence. The pTSFS based methodology shows sufficient sensitivity to extract signals from Tyr emission which is key to being able to observe and measure subtle changes in protein structure [8]. This is of particular importance for understanding the mechanisms of aggregation [48] and protein degradation which are fundamental to the delivery of safe bio-therapeutic agents.

Declaration of competing interest

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Acknowledgements

This publication 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 under Grand number (14/IA/2282, Advanced Analytics for Biological Therapeutic Manufacture, to AGR). We also thank Agilent Technologies (Mulgrave Victoria, Australia) for the loan of a fluorescence spectrometer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2019.103871.

References

[1] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, third ed. ed., Springer, New York, 2006.
[2] J.M. Reichert, Marketed therapeutic antibodies compendium, mAbs 4 (2012) 413–415.
[3] D.M. Ecker, S.D. Jones, H.L. Levine, The therapeutic monoclonal antibody market, mAbs 7 (2015) 9–14.
[4] B. Sharma, Immunogenicity of therapeutic proteins. Part 1: impact of product handling, Biotechnol. Adv. 25 (2007) 310–317.
[5] K. Wiberg, A. Sterner-Molin, S.P. Jacobsson, Simultaneous determination of albumin and immunoglobulin G with fluorescence spectroscopy and multivariate calibration, Talanta 62 (2004) 567–574.
[6] E.A. Burstein, N.S. Vedenkina, M.N. Ivkova, Fluorescence and location of tryptophan residues in protein molecules, Photochem. Photobiol. 18 (1973) 263–279.
[7] P. Garidel, M. Hegyi, S. Bassarab, M. Weichel, A rapid, sensitive and economical assessment of monoclonal antibody conformational stability by intrinsic tryptophan fluorescence spectroscopy, Biotechnol. J. 3 (2008) 1201–1211.
[8] L. Liu, L.J. Braun, W. Wang, T.W. Randolph, J.F. Carpenter, Freezing-induced perturbation of tertiary structure of a monoclonal antibody, J. Pharm. Sci. 103 (2014) 1979–1986.
[9] R.C. Groza, B. Li, A.G. Ryder, Anisotropy resolved multidimensional emission spectroscopy (ARMES): a new tool for protein analysis, Anal. Chim. Acta 886 (2015) 133–142.
[10] K. Ohadi, R.L. Legge, H.M. Budman, Intrinsic fluorescence-based at situ soft sensor for monitoring monoclonal antibody aggregation, Biotechnol. Prog. 31 (2015) 1423–1432.
[11] G. Weber, Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds, Biochem. J. 75 (1960) 335–345.
[12] J. Eisinger, Intramolecular energy transfer in adrenocorticotropin, Biochemistry 8 (1969) 3902–&.
[13] H.C. Chiu, R. Bersohn, Electronic-energy transfer between tyrosine and tryptophan in peptides trp-(pro)N-Tyr, Biopolymers 16 (1977) 277–288.
[14] S. Bobone, M. van de Weert, L. Stella, A reassessment of synchronous fluorescence in the separation of Trp and Tyr contributions in protein emission and in the determination of conformational changes, J. Mol. Struct. 1077 (2014) 68–76.
[15] I.M. Warner, G.D. Christian, E.R. Davidson, J.B. Callis, Analysis of multicomponent fluorescence data, Anal. Chem. 49 (1977) 564–573.
[16] D. Patra, A.K. Mishra, Recent developments in multi-component synchronous fluorescence scan analysis, Trac. Trends Anal. Chem. 21 (2002) 787–798.
[17] K. Kumar, A.K. Mishra, Analysis of dilute aqueous multifluorophoric mixtures using excitation-emission matrix fluorescence (EEMF) and total synchronous fluorescence (TSF) spectroscopy: a comparative evaluation, Talanta 117 (2013) 209–220.
[18] B. Li, P.W. Ryan, M. Shanahan, K.J. Leister, A.G. Ryder, Fluorescence excitation-emission matrix (EEM) spectroscopy for rapid identification and quality evaluation of cell culture media components, Appl. Spectrosc. 65 (2011) 1240–1249.
[19] B. Li, M. Shanahan, A. Calvet, K.J. Leister, A.G. Ryder, Comprehensive, quantitative bioprocess productivity monitoring using fluorescence EEM spectroscopy and chemometrics, Analyst 139 (2014) 1661–1671.
[20] Y. Casamayou-Boucau, A.G. Ryder, Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters, validation standards, and Rayleigh scatter removal methods, Methods Appl. Fluoresc. 5 (2017), 037001.
[21] H. Abdollahi, R. Tauler, Uniqueness and rotation ambiguities in multivariate curve resolution methods, Chemometr. Intell. Lab. Syst. 108 (2011) 100–111.
[22] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: a review of advanced and tailored applications and challenges, Anal. Chim. Acta 765 (2013) 28–36.
[23] R. Bro, PARAFAC. Tutorial and applications, Chemometr. Intell. Lab. Syst. 38 (1997) 149–171.
[24] C.M. Andersen, R. Bro, Practical aspects of PARAFAC modeling of fluorescence excitation-emission data, J. Chemom. 17 (2003) 200–215.
[25] K.R. Murphy, C.A. Stedmon, D. Graeber, R. Bro, Fluorescence spectroscopy and multi-way techniques. PARAFAC, Anal. Methods 5 (2013) 6557–6566.
[26] S.D. Brown, L.A. Sarabia, J. Trygg, Comprehensive Chemometrics Chemical and Biochemical Data Analysis, Elsevier, Amsterdam, 2009. Amsterdam.

[27] J.M. Amigo, F. Marini, Multiway methods, in: F. Marini (Ed.), Chemometrics in Food Chemistry, Elsevier Science Bv, Amsterdam, 2013, pp. 265–313.
[28] R.C. Groza, A. Calvet, A.G. Ryder, A fluorescence anisotropy method for measuring protein concentration in complex cell culture media, Anal. Chim. Acta 821 (2014) 54–61.
[29] M. Steiner-Browne, S. Elcoroaristizabal, Y. Casamayou-Boucau, A.G. Ryder, Investigating native state fluorescence emission of Immunoglobulin G using polarized Excitation Emission Matrix (pEEM) spectroscopy and PARAFAC, Chemometr. Intell. Lab. Syst. 185 (2019) 1–11.
[30] K. Kumar, A.K. Mishra, Application of parallel factor analysis to total synchronous fluorescence spectrum of dilute multifluorophoric solutions: addressing the issue of lack of trilinearity in total synchronous fluorescence data set, Anal. Chim. Acta 755 (2012) 37–45.
[31] K. Kumar, A.K. Mishra, Parallel factor (PARAFAC) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets in excitation-emission matrix fluorescence (EEMF) layout: certain practical aspects, Chemometr. Intell. Lab. Syst. 147 (2015) 121–130.
[32] S. Elcoroaristizabal, R. Bro, J.A. Garcia, L. Alonso, PARAFAC models of fluorescence data with scattering: a comparative study, Chemometr. Intell. Lab. Syst. 142 (2015) 124–130.
[33] K. Kumar, A.K. Mishra, Multivariate curve resolution alternating least square (MCR-ALS) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets: comparing certain ways of arranging TSFS-based three-way array, Chemometr. Intell. Lab. Syst. 147 (2015) 66–74.
[34] Y. Casamayou-Boucau, A.G. Ryder, Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR), Anal. Chim. Acta 1000 (2018) 132–143.
[35] K. Kumar, A.K. Mishra, Understanding the effect of calibration set design for the application of MCR-ALS analysis on excitation-emission matrix fluorescence (EEMF) data sets under commonly used non-negativity constraints, Chemometr. Intell. Lab. Syst. 149 (2015) 70–77.
[36] S. Elcoroaristizabal, A. de Juan, J.A. Garcia, N. Durana, L. Alonso, Comparison of second-order multivariate methods for screening and determination of PAHs by total fluorescence spectroscopy, Chemometr. Intell. Lab. Syst. 132 (2014) 63–74.
[37] M.R. Eftink, L.A. Selvidge, P.R. Callis, A.A. Rehms, Photophysics of indole-derivatives - experimental resolution of La and Lb transition and comparison with theory, J. Phys. Chem. 94 (1990) 3469–3479.
[38] M. Ameloot, M. vandeVen, A.U. Acuna, B. Valeur, Fluorescence anisotropy measurements in solution: methods and reference materials (IUPAC Technical Report), Pure Appl. Chem. 85 (2013) 589–608.
[39] D.N. Kothawala, K.R. Murphy, C.A. Stedmon, G.A. Weyhenmeyer, L.J. Tranvik, Inner filter correction of dissolved organic matter fluorescence, Limnol Oceanogr. Methods 11 (2013) 616–630.
[40] M. Tarai, A.K. Mishra, Inner filter effect and the onset of concentration dependent red shift of synchronous fluorescence spectra, Anal. Chim. Acta 940 (2016) 113–119.
[41] G. Tomasi, R. Bro, PARAFAC and missing values, Chemometr. Intell. Lab. Syst. 75 (2005) 163–180.
[42] L.G. Thygesen, A. Rinnan, S. Barsberg, J.K.S. Moller, Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area, Chemometr. Intell. Lab. Syst. 71 (2004) 97–106.
[43] B. Li, A.G. Ryder, Similarity Index: a Rapid Classification Method for Multivariate Data Arrays, 2011. US7983874B2.
[44] D.C. Hanson, J. Yguerabide, V.N. Schumaker, Segmental flexibility of immunoglobulin-G antibody molecules in solution - a new interpretation, Biochemistry 20 (1981) 6842–6852.
[45] K. Lim, D.M. Jameson, C.A. Gentry, J.N. Herron, Molecular-dynamics of the antifluorescein-4-4-20 antigen-binding fragment. 2. Time-resolved fluorescence spectroscopy, Biochemistry 34 (1995) 6975–6984.
[46] T.J. Kamerzell, J.D. Ramsey, C.R. Middaugh, Immunoglobulin dynamics, conformational fluctuations, and nonlinear elasticity and their effects on stability, J. Phys. Chem. B 112 (2008) 3240–3250.
[47] V.P. Timofeev, I.V. Dudich, Y.K. Sykulev, R.S. Nezlin, Rotational correlation times of IgG and its fragments spin-labeled at carbohydrate or protein moieties - spatially fixed position of fc carbohydrate, FEBS Lett. 89 (1978) 191–195.
[48] Z. Hamrang, N.J.W. Rattray, A. Pluen, Proteins behaving badly: emerging technologies in profiling biopharmaceutical aggregation, Trends Biotechnol. 31 (2013) 448–458.