



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Anisotropy resolved multidimensional emission spectroscopy (ARMES) for the analysis and resolution of insulin oligomer emission
Author(s)	Casamayou-Boucau, Yannick
Publication Date	2018-11-14
Publisher	NUI Galway
Item record	<a href="http://hdl.handle.net/10379/14931">http://hdl.handle.net/10379/14931</a>

Downloaded 2024-04-24T08:36:09Z

Some rights reserved. For more information, please see the item record link above.





National University of Ireland – Galway  
College of Science, School of Chemistry

**Anisotropy Resolved Multidimensional  
Emission Spectroscopy (ARMES) for the  
analysis and resolution of insulin oligomer  
emission**

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

by **Yannick Casamayou-Boucau, MSc**

Supervisor: **Prof. Alan Ryder**

– November 2018 –

# Acknowledgments

Several events in my life brought me to Ireland and strangely brought me back to chemistry. Without being too philosophical about it, I believe that sometimes things are meant to be, and I sincerely thank Alan for giving me a chance to join his team back in 2014. At that time, I was hired as a research assistant, barely spoke English, and decided to start a PhD in a field that was almost unknown to me. I had to learn a lot during the last years, in terms of science of course but not only as the PhD is above all, one of the best self-awareness experiences that one can do. So, for guiding me throughout the thesis but also for leading me towards a life experience that I would have otherwise missed<sup>a</sup>, I thank again my supervisor Alan Ryder.

I also want to gratefully acknowledge the financial support from the Irish Research Council, without which this work would have not been possible. My sincere gratitude goes also towards my examiners Prof. Peter Crowley and Åsmund Rinnan, for making my viva very enjoyable with a stimulating discussion and interesting suggestions.

Next on the list are all my lab colleagues that are for most my friends too. I am especially thankful to Boyan, who patiently taught me some rudiments about data analysis when I started. I also want to express my gratitude to the “old” lab team Amandine, Cheryl, Bridget for being good instructors and guiding me in the lab. A new and very international (multidimensional?) group formed since, and to all of you, I thank you for your support and the craic you all brought to the lab and my life in Galway.

It is difficult to dedicate this work to a single person. My family has always been the core of my life, and I know how lucky I am to have them. Ama, Aita, Audrey-zink eta Amañi, this work is for you and I profoundly thank you for always trusting in me. I also want to thank and dedicate this work to Helen, the ledge, who brought me all the help and support I could ever hope in Ireland. To complete the circle of acknowledgements, I also want to mention my mates at home, that are crucial in my life for many years and having them is also part of it all: Couro, Romano, Babaxteko, Jeje, Chiki, Bebe, Fab, Andde, Funky, Seb, Charlo, Beñat, this work is also for you.

*“Aditu metatzeko laster egin garraitz eko”*

---

<sup>a</sup> No, I am not a masochist!

# Contents

<b>Abstract.....</b>	<b>9</b>
<b>1. General concepts about fluorescence.....</b>	<b>11</b>
1.1 Basic principle of fluorescence .....	11
1.2 Fluorophores and their characteristics.....	14
1.2.1 Excitation, emission, and Stokes shift .....	14
1.2.2 The fluorescence quantum yield .....	15
1.2.3 The fluorescence lifetime.....	16
1.3 Influence of intermolecular photophysical processes on the emission properties of fluorophores .....	18
1.3.1 Static and dynamic quenching .....	18
1.3.2 Förster Resonance Energy Transfer (FRET) .....	21
1.3.3 Inner filter effects (IFE).....	25
<b>2. Material and Methods.....</b>	<b>27</b>
2.1 Reagents .....	27
2.2 General material and instrumentation .....	28
2.3 Anisotropy standards solutions .....	30
2.3.1 Stock solutions for filter validation.....	30
2.3.2 Fluorophore mixtures for chemometric modelling.....	31
2.3.3 Parameters for the data collection.....	32
2.4 Insulin study .....	33
2.4.1 Human insulin solutions .....	33
2.4.2 L-Tyrosine solutions in “insulin buffers” .....	34
2.4.3 Parameters for the data collection.....	35

<b>3. Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurement: better filters and validation standards.....</b>	<b>36</b>
3.1 Fluorescence polarization and anisotropy .....	36
3.1.1 Applications of conventional fluorescence polarization spectroscopy and polarization imaging .....	40
3.1.2 Multidimensional fluorescence (MDF) spectroscopy and the development of ARMES.....	42
3.1.3 Chapter goals .....	45
3.2 Adaption of wire-grid polarizers to spectrofluorometer .....	46
3.3 Perylene: a potential ARMES standard.....	49
3.4 Comparison of filter efficiency using perylene: <i>conventional 2-D anisotropy</i> .....	52
3.5 Temperature dependence study to confirm dWGP accuracy .....	56
3.6 Validation of dWGP accuracy in the 450 to 640 nm range .....	58
3.7 Conclusions .....	61
<b>4. Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR).....</b>	<b>62</b>
4.1 Introduction .....	62
4.1.1 MCR decomposition .....	63
4.1.2 MCR: a bilinear model.....	65
4.1.3 Iterative and non-iterative approaches .....	66
4.1.4 Importance of initial estimates for MCR decomposition.....	68
4.1.5 Ambiguities in MCR.....	69
4.1.6 Constraints and advantages of MCR-ALS.....	71
4.1.7 Extension of MCR to multi-way data .....	73
4.1.8 Concept of trilinearity and multilinear constraints .....	76
4.2 The EEM and TSFS measurements .....	78
4.3 Accurate Rayleigh scatter removal in EEM.....	80
4.3.1 Rinnan's method and threshold level.....	81

4.3.2	Analysis of corrected <i>aniso</i> -MDF spectra: increase of perylene's working range as multidimensional anisotropy standard.....	83
4.4	Anisotropy recovery from fluorophore mixtures using MCR-ALS.....	84
4.4.1	Problem statement.....	84
4.4.2	Materials and data analysis.....	86
4.4.3	Results and Discussion.....	87
4.5	Conclusions.....	102
<b>5.</b>	<b>Elucidating the intrinsic fluorescence of human insulin and its different association states.....</b>	<b>103</b>
5.1	Introduction.....	103
5.2	The molecule of insulin.....	104
5.2.1	Insulin structure.....	104
5.2.2	The natural aggregation surfaces of insulin.....	112
5.3	Measuring protein aggregation: from traditional methods to motivations for new alternatives.....	116
5.4	The steady-state fluorescence of Tyrosine and Insulin.....	120
5.4.1	Tyrosine fluorescence.....	120
5.4.2	Fluorescence methods used to measure insulin aggregation.....	123
5.5	Materials and data analysis.....	128
5.6	Behaviour of Tyr in various buffers.....	132
5.7	Analysis of insulin UV-Vis data.....	134
5.8	PARAFAC modelling of insulin monomer emission.....	141
5.9	Spectroscopic analysis of the insulin ARMES data at 0.2 and 2 mg/mL, or how to dissect the measurement.....	149
5.9.1	Utilization of the 1 <sup>st</sup> order Rayleigh scatter band.....	149
5.9.2	Appearance of slight shifts in the fluorescence.....	156
5.9.3	Analysis of the 2-D <i>aniso</i> and <i>aniso</i> -MDF data.....	157

5.10	Chemometric analysis of ARMES data (0.2 mg/mL): <i>towards resolving the complex Tyr overlap in insulin fluorescence</i> .....	161
5.10.1	EEM <sub>  </sub> .....	161
5.10.2	EEM <sub>⊥</sub> .....	172
5.11	Models using 2 mg/mL samples: non-IFE corrected data for separating highly concentrated solutions.....	174
5.12	Conclusions .....	178
<b>6.</b>	<b>Conclusions.....</b>	<b>179</b>
<b>7.</b>	<b>Appendices.....</b>	<b>183</b>
7.1	Further analysis on filters transmittance spectra .....	183
7.2	AutoCAD design for dWGP holder .....	184
7.3	Influence of integration time .....	185
7.4	<i>Aniso</i> -MDF spectra, threshold and measurement dispersion.....	186
7.5	Slight modifications of Rinnan's method to deal with 2 <sup>nd</sup> order Rayleigh scatter..	187
7.6	Presence of an impurity in the glycerol.....	188
7.7	CWA-MCR modelled perylene data .....	188
7.8	SWA-MCR modelled perylene data .....	195
7.9	Fluorophore mixture modelling .....	198
7.10	Some historical and general information about insulin.....	201
7.10.1	The history behind the discovery of insulin.....	201
7.10.2	Evolution of insulin formulations .....	203
7.10.3	The current insulin market .....	205
7.11	The primary structure elucidation of insulin .....	206
7.12	Historical details behind the first elucidation of insulin's crystal structure .....	208
7.13	Insulin: a molecule with similar crystal and solution structures .....	208
7.14	The T/R transition: impacts on insulin hexamer structure and biological relevance.....	210
7.15	Details about Tryptophan steady-state fluorescence .....	214

7.16	Certificate analysis of Human insulin lot used for this thesis .....	215
7.17	Details for the G factor denoising.....	216
7.18	PARAFAC models on the L-Tyr data .....	218
7.19	Details on the PCA/MCR models built on the UV-Vis data .....	224
7.20	Repeated PARAFAC models obtained on the EEM <sub>  </sub> monomeric insulin data at 0.2 mg/mL.....	230
7.21	Fluorescence of Tyrosinate in the various buffers used .....	230
7.22	Details about the analysis of insulin <i>aniso</i> -MDF spectra .....	231
7.23	Robustness of the PARAFAC model built using the EEM <sub>  </sub> insulin oligomers data, at 0.2 mg/mL.....	232
7.24	Robustness of the PARAFAC model built using the EEM <sub>⊥</sub> insulin oligomers data, at 0.2 mg/mL.....	233
7.25	Further information about the PARAFAC models built using the 2 mg/mL insulin samples.....	235
7.26	Denaturation of insulin monomers by GuHCl.....	236
<b>References.....</b>		<b>239</b>
<b>List of scientific achievements related to this work.....</b>		<b>274</b>

## Commonly used abbreviations

**ARMES:** Anisotropy resolved multidimensional emission spectroscopy

**AVG:** Average

**CD:** Circular dichroism

**CWA:** Column wise augmentation

**DLS:** Dynamic light scattering

**dWGP:** Dual wire grid polarizers

**EB:** Erythrosin B

**EEM:** Excitation emission matrix

**FRET:** Förster resonance energy transfer

**G:** G factor

**HDI:** Human insulin dimer

**HHX:** Human insulin hexamer

**HMO:** Human insulin monomer

**HTE:** Human insulin tetramer

**IFE:** Inner filter effects

**MCR:** Multivariate curve resolution

**MDF:** Multidimensional fluorescence

**PARAFAC:** Parallel factor analysis

**PAT:** Process analytical technology

**PB:** Phloxine B

**PCA:** Principal component analysis

**Phe:** Phenylalanine

**r:** Anisotropy

**RSD:** Relative standard deviation

**S<sub>1</sub>:** First singlet excited state

**SEC:** Size exclusion chromatography

**SNR:** Signal sur noise ratio

**STD:** Standard deviation

**SWA:** Sample wise augmentation

**TFP:** Thin film polarizers

**Trp:** Tryptophan

**TSFS:** Total synchronous fluorescence spectroscopy

**Tyr:** Tyrosine

**UV-Vis:** Ultraviolet-Visible spectroscopy

## Abstract

Protein aggregation is one of the biggest challenges affecting the manufacture and safe use of biopharmaceuticals. Despite efforts made to enhance protein stability<sup>1</sup>, aggregation still remains a critical factor in product safety<sup>2</sup> as it can affect bioavailability and increase immunogenicity problems<sup>3-4</sup>. Aggregation can occur at every step of the bio-pharma process, from production to protein delivery and can be induced by a multitude of factors<sup>5</sup>. Detecting and quantifying protein aggregation is a considerable analytical challenge, due to the often unknown nature of the aggregates, their wide size range (nm to mm)<sup>6</sup> and the varying concentration ranges encountered. Within the aggregate family, being able to detect the soluble nm size aggregates is probably the most difficult/critical part as they form the seeds of bigger aggregation issues<sup>7-8</sup>. Many current methods for aggregation monitoring are either destructive, only capable of measuring large precipitated aggregates, expensive/time-consuming and/or affect the original distribution of oligomers in the sample<sup>9</sup>. Therefore, better analytical methods that are capable of identifying and quantifying small soluble aggregates are needed. The methods must deliver good robustness, high sample throughput, ease of use and low cost required for use as a process analytical technology (PAT) tools.

Protein analysis by intrinsic fluorescence is very attractive because in addition to being fast, non-destructive, sensitive and inexpensive, it is also non-invasive and completely preserves the structure and behavior of proteins as well as the potential distribution of oligomers<sup>10</sup>. The major issue however is that the multiple fluorophores present in proteins are generally in close proximity (<15 nm) and thus interact photophysically, resulting in a complex emission space and highly overlapping signals. Multidimensional fluorescence spectroscopy (MDF) measurement techniques like excitation-emission matrix (EEM)<sup>11</sup> and total synchronous fluorescence scan (TSFS)<sup>12</sup>, can be coupled with factor based chemometric methods to partially overcome this issue by deconvoluting the MDF data and resolving fluorophore contributions. Recent work in Alan Ryder's group has shown that more and better information could be obtained if MDF is combined with anisotropy ( $r$ )<sup>13-14</sup>, which is related to rotational speeds and hydrodynamic volumes, and thus molecular size. This method was called Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES)<sup>14</sup> and is the foundation for this research.

The first part of this thesis project was to improve the ARMES hardware and enable the acquisition of polarized emission in the important tyrosine/tryptophan region. This was solved by replacing the classically fitted thin film polarizers with wire grid polarizers that have a better UV transmittance. To correct for the weak efficiency of these filters in the important protein region, the wire grid polarizers had to be used in pairs. The efficacy of these new filters was validated with various anisotropy standards, and a homebuilt fitting was designed to mount them in the sample compartment of the spectrofluorometer (**Chapter 3, Casamayou-Boucau *et al.*<sup>15</sup>, 2017**).

The use of UV transmitting filters resulted however in the elimination of the spectral reshaping previously observed in the TSFS-ARMES studies<sup>13-14</sup>, which had consequences on the structure of the TSFS data being generated. Two mixtures of non-interacting fluorophores were used to investigate that issue and standardize the chemometric methodology to be used. The results from this study showed for the first time how to accurately model component anisotropy using multivariate curve resolution (MCR) from data collected using TSFS or EEM. The change in TSFS data structure was also confirmed and their non-trilinear behaviour restored. While the correct data augmentation strategy enabled TSFS to yield acceptable results, EEM data proved to be more accurate once an efficient method was found to remove Rayleigh scatter (**Chapter 4, Casamayou-Boucau *et al.*<sup>16</sup>, 2018**).

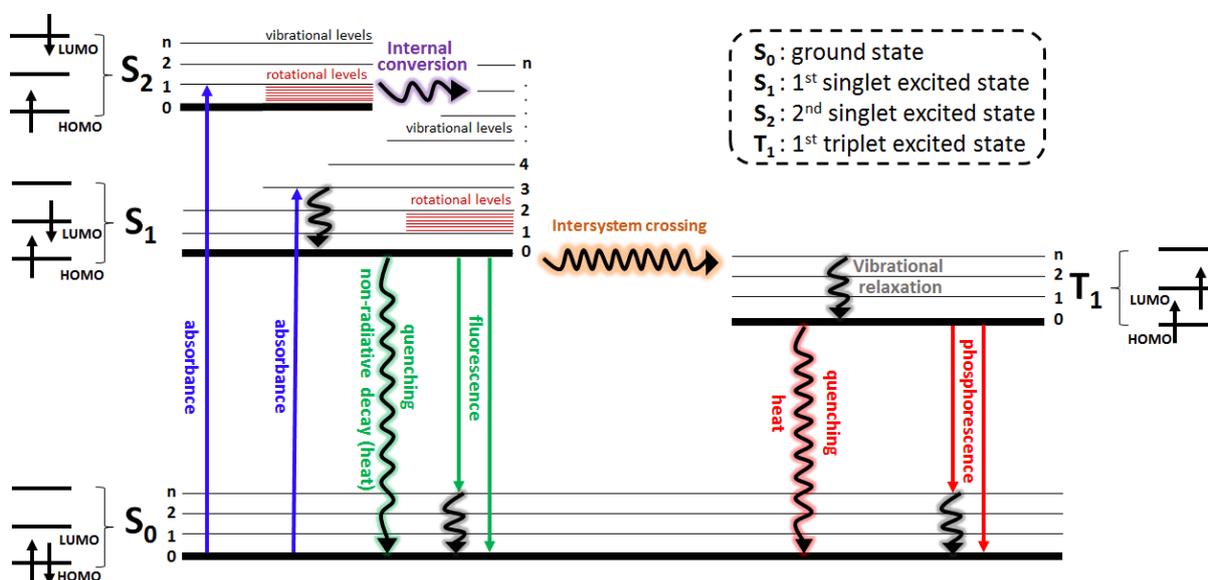
Having improved and validated ARMES hardware and chemometrics, the method was ready to be applied to intrinsic fluorescence analysis of insulin. The work in this thesis focused on the analysis of small soluble non-covalent aggregates of human insulin, which are difficult to differentiate spectroscopically<sup>17</sup> and not suitable for conventional size exclusion chromatography (SEC)<sup>18</sup>. By carefully dissecting the EEM-ARMES measurement into pure scatter and fluorescence information, unique tools such as the polarized 1<sup>st</sup> order Rayleigh scatter band, the scatter depolarization, and the denoised *aniso*-MDF spectra could be used to assess, either qualitatively or quantitatively, the degree of insulin aggregation (**Chapter 5**). Furthermore, due to the location of insulin's tyrosine residues in the regions responsible for dimer and hexamer formation, it was possible for the first time to resolve the intrinsic fluorescence of insulin and detect clearly the presence of internal energy transfer (**Chapter 5**). The linear dependencies usually observed were broken, permitting the building of robust models for discriminating the small soluble insulin aggregates. Thus **Chapter 5** demonstrated the potential of ARMES for protein analysis and showed how multiple probes were obtained in a single non-invasive measurement.

# 1. General concepts about fluorescence

The work developed in this thesis is quite inter-disciplinary with optical, biophysical, mathematical, and protein structural considerations. As such, we believed it was better to introduce in each chapter separately all the concepts necessary to understand the subsequent work. The basis of the analytical method is, however, fluorescence, and this short chapter focuses on introducing some of the general fluorescence concepts that are encountered throughout the thesis.

## 1.1 Basic principle of fluorescence

In the last few decades, the use of fluorescence spectroscopy has grown remarkably in the biochemical and biophysical fields<sup>19</sup>. This technique presents many advantages such as being highly sensitive, non-destructive, non-invasive and very responsive to changes in the microenvironment of the molecule studied. For all these reasons, fluorescence is nowadays one of the main methods used to characterise biomolecules<sup>10,20</sup>.



**Figure 1:** This illustration represents one form of the Jablonski diagram<sup>19</sup>.

The earliest observations of photoluminescence date back to the middle ages, as several texts reported minerals glowing in the dark<sup>21</sup>. The topic of photoluminescence is large and encompasses all the various phenomena (*e.g.* fluorescence, phosphorescence, thermoluminescence etc.) that cause a molecule to emit light after being photo-excited. The

particular concept of fluorescence was however only starting to be described in the middle of the 19<sup>th</sup> century, following the work<sup>22</sup> of Sir G. G. Stokes. Along with E. Becquerel, he was the first to have found that the wavelength of emitted light is always of longer wavelength (*i.e.* lower energy) than the exciting light<sup>21</sup>. This difference of energy, or Stokes shift, is still nowadays one of the fundamental concepts in fluorescence. Since Stokes, many other scientists have significantly contributed to the development of this field such as Stern & Volmer, Jablonski (diagram presented Figure 1), Lewis, Kasha, Förster, etc.

At room temperature, molecules normally occupy the lowest vibrational state of the ground state ( $S_0$ ). When the light exciting a molecule has an energy corresponding to a possible electronic transition, the photons are absorbed and electrons are promoted to a higher molecular orbital. Depending on the energy of the photons, electrons occupying the HOMO<sup>b</sup> ( $S_0$ ) are promoted to the first ( $S_1$ ) or a higher singlet excited state ( $S_n$ ) which defines the LUMO<sup>23</sup>. The excited molecules can in fact populate any of the vibrational or rotational sublevels that are associated with each electronic state. Each of these transitions should form a distinct band in the absorption spectrum as the molecules absorb the light in discrete amounts (or quanta), but in reality a broad spectrum is obtained which can be imagined as an almost continuous succession of sublevel transitions<sup>23</sup>.

On a time scale, the absorption of light is the fastest<sup>19,23</sup> of all processes presented in Figure 1 and occurs in  $\sim 10^{-15}$  s. Due to the instability of the excited state however, the molecules always return to the ground state, that is, the one that is thermodynamically stable at room temperature. Molecules have many different paths to return to the ground state (Figure 1). Most of the time, they are excited to some vibrational levels and electronic states that are higher than the lowest vibrational state of the 1<sup>st</sup> excited state ( $S_1$ ). In that case, a non-radiative and fast internal conversion ( $10^{-12}$  s or less<sup>19,23</sup>), possible only between electronic states sharing the same spin multiplicity, brings the molecule from *e.g.* the  $S_2$  to the  $S_1$  state. This is directly followed by vibrational relaxation ( $10^{-12}$  to  $10^{-10}$  s), enabling the molecule to reach the lowest vibrational level of the  $S_1$  state while the excess energy produced by the relaxation is dissipated as heat in the solvent. At this point, the gap of energy between the  $S_1$  and  $S_0$  state is generally too big for internal conversion to occur, but a variety of other mechanisms can bring the molecule to the ground state.

---

<sup>b</sup> The HOMO and LUMO refer respectively to the highest occupied molecular orbital and to the lowest unoccupied molecular orbitals.

One of these mechanisms is called fluorescence and corresponds to the relaxation from the  $S_1$  to the  $S_0$  state by the radiative emission of a photon<sup>19,23</sup>. Due to the previous vibrational relaxation that decreased the energy of the molecule, the emitted photon is always of lower energy (longer wavelength) than the one absorbed (which brings us back to the Stokes shift). As shown in Figure 1, the molecule can then reach any of the vibrational levels associated with the ground state  $S_0$ . This property is responsible for the structure of perylene emission for example<sup>19</sup>, and more details can be found in section 3.3 (chapter 3). Similarly, as stated in the previous paragraph, fluorescence occurs from the lowest vibrational level of the  $S_1$  state, but there are exceptions to this rule and more information is given in the same section. Compared to the processes described previously, fluorescence is slower and occurs typically in  $10^{-10}$  to  $10^{-7}$  seconds. This time corresponds to the average time that the molecules spend in the excited state prior to emission and is called the fluorescence lifetime. The photon emission itself on the other hand is as fast as the photon absorption ( $10^{-15}$  s)<sup>23</sup>.

The other mechanism that can relax a molecule from the  $S_1$  state is caused by intersystem crossing towards the triplet ( $T_1$ ) state (Figure 1), which involves the spin inversion of the excited electron. In the ground state, two electrons can populate the HOMO only if they have opposite spins ( $1/2, -1/2$ ) which is the exclusion principle of Pauli<sup>24</sup>. The total spin ( $s$ ) is thus zero defining a spin multiplicity ( $M = 2 \times s + 1$ ) of one. For that reason,  $S_0$  but also  $S_1, S_2$  etc. are called singlet states, and the absorption and fluorescence emission of a photon are two phenomena that are spin-allowed in quantum theory, as they engage electronic states ( $S_0, S_1$ ) that have the same multiplicity ( $M = 1$ )<sup>25</sup>. However, in some cases, especially when the spin-orbit coupling is strong enough, an electron can undergo a spin inversion so that its orientation is the same as the HOMO electron ( $s = 1/2 + 1/2 = 1$ ). This conversion from the  $S_1$  state to a state of multiplicity ( $M = 3$ ) called the 1<sup>st</sup> triplet state  $T_1$  is known as intersystem crossing<sup>19,25</sup> (Figure 1). It is described as a non-radiative transition between isoenergetic vibrational states of  $S_1$  and  $T_1$ , which occurs in  $10^{-10}$  to  $10^{-8}$  seconds. Once in the  $T_1$  state, photons can eventually be emitted so that the molecule returns to the ground state: this process is known as phosphorescence. The emission generally occurs from the lowest vibrational state of  $T_1$  which has a lower energy than  $S_1$ , explaining why the phosphorescence of a molecule is always occurring at a longer wavelength than its fluorescence<sup>25</sup>. Another aspect of this phenomena is that it is very slow, with lifetimes that range from  $10^{-5}$  seconds to seconds<sup>19,25</sup>. The transition from  $T_1$  to  $S_0$  is in fact forbidden in quantum theory (both states have different multiplicity)

which results in rate constants for triple state emission that are several orders of magnitude smaller than *e.g.* those for  $S_1$ .

There are other pathways for the excited molecules to return to the ground state (Figure 1). Instead of releasing the energy by emitting a photon, the energy can, for example, be dissipated as heat to the surrounding solvent molecules. Another common mechanism is when excited molecules transfer their energy to other molecules via collisions or dipole-dipole coupling. These types of non-radiative decays are thus competing with the fluorescence or phosphorescence and are known as quenching effects (see section 1.3 for more details).

## 1.2 Fluorophores and their characteristics

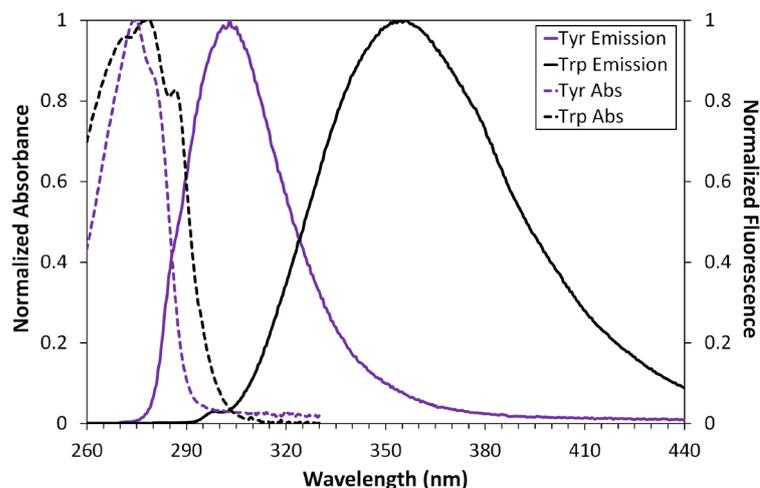
Fluorophores are molecules that absorb and emit light. Many fluorophores are small organic molecules, synthesized to have very precise fluorescence properties (*e.g.* rhodamines), and are used as extrinsic probes, for example, in the study of tissues and cells but also for studying macromolecules<sup>26-27</sup>. However many proteins are also intrinsically fluorescent, due to the presence of three amino-acids: tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe)<sup>28</sup>. The complex fluorescence properties of these amino-acids as well as their dependence on their surrounding microenvironment is discussed in detail in the sections 5.4.1 and 7.15. Nevertheless, whether the fluorophores are extrinsic or intrinsic, they are all defined by five specific characteristics: their excitation / emission spectra (and Stokes shift), quantum yield, lifetime, and anisotropy. A more detailed discussion on anisotropy can be found in the section 3.1. The other characteristics are detailed below.

### 1.2.1 Excitation, emission, and Stokes shift

The emission spectrum of a fluorophore is measured by fixing the excitation wavelength  $\lambda_{\text{ex}}$  and measuring the fluorescence intensity over a range of emission wavelengths.  $\lambda_{\text{ex}}$  is preferably chosen to generate the maximum fluorescence intensity (Figure 2).

The excitation spectrum is generated by the reverse process, *i.e.* the wavelength of observation  $\lambda_{\text{em}}$  is kept constant while the wavelength of excitation is varied to cover the entire absorbance spectrum of the molecule studied. The intensity measured in the excitation spectrum is thus reflecting the amount of photons that are absorbed with the wavelength. If the species that absorbs the light is also the one responsible for the fluorescence, then the excitation spectrum (once corrected for the wavelength dependency of the lamp intensity and transmission efficiency of the excitation monochromator) should be superimposable with the absorption spectrum<sup>25,29-30</sup> (Figure 2). This is the case if *e.g.* only one type of molecule is

present in solution. If, however, several species are populating the ground state, then excitation and absorbance spectra are highly likely to differ<sup>25,29-30</sup>.



**Figure 2: Overlay of the absorption and emission spectra of tyrosine (Tyr) and tryptophan (Trp), obtained in 0.1 M phosphate buffer solution (pH 7) at room temperature. Data from PhotoChemCad 2.1 software<sup>31-32</sup>.**

The Stokes shift represents the difference between the wavelength of maximum emission and the wavelength of maximum absorbance<sup>25,29-30</sup>. This parameter is interesting as it increases with solvent polarity for fluorophores that have a higher dipole moment in the excited rather than the ground state<sup>25</sup>. Trp is an example of a fluorophore that has a very large Stokes shift with increasing solvent polarity (see section 7.15 for more details). This shift directly reflects important changes in the nature of the Trp microenvironment and is thus a very useful probe to look at protein structural changes.

### 1.2.2 The fluorescence quantum yield

The quantum yield of fluorescence ( $\Phi$ ) is defined as the ratio between the quantity of fluorescence photons emitted and the number of photons originally absorbed (Equation 1). The maximum value for  $\Phi$  is thus one and the closer the value is to one, the brighter is the molecule<sup>25,30</sup>.

#### Equation 1

$$\Phi = \frac{\text{number of fluorescence photons emitted}}{\text{number of photons absorbed}}$$

Another way to express the quantum yield is to consider the rate constants from all the processes that can participate in the depopulation of the excited state<sup>25,30</sup>. If one considers  $k_f$  as the fluorescence radiative rate constant, and  $k_{nr}$  as the rate constant associated with all the non-radiative ways to depopulate  $S_1$ , then:

**Equation 2**

$$\Phi = \frac{k_f}{\sum k} = \frac{k_f}{k_f + k_{nr}}$$

The quantum yield is thus highly dependent on the efficiency of the non-radiative processes, in particular, quenching processes, which are competing with the fluorescence. Trp in various proteins was found, for example, to have a quantum yield and lifetime that could decrease by 30-fold<sup>10</sup>, depending on the quenching efficiency of the charge transfer mechanism operating between the indole ring and the local amide carbonyl groups from the peptide backbone<sup>33-34</sup>. A perfect parallel can be drawn between Tyr in proteins, for which equally variable quantum yields and fluorescence lifetimes are observed due to the same quenching process<sup>34-35</sup>. Therefore, both fluorophore quantum yield and lifetime (see below) are good indicators of microenvironment changes and are extremely sensitive to quenching mechanisms.

**1.2.3 The fluorescence lifetime**

The fluorescence lifetime ( $\tau$ ) defines the average time that a fluorophore spends in the excited state prior to emission, which typically ranges from tens of picoseconds to hundreds of nanoseconds. This is the time window available for the fluorophore to interact with its environment in the excited state ( $S_1$ ) before relaxing to the ground state ( $S_0$ ). This is thus the time during which dynamic processes can be observed<sup>25,30</sup>.

Let us consider a dilute solution of fluorophore X of concentration [X]. Upon excitation with a very short light pulse, at time  $t = 0$ , a certain amount of X molecules populating  $S_0$  are promoted to the  $S_1$  state. These excited molecules, denoted  $X^*$ , follow then either a radiative ( $k_f$ ) or a variety of non-radiative ( $k_{nr}$ ) decay routes to return to the ground state. By simplifying the case and ignoring intermolecular interactions (dilute solution), the rate at which  $[X^*]$  depopulates the excited state follows a classical first order kinetic law<sup>25</sup>, and can thus be expressed as:

**Equation 3**

$$-\frac{d[X^*]}{dt} = (k_f + k_{nr})[X^*]$$

Considering that  $[X^*]_0$  was the concentration of excited molecules at  $t = 0$ , then integration of this equation gives:

**Equation 4**

$$[X^*] = [X^*]_0 e^{-t/(k_f + k_{nr})} = [X^*]_0 e^{-t/\tau}$$

The lifetime of the excited state ( $\tau$ ) is thus expressed as:

**Equation 5**

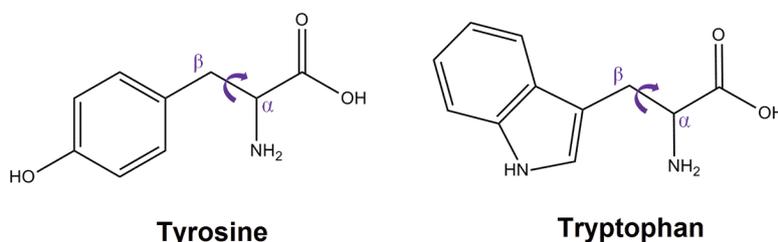
$$\tau = \frac{1}{k_f + k_{nr}}$$

The lifetime ( $\tau$ ) is thus dependant on the fluorescence radiative rate constant ( $k_f$ ), but also on the rate constants of all the other non-radiative routes that depopulate  $S_1$ . Ideally,  $\tau$  should however depend only on  $k_f$  such that  $\tau_f = 1/k_f$ , which is called the radiative lifetime<sup>25</sup>. Using Equation 2, the lifetime ( $\tau$ ) and quantum yield of fluorescence ( $\Phi$ ) can easily be related and the radiative lifetime ( $\tau_f$ ) appears also easily retrievable from these two values:

**Equation 6**  $\Phi = k_f \times \tau = \tau/\tau_f$

For a given radiative rate constant, lifetime and quantum yield are thus directly proportional. This explains why the quenching of Trp/Tyr fluorescence caused by charge transfer to the peptide backbone was characterised by a simultaneous decrease in both quantum yield and lifetime<sup>33-35</sup>. Similarly when Tyr or Trp are in close proximity to any other efficient fluorescence quenchers such as disulfide bridges, the same observations are made<sup>36</sup>. The single Trp residue of the cutinase enzyme is, for example, in contact with one of the disulfide bridges and shows a very short lifetime of 40 ps and a very low quantum yield of 0.002<sup>37</sup>. This is in contrast with other proteins where yields close to 0.35 and lifetimes of a few nanoseconds are obtained<sup>10</sup>.

Most of the time however, the fluorescence decay from a complex macromolecule like a protein is usually a non-mono-exponential decay. Therefore to extract the fluorescence lifetime one has to fit the decay data to a multi-exponential function<sup>25,38</sup>. The first reason is of course because proteins often have multiple chromophores, but also because each chromophore has a unique microenvironment associated with various quenching interactions, different solvent exposure and solvation effects<sup>25,38</sup>. Surprisingly, proteins with a single chromophore also often require to be fitted with multi-exponential fit function. These exponentials are usually associated with a different rotamer form of the chromophore obtained by rotation at the  $C_\alpha$ - $C_\beta$  bond (Figure 3).



**Figure 3: Illustration of the  $C_\alpha$ - $C_\beta$  rotation believed to generate the different rotamers of Trp and Tyr.**

For proteins containing a single Trp residue, bimodal or trimodal decays obtained either in the native or denatured state were attributed to various Trp rotamers<sup>39</sup>. For proteins containing a single Tyr residue, up to four rotamers were used recently to explain the complex fluorescence decay obtained during the aggregation of the  $\beta$ -amyloid (A $\beta$ ) protein<sup>40-41</sup>. This molecule, like insulin, can form under destabilising conditions insoluble amyloid fibrils, leading in the case of A $\beta$  to Alzheimer's disease.

### 1.3 Influence of intermolecular photophysical processes on the emission properties of fluorophores

We saw in the previous section that fluorophores are interesting probes due to the unique sensitivity they show to changes in their microenvironment. Each fluorophore is solvated in a complex medium, and any physical/chemical modification of that medium affects how the surrounding molecules interact with the fluorophore<sup>10,25,28</sup>. This has of course direct consequences on the efficiency of the various processes that compete with the de-excitation of the S<sub>1</sub> state. For example, an increase in temperature decreases the quantum yield/lifetime of Tyr and Trp<sup>42</sup>, because it promotes non-radiative decays related to thermal agitation (*e.g.* collisions with solvent molecules, intramolecular motions<sup>25</sup>). Similarly, a change in pH and/or solvent polarity have important consequences on the fluorophore emission properties, which have been discussed in detail for Tyr in section 5.4.1 (and for Trp in section 7.15).

Detailed below are the main photophysical intermolecular processes occurring in the fluorophore microenvironment that are responsible for these changes and that have major consequences for the emission properties of fluorophores. These mechanisms are always encountered when studying intrinsic protein fluorescence.

#### 1.3.1 Static and dynamic quenching

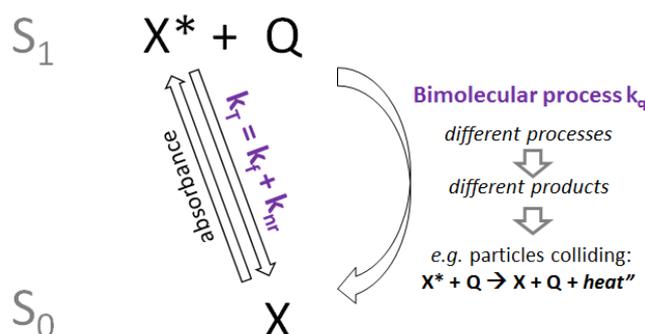
Fluorescence quenching refers to any processes that occur between the fluorophore and its environment, leading to a decrease in the quantum yield of fluorescence. Electron/proton/energy transfers as well as particles colliding or exciplex<sup>c</sup> formation are the main photophysical mechanisms responsible for fluorescence quenching<sup>43</sup>, and all have in common that another particle (called the quencher, Q) is interacting with the fluorophore.

---

<sup>c</sup> Exciplex refers to complexes that are formed in the excited-state<sup>19</sup>.

Quenching is thus the result of intermolecular interactions, which can happen either in the ground state (static quenching) or in the excited-state (dynamic quenching)<sup>43-44</sup>.

Let us consider first the case of dynamic quenching, where the quencher molecules (Q) interact with the molecules (X\*) during their excited state (Figure 4).



**Figure 4:** Scheme illustrating the principle of dynamic quenching, characterised by the appearance of a bimolecular decay rate from the first excited state. Various processes such as electron/proton/energy transfers or particles colliding can cause this bimolecular decay.

If we name  $k_T$  the sum of the radiative/non-radiative rate constants that can depopulate  $S_1$  in the absence of quenchers ( $k_T = k_f + k_{nr}$ ), and name  $k_q$  the rate constant for the bimolecular process between  $X^*$  and  $Q$  (Figure 4), Equation 3 can be modified so that the depopulation rate of  $S_1$  accounts for the intermolecular interaction:

**Equation 7** 
$$-\frac{d[X^*]}{dt} = (k_T + k_q[Q])[X^*]$$

Integrating this equation gives of course a similar answer as Equation 4, except that the lifetime of the excited state ( $\tau$ ) is now equal to:

**Equation 8** 
$$\tau = \frac{1}{k_T + k_q[Q]}$$

If we call  $\tau_0$  the lifetime of the excited state obtained in the absence of a quencher (see Equation 5), then  $\tau_0 = 1/k_T$  and Equation 8 can be rearranged so that:

**Equation 9** 
$$\frac{\tau_0}{\tau} = 1 + k_q[Q]\tau_0$$

The quantum yield of fluorescence, as defined in Equation 6, is equal to the product of the excited state lifetime and the fluorescence radiative rate constant of  $X^*$  ( $k_f$ ). However as shown in Figure 4,  $k_f$  is  $Q$  independent which means that:

**Equation 10** 
$$\frac{\Phi_0}{\Phi} = \frac{I_0}{I} = 1 + k_q[Q]\tau_0 = 1 + K_{SV}[Q]$$

Where  $\Phi_0$  and  $I_0$  are the quantum yield and fluorescence intensity obtained in the absence of the quencher, while  $\Phi$  and  $I$  are their equivalents in the presence of the quencher. This final equation is called the Stern-Volmer equation, and the constant  $K_{sv}$  the Stern-Volmer constant<sup>43-44</sup>. By plotting  $I_0/I$  versus  $[Q]$ , a linear plot is usually obtained for which the slope is equal to  $K_{sv}$ . Depending on the value of this constant, one can make a judgment about the accessibility of the fluorophore to the quencher. Typically, Trp or Tyr free in solution have a high  $K_{sv}$  constant ( $\sim 8 \text{ M}^{-1}$ ) because they are completely accessible to the quencher whereas a fluorophore buried inside a protein has a low  $K_{sv}$  value ( $0.33 \text{ M}^{-1}$  for the three Trp residues in ovalbumin)<sup>45</sup>.

This tool is thus particularly interesting when applied to protein analysis as it gives quantitative information about the degree of exposure of the fluorophores to the environment<sup>43-44</sup>. In fact, by using a variety of quenchers which differ by their size or charge (*e.g.* iodide, oxygen and acrylamide), a selective quenching operation, targeting only certain fluorophores in the protein sequence, can be implemented. This type of experiment enables the differentiation of exposed and buried residues and enables the monitoring of protein conformational changes<sup>44-46</sup>. In that case, instead of being linear, the Stern-Volmer plot shows different slopes associated with different  $K_{sv}$  values. Möller *et al.*<sup>45</sup> obtained for example a biphasic plot during the quenching of bovine serum albumin (BSA) with iodide, reflecting the different accessibility of both Trp residues present in BSA.

Static quenching refers to the formation of a dark complex<sup>30,43-44,46</sup> in the ground state between the fluorophore  $X$  and the quencher  $Q$  such as  $X + Q \leftrightarrow XQ$ . This equilibrium is characterised by the constant of association  $K_a$ :

**Equation 11**  $K_a = \frac{[XQ]}{[X][Q]}$

Depending on the value of  $K_a$ , at any time  $t$  there is a portion of fluorophores  $[X]$  that remains unbound and fluoresces as if no quencher was present. Using Equation 11 and the mass conservation principle stipulating that the initial quantity of molecules  $[X]_0$  is equal at any time  $t$  to  $[X] + [XQ]$ , then  $[X]$  can be defined as:

**Equation 12**  $[X] = \frac{[X]_0}{1+K_a[Q]}$

Assuming that the fluorescence intensity is proportional to the concentration (dilute conditions) of the unbound fluorophore  $[X]$ , Equation 12 can be rearranged such as:

**Equation 13**

$$\frac{I_0}{I} = 1 + K_a[Q]$$

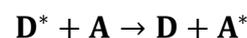
The final equation obtained is thus highly similar to the Stern-Volmer equation (see Equation 10): a linear relationship is obtained between the intensity ratio and the concentration of quencher [Q]. The main difference however is that Equation 13 does not depend upon the lifetime of the excited state. Static quenching leads thus to a decrease in the quantum yield but does not affect the lifetime of the excited state<sup>30,43-46</sup>, which makes sense considering that the formation of a dark complex only decreases the number (and not the properties) of molecules X that reach S<sub>1</sub>. Said differently, “static” means that both X and Q are at a fixed distance and their interaction does not depend on the time window ( $\tau$ ) normally necessary for molecules to diffuse and interact. Plotting fluorescence lifetime versus quencher concentration offers a way to distinguish between static and dynamic quenching<sup>30,43-46</sup>.

The molecular contact in dynamic quenching requires the quencher to diffuse to the fluorophore during its lifetime. The process is thus distance dependant, and the quencher can diffuse further if the lifetime is longer. Increasing the temperature also results in a higher dynamic rate constant, as the quencher diffuses faster causing more collisions<sup>43-44</sup>. On the other hand, a higher temperature decreases the static rate constant as it leads to the dissociation of the weak dark complexes formed. This different dependency to the temperature is thus another way to differentiate these two types of quenching.

We saw that quenchers could be used to probe the accessibility of chromophores in proteins. Some studies used it to look at protein aggregation: Kayser *et al.* measured a decrease in K<sub>sv</sub> during the pH-triggered self-assembly of  $\beta$ -sheet ribbons, interpreted as Trp residues being buried (*i.e.* less accessible) upon aggregation<sup>47</sup>. Dusa *et al.* also observed that the K<sub>sv</sub> decreased during the fibrillation of  $\alpha$ -synuclein, suggesting the burial of Trp39 during the aggregation<sup>48</sup>. This protein, like insulin, can form amyloid fibrils under destabilising conditions that lead in the case of  $\alpha$ -synuclein to Parkinson’s disease. Specific cases of Stern-Volmer plots applied to the study of insulin denaturation and aggregation can be found in section 5.4.2, where the intrinsic quenching mechanisms of Tyr fluorescence in proteins are also discussed.

### 1.3.2 Förster Resonance Energy Transfer (FRET)

FRET is the non-radiative transfer of energy between an excited donor (D\*) and an acceptor (A) molecule, such as:

**Equation 14**

More precisely, if the donor and acceptor molecules are close enough, their dipoles can couple, which permits the energy to flow from one molecule to another by resonance<sup>49-50</sup>. Looking at Equation 14, FRET can be seen as a competing pathway to fluorescence. FRET is therefore a form of dynamic quenching<sup>43,51</sup> and considering  $k_D^0$  the fluorescence rate constant of  $D^*$  in the absence of acceptor and  $k_{ET}$  the rate of energy transfer, the FRET efficiency ( $E$ ) is simply defined as:

**Equation 15**

$$E = \frac{k_{ET}}{k_D^0 + k_{ET}}$$

The Förster theory established that the rate of energy transfer ( $k_{ET}$ ) can be expressed as<sup>43</sup>:

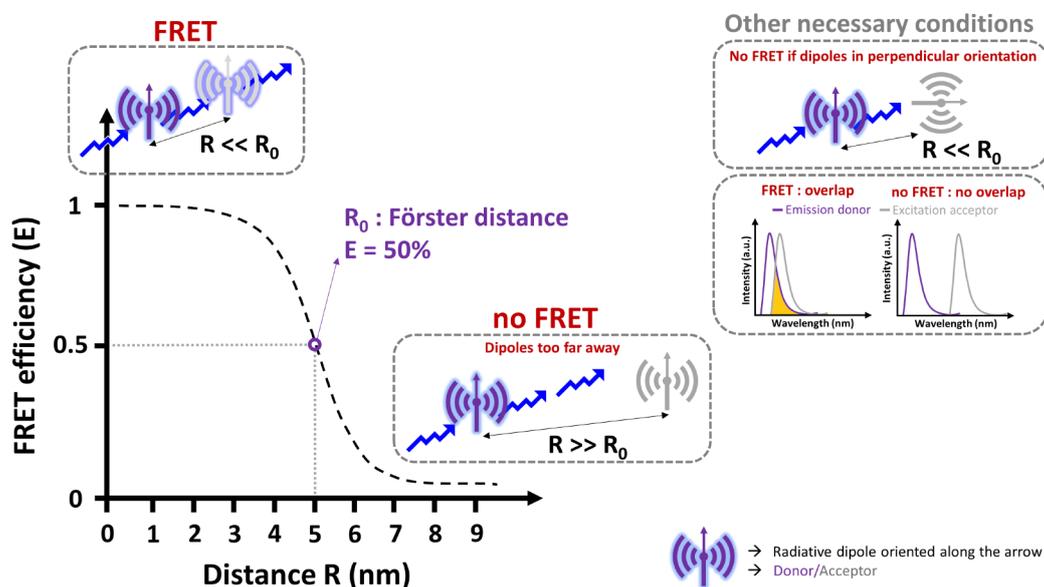
**Equation 16**

$$k_{ET} = k_D^0 \left(\frac{R_0}{R}\right)^6 = \frac{1}{\tau_D^0} \left(\frac{R_0}{R}\right)^6$$

Where  $\tau_D^0$  is the donor lifetime in the absence of acceptor,  $R$  is the distance between  $D^*$  and  $A$ , and  $R_0$  is the Förster distance. By combining these two latter equations, we obtain a new expression for the FRET efficiency ( $E$ ):

**Equation 17**

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



**Figure 5: Schematic representation of the FRET process that can happen between two dipoles (represented as antennas). The FRET efficiency highly depends on the distance between fluorophores, but also on the angle between dipoles and the extent of overlap (shown in orange) between the donor emission and acceptor excitation spectra.**

$E$  and  $k_{ET}$  are thus strongly dependent on the distance ( $R$ ) separating the donor/acceptor pair<sup>52-53</sup> (Figure 5), and rapidly decrease with increasing  $R$  as both are proportional to the

inverse of its 6<sup>th</sup> power. A short distance (typically  $\leq \sim 20$  nm for Tyr/Trp) is then necessary to enable coupling of the oscillating dipole moments.

FRET depends on several other factors (Figure 5) such as the extent of spectral overlap between donor emission and acceptor absorbance necessary to have matching dipole frequencies (*i.e.* resonance) and the degree of alignment of these dipoles<sup>49-50</sup>. These dependencies are contained in the Förster distance ( $R_0$ ), which corresponds to the distance for which the FRET efficiency ( $E$ ) of a D-A pair is equal to 50%.  $R_0$  can be calculated from the spectroscopic and dipole orientation parameters of the donor and acceptor such that<sup>43,50</sup>:

**Equation 18** 
$$R_0 = [2.8 \times 10^{17} \times k^2 \times \Phi_D^0 \times \epsilon_A \times J(\lambda)]^{1/6}$$

Where  $\epsilon_A$  is the maximum extinction coefficient of the acceptor,  $\Phi_D^0$  is the quantum yield of the donor in the absence of acceptor,  $k^2$  the dipole orientation factor, and  $J(\lambda)$  the integral of the spectral overlap between the normalised donor emission ( $F_D$ ) and the acceptor excitation ( $E_A$ ) spectra:

**Equation 19** 
$$J(\lambda) = \int E_A(\lambda) F_D(\lambda) \lambda^4 d\lambda$$

The FRET efficiency is thus higher for fluorophores having large  $J(\lambda)$ , *i.e.* large spectral overlap. The  $k^2$  coefficient is a bit more complex to determine and depends on the angle between the donor/acceptor dipoles, relating the FRET efficiency to their degree of alignment. As shown in Figure 5, the FRET coupling is maximal if the dipoles are parallel to each other but is negligible for orthogonal orientations<sup>43,50</sup>. The  $k^2$  coefficient is however very difficult to estimate as in almost every realistic case, the fluorophores never have a fixed orientation. Instead they rotate during the excited state lifetime, leading to dipoles having isotropic orientations. For this reason, most studies consider  $k^2$  to be equal to 2/3, which is the average value obtained by integrating over all possible dipole angles<sup>43,50</sup>.

If the  $R_0$  value is known for a given D-A pair, the measurement of the FRET efficiency allows one to use Equation 17 to calculate the distance,  $R$ , separating donor and acceptor. The distance range can vary between 10 and 100 Å depending on the spectroscopic characteristics of the donor and acceptor. For that reason, FRET is a very powerful method to study proteins and their assembly and is often referred to as a spectroscopic ruler<sup>49,54</sup>.

For proteins labelled with a suitable donor-acceptor pair, FRET can be used for example, to identify important conformational changes and aggregation processes induced by various stress factors. Teoh *et al.*<sup>55</sup> examined the effect of shear-flow on amyloid fibril formation by

human apolipoprotein C-II (apoC-II). This protein is essential in the body to activate the enzyme that hydrolyses triglycerides, providing free fatty acids for cells. Like insulin, it is a small peptide (79 residues) that can form amyloid fibrils. This molecule also has the particularity to have a single Trp residue. The authors used this Trp as the FRET donor and covalently attached the acceptor 1,5-I-AEDANS on apoC-II cysteine mutants to complete the FRET pair ( $R_0 = 22 \text{ \AA}$ ). Once subjected to shear-flow, the FRET measurements show an increase in the distance between the fluorophores (+2  $\text{\AA}$ ), indicative of an irreversible misfolding in the monomer species preceding the appearance of fibrils.

Similarly, the early aggregation steps in the  $\alpha$ -synuclein fibrillation were followed by FRET<sup>56</sup>, by labelling the protein's cysteine mutant with Alexa-488 (donor) and tetra-methyl rhodamine (TMRh) (acceptor). The FRET signal, believed to be intra and not intermolecular, appeared only upon oligomer formation which, according to the authors, meant that a conformational change must be associated with oligomerization. To prove it, they destabilized the aggregates in alkaline pH and observed a recovery of the donor fluorescence that was concentration dependent. A specific FRET study applied to the oligomerization of labelled insulin can be found in section 5.4.2.

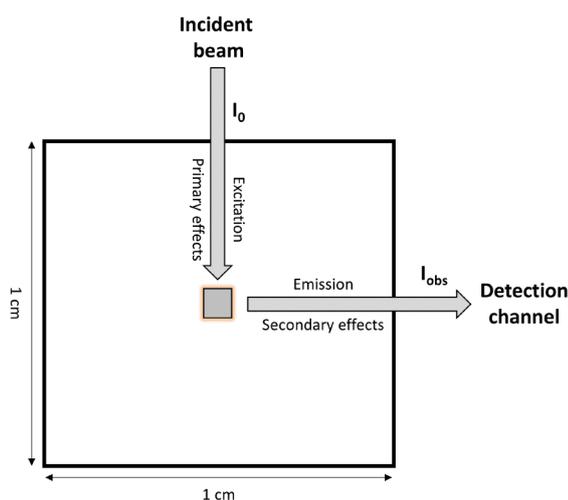
However FRET studies from proteins labelled with external fluorophores are not ideal, first because mutants are often used instead of the wild protein, but also because the attachment of external molecules and the steps necessary for their attachment may disturb the protein native state<sup>20</sup> (see section 5.3). Instead it may be possible to use the fluorophores already present in the protein sequence as FRET pairs. The only issue (see chapter 5) is that in the absence of multivariate data analysis, the presence of strong spectral overlaps limits the use of internal FRET to certain protein cases.

Human interferon- $\gamma$  is an interesting example and involved the Tyr-Trp pair ( $R_0 = 10\text{--}15 \text{ \AA}$ ), probably the most commonly observed intrinsic FRET pair<sup>28</sup>. This protein is produced by activated lymphocytes and has antiviral, antitumor, and immunoregulatory properties. It is biologically active as a dimer and formed non-covalently by two monomers, each containing one Trp and four Tyr. By increasing the temperature, the dimer dissociated and the authors observed a decrease in the quantum yield associated with a loss in the Tyr-to-Trp FRET efficiency<sup>57</sup>. Concomitantly the relative Tyr intensity increased and went up from being equal to 20% of Trp fluorescence in the dimer to 50% in the monomer, as less energy is transferred from Tyr to Trp<sup>57</sup>.

FRET is thus highly sensitive to any process (*e.g.* unfolding, association, etc.) that significantly affects the distances between protein's fluorophores.

### 1.3.3 Inner filter effects (IFE)

Under optically dilute conditions ( $A < 0.1$ )<sup>58</sup>, the fluorescence intensity is a linear function of the fluorophore concentration. With the usual right-angle spectrofluorometers<sup>29</sup> (as used in this thesis), this fluorescence intensity emanates from the portion of the sample that is illuminated at the centre of the cuvette (Figure 6). This implies that before and after exciting the sample, the light beam must travel through a certain volume of solution, which automatically generates possibilities for the light to be attenuated. Depending on the pathlength and the density of molecules that the light must cross, the excitation and/or emission radiation can be absorbed, resulting in an apparent decrease in the quantum yield and a nonlinear relationship between observed intensity and fluorophore concentration<sup>29,58</sup>. This is called the inner filter effect (IFE).



**Figure 6: Illustration of the primary and secondary inner filter effects affecting the fluorescence of a sample illuminated in the centre of a 1cm×1cm cuvette and observed with a right-angle geometry.**

There are in fact two types of IFEs in the right-angle geometry. The first one, called primary IFE corresponds to the absorption of the excitation light. The second one called secondary IFE, corresponds to the re-absorption of the fluorescence emission (Figure 6)<sup>29,58</sup>. This occurs if the absorption and emission band of the fluorophore are partially overlapping (Stoke's shift dependant), generating spectral shapes that are distorted in the overlap region<sup>29,58</sup>. IFE are however not only caused by the fluorophore of interest, but can be generated by other molecules (fluorescent or not) as long as they absorb in the same wavelength range as the

fluorophore<sup>59</sup>. IFE can be difficult to correct for. For this reason, most authors recommend, if possible, avoiding IFEs by working in dilute conditions<sup>29,58</sup>.

Alternatively Lakowicz<sup>58</sup> proposed a formula (Equation 20) that can approximately correct for the presence of both types of IFEs at each data point. Supposing an optical density  $A_{\lambda_{ex}}$  and  $A_{\lambda_{em}}$  at the wavelength pair  $(\lambda_{ex}/\lambda_{em})$ , the author considered that for a standard 1×1 cuvette, the optical density attenuates the light by a factor of  $10^{-0.5A_{\lambda_{ex}}}$  and  $10^{-0.5A_{\lambda_{em}}}$  respectively. Using an absorbance-based approach (ABA) and the collection of a separate absorbance measurement, the author proposed to compensate mathematically this attenuation by building a correction matrix ( $\mathbf{M}$ ), at each  $\lambda_{ex}/\lambda_{em}$  point, which transforms the observed intensity ( $I^{obs}$ ):

**Equation 20** 
$$I_{\lambda_{ex},\lambda_{em}}^{corr} = I_{\lambda_{ex},\lambda_{em}}^{obs} \times \mathbf{M} = I_{\lambda_{ex},\lambda_{em}}^{obs} \times 10^{(0.5 \times (A_{\lambda_{ex}} + A_{\lambda_{em}}))}$$

The ABA is probably the most widely used approach to correct for IFEs, even if it has several drawbacks such as being dependent on absorbance measurements that are less sensitive and bringing an additional error source to the fluorescence data<sup>60-61</sup>. Also, without considering the determination of the geometrical parameters proper to the excitation/emission beam<sup>62</sup>, ABA seems limited to total absorbance values ( $= A_{\lambda_{ex}} + A_{\lambda_{em}}$ ) that do not exceed 1.5<sup>60-61</sup>. Other methods have been proposed to correct for IFE such as the controlled dilution approach, which consists of using a diluted version of the sample to build the correction matrix, avoiding the need for UV-Vis data<sup>60</sup>. A recent review summarized all the methods available to correct for IFEs, discussing their advantages and inconveniences<sup>63</sup>.

If possible, the safer option is still to avoid IFE instead of having to correct them mathematically. But since working in dilute solutions might not always be feasible, one simple/efficient way to attenuate IFE is to reduce the cuvette pathlength<sup>58,29</sup>. Using a shorter excitation pathlength is for example very efficient in reducing the strong primary IFE<sup>64</sup>. If IFE still remains after reducing the pathlength, then Equation 20 can be used to do the correction by changing the length parameters according to the cuvette used. Other alternatives involving different geometrical arrangements such as front face illumination<sup>65</sup> or off-centre illumination<sup>58</sup> were also proposed to avoid IFEs, but suffer from other drawbacks (light reflection, lack of reproducibility). Finally, some studies explored a completely different approach and proved that chemometric methods can directly handle data affected by IFEs and still perform well in *e.g.* predicting analyte concentration<sup>66-67</sup>. This point is interesting and acknowledges that for real samples in the real world, IFEs might not always be avoidable.

## 2. Materials and Methods

The goal in this short chapter is to describe the reagents, the sample preparation as well as the various instrumentation and parameters used to collect all the data presented in this thesis. The part specific to the data analysis was kept separate and can be found in the respective chapters, as the method evolved during the study.

### 2.1 Reagents

All reagents were used without further purification (Table 1) and purchased as analytical grade where possible. The water used to make all solutions was of high purity (Chromasolv™ Plus, HPLC grade) and did not show any impurities that were detectable during the fluorescence measurements.

**Table 1: Description of all the reagents used in the thesis.**

Substance name	Description	CAS number	Company
Acetic acid	Fluka TraceSELECT™ ≥99%	64-19-7	Sigma-Aldrich
Ammonium formate	for HPLC, ≥99.0%	540-69-2	Sigma-Aldrich
Erythrosin B	Analytical standard	16423-68-0	Sigma-Aldrich
Ethanol	HPLC grade	64-17-5	Sigma-Aldrich
Formic acid	Fluka TraceSELECT™ ≥88%	64-18-6	Sigma-Aldrich
Glycerol	Spectrophotometric grade ≥ 99.5%)	56-81-5	ACROS Organic
HEPES	≥99.5% (titration)	7365-45-9	Sigma-Aldrich
Human insulin (I0908)	- meets USP specifications - expressed in <i>Saccharomyces cerevisiae</i>	11061-68-0	Sigma-Aldrich
Hydrochloric acid	Fluka TraceSELECT™ ≥30%	7647-01-0	Sigma-Aldrich
Isopropanol	Chromasolv™ for HPLC 99.9 %	67-63-0	Sigma-Aldrich
L-Tyrosine	Reagent Grade 98%	60-18-4	Sigma-Aldrich
Nitric acid	SG 1.42 (70 %)	7697-37-2	Fisher-Scientific
Perylene	Analytical standard	198-55-0	Sigma-Aldrich
Phloxine B	Antibacterial fluorescent dye	18472-87-2	Sigma-Aldrich
Sodium chloride	Fluka TraceSELECT™	7647-14-5	Sigma-Aldrich

Sodium hydroxide	BioXtra, $\geq 98\%$ (acidimetric), pellets (anhydrous)	1310-73-2	Sigma-Aldrich
Trizma® base	Fluka TraceSELECT™ inorganic trace analysis, $\geq 99.9995\%$	77-86-1	Sigma-Aldrich
Trizma® HCl	Reagent grade, $\geq 99.0\%$ (titration), crystalline	1185-53-1	Sigma-Aldrich

## 2.2 General material and instrumentation

Several precautions were taken to make protein and amino acid solutions. To prevent contamination, the water used to make the solutions as well as the pipette tips were sterilized using an Omega Media autoclave (Prestige medical, part no. 220140). The solutions were prepared in a laminar flow hood and sterile filtered using 5 mL syringes (BD Plastik™) equipped with Minisart filters (Sartorius, 0.2  $\mu\text{m}$  pore size). All the other solutions that involved small organic dyes were done on the bench, in a standard manner.

For the analysis, all samples were pipetted into 0.4cm  $\times$  1cm quartz cuvettes (Lightpath Optical, UK). These cuvettes were preferred to the standard 1cm  $\times$  1cm variety because, as explained in section 1.3.3, a shorter excitation pathlength permits the minimization of IFE. The other advantage is that a smaller volume of sample is required to fill up the cuvette (1 mL versus 3 mL for the 1cm  $\times$  1cm format).

To avoid photodegradation, the samples out on the bench were always kept in the dark. The other samples, waiting to be analysed, were kept in a 2-8°C fridge, trying to minimize especially for proteins any sort of physical stress prior to analysis. Straight after the measurements, the cuvettes were cleaned using distilled water first, before being rinsed several times with high quality water, isopropanol and ethanol. Once every couple of weeks, the cuvettes were immersed for few hours in a 30% nitric acid solution (v/v) and a 2 M sodium hydroxide solution to remove or prevent any potential stains on the cuvettes. The protein samples were stored in a -70°C freezer while the samples less prone to degradation were kept in a -20°C freezer.

UV-visible absorbance measurements were done on a Cary 60 UV-Vis spectrophotometer (Agilent Technologies, part no. G6860A). Temperature control of the sample was carried out using a Single Cell Peltier Accessory (part no. SPV 1X0).

The fluorescence measurements were performed on a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, part no. G9800A), equipped with a multicell holder (Agilent Technologies, part no. G9808A) combined with a temperature controller (Agilent Technologies, part no. G9844A). In this spectrophotometer, the polarization of excitation and emission light is done by thin film polarizers (TFP) mounted either on a Manual Polarizer Accessory (part no. 00-100761-00) or on Automated Polarizer Accessory (part no. 8510187100). Due to the difficulty of measuring multidimensional data with the automated assembly, only the manual accessory was used in the thesis. Ultimately these filters were replaced by dual wire grid polarizers (dWGP), more suited to the analysis of the intrinsic fluorescence of proteins, as explained in detail in chapter 3.

Before collecting any fluorescence data, the spectrometer was always validated according to an in-house protocol. Using the “validate” software delivered with the machine, several parameters were checked such as:

- The accuracy and reproducibility of excitation and emission wavelengths (Xenon lamp).
- The accuracy of spectral bandwidth for various excitation and emission slits (Xenon lamp).
- The 0% transmittance at the Rayleigh and Raman band with closed emission shutter (using a certified Starna Raman water standard, part no. 3Q-10-Water).
- The absence of stray light using ground silica diffuser.
- The Raman water band sensitivity at two excitations (350 and 500 nm), ensuring that the signal-to-noise ratio (SNR) was higher than a defined threshold.

Furthermore, considering that all the measurements in this thesis were temperature controlled, it was also important at the start to check that the temperatures delivered by the multicell holder were reliable.

The test was simple and consisted of monitoring the temperature of a cuvette filled with water using an independent temperature probe (Agilent Technologies, part no. 10040800) and comparing the results to the temperatures claimed by the software. Covering the range from 20 to 70°C with a 10°C increment, the conclusion was that the temperature controller worked well. It reached the target temperature in less than a minute while it took a bit more than five minutes for the cuvette to thermally stabilise to the set temperature, reaching values within 0.5°C of the set value. In conclusion, any sample measured on this spectrometer in an aqueous matrix was

left between 5 and 10 minutes in the cell holder to equilibrate prior to making the spectral measurement.

## 2.3 Anisotropy standards solutions

### 2.3.1 Stock solutions for filter validation

All samples described in this section were prepared in triplicate and measured on the day of their preparation (fresh). The solutions made in glycerol (highly viscous) were left sufficiently longer (~ 2h) in the multicell holder to settle and equilibrate to the right temperature. This also permitted the removal of potential air bubbles trapped in the viscous matrix. After analysis, the samples were stored in a -20°C freezer.

In order to obtain a 1  $\mu\text{M}$  stock solution of perylene in 100% glycerol, ~ 0.006 g of perylene were weighed and mixed with 25 mL of glycerol. Due to the low solubility of perylene in glycerol<sup>d</sup>, the volumetric flask was left over an hour in the sonicator (Branson, model 1510) and the remaining non-dissolved particles were removed by centrifugation at 13 000 rpm for 15 minutes (Thermo Scientific, model Heraeus Fresco 17). At that point an absorbance spectrum was collected, indicating that the solution concentration was 6  $\mu\text{M}$ . By doing a simple dilution (mixing 5 mL of this solution with 25 mL of glycerol) the 1  $\mu\text{M}$  stock solution was obtained.

This solution was used in chapter 3 to measure the accuracy of various sets of polarizers. The samples were measured at various temperatures: 10, 15, 20 and 25°C.

To further validate the accuracy of the new filters developed, xanthene dye solutions that emit in the longer wavelength region were used. We prepared a 6  $\mu\text{M}$  solution of Erythrosin B (EB) and a 0.6  $\mu\text{M}$  solution of Phloxine B (PB) in 50 mM HEPES buffer (pH 7.2), that were measured at 25 °C.

The buffer was obtained by simply dissolving 2.383 g of HEPES with high purity water in a 200 mL flask, the pH being adjusted to 7.2 using small aliquots of 2 M NaOH solution. Two stock solutions ( $10^{-4}$  M) were first prepared for EB and PB by dissolving respectively 0.0022 and 0.0021 g in 25 mL flasks, filled to the mark with the HEPES buffer freshly prepared. The final 6 and 0.6  $\mu\text{M}$  solutions were obtained by pipetting respectively 1.5 mL and 0.15 mL from the EB and PB stock solutions into new 25 mL flasks, filled also with the HEPES buffer.

---

<sup>d</sup> Kalman *et al.*<sup>68</sup> observed also that perylene dissolves in glycerol at micromolar concentrations.

Due to the aqueous nature of this buffer, only a few minutes were necessary for the samples to equilibrate to the desired temperature prior to measurement.

### 2.3.2 Fluorophore mixtures for chemometric modelling

Two binary mixtures of non-interacting fluorophores were used in chapter 4 to investigate the best measurement/chemometric strategies to be used for ARMES.

For these mixtures, each sample was prepared in triplicate and measured at 25°C. Because the number of samples that had to be analysed was too large to be measured fresh, all the samples were stored first in -20°C freezer. Samples were removed in batches for analysis and placed in a fridge (2-8°C) overnight to thaw. By doing so, we ensured reproducible preparation and sample handling. After analysis, the samples were placed back in the freezer.

The first standard sample set was composed of various perylene concentrations in 100% glycerol. In the wavelength range used for the data collection, a fluorescent impurity from the glycerol was detected, which represented the second species of the mixture. The 1  $\mu\text{M}$  stock solution of perylene was used as the highest solution concentration and diluted to obtain 5 mL of 0.8, 0.5, and 0.2  $\mu\text{M}$  solutions. The volumes pipetted are shown in Table 2 below.

**Table 2: Preparation of the first fluorophore mixture, the perylene dataset.**

<b>Perylene in 100% glycerol (<math>\mu\text{M}</math>)</b>	<b>1 <math>\mu\text{M}</math> stock solution (mL pipetted)</b>	<b>Glycerol (mL pipetted)</b>	<b>Final volume (mL)</b>
1 $\mu\text{M}$	5	0.0	5
0.8 $\mu\text{M}$	4	1.0	5
0.5 $\mu\text{M}$	2.5	2.5	5
0.2 $\mu\text{M}$	1	4.0	5

In order to remove potential air bubbles but also give enough time for the solution to settle and equilibrate to 25°C, the samples were left ~ 2h hours in the multicell holder prior to measurement, as was done for the analysis of the 1  $\mu\text{M}$  perylene solution.

The second standard sample set was built by combining in various ratios the Erythrosin B (~ 6  $\mu\text{M}$ ) and Phloxine B (~ 0.6  $\mu\text{M}$ ) stock solutions. The volumes were pipetted according to Table 3 so that 5 mL of solution was obtained for each fluorophore mixture concentration.

**Table 3: Preparation of the second fluorophore mixture, the xanthene dyes dataset. There was no particular reason for choosing a closed system ( $\Sigma C_i = 100\%$ ), except that it was a worst-case example for MCR.**

Ratio PB/EB (v/v, in %)	Phloxine B (0.6 $\mu\text{M}$ ) (mL pipetted)	Erythrosin B (6 $\mu\text{M}$ ) (mL pipetted)	Final volume (mL)
100 / 0	5	0	5
80 / 20	4	1	5
60 / 40	3	2	5
40 / 60	2	3	5
20 / 80	1	4	5
0 / 100	0	5	5

Due to the aqueous nature of the buffer used, these samples only needed a few minutes to equilibrate to 25°C.

### 2.3.3 Parameters for the data collection

UV-Vis absorbance spectra (200–800 nm) were collected along the 0.4 cm pathlength, using a 2 nm step at a scan rate of 1200 nm min<sup>-1</sup>. Fluorescence data were also collected using a 2 nm step at a scan rate of 1200 nm min<sup>-1</sup>, for which the rationale is explained in section 7.3 (Appendices). The short cuvette pathlength was used along the excitation.

For the analysis of the stock solutions, the Eclipse fluorescence spectrophotometer was fitted with various polarizer configurations (see chapter 3) whereas the fluorophore mixtures were only analysed with dWGP. For perylene, EEM-ARMES data were collected for a  $\lambda_{\text{ex}}/\lambda_{\text{em}}$  range of 230–470/420–530 nm with 5 and 10 nm excitation/emission slit widths. TSFS-ARMES measurements were collected using  $\lambda_{\text{ex}} = 230\text{--}470$  nm and  $\Delta\lambda$  intervals (offset) of 20–210 nm. EEM-ARMES data of the xanthene dyes were collected using 10/10 nm slit widths and  $\lambda_{\text{ex}}/\lambda_{\text{em}}$  range of 470–570/500–650 nm. The photomultiplier tube (PMT) detector voltage was fixed at 600V in all cases.

For each sample, four different polarized spectra were collected: vertical-vertical (VV), vertical-horizontal (VH), horizontal-vertical (HV) and horizontal-horizontal (HH). The anisotropy ( $r$ ) at each emission wavelength was calculated using the anisotropy formula (see section 3.1, Equation 25) and used to construct the 2-D anisotropy plots or the multidimensional anisotropy (*aniso*-MDF) plots. Spectra were not corrected for instrument response.

## 2.4 Insulin study

In this section, all sample solution preparation was carried out in a laminar flow hood using aseptic techniques to minimize contamination. The water and pipette tips were sterilized and the solutions were sterile filtered as explained in section 2.2.

### 2.4.1 Human insulin solutions

Solutions containing mainly monomers, dimers, tetramers or hexamers were created using the conditions described in Table 4, conditions that were described in other studies interested also in the various oligomers of insulin<sup>17,69-73</sup>.

**Table 4: Description of the buffers used to generate solutions of human insulin containing mainly monomers (MO), dimers (DI), tetramers (TE) and hexamers (HX).**

Association state	Buffers	pH
Monomer (MO)	20% acetic acid, 0.1 M NaCl	2.0
Dimer (DI)	25 mM HCl, 0.1 M NaCl	1.6
Tetramer (TE)	20 mM formate, 0.1 M NaCl	3.0
Hexamer (HX)	20 mM Tris, 0.1 M NaCl	7.4

With an isoelectric point of 5.3<sup>74</sup>, native wild-type insulin has a low solubility in neutral pH<sup>e</sup> but can be solubilised up to 10-20 mg/mL in dilute acidic conditions (pH ~ 2/3)<sup>74</sup>. For experiments starting with monomeric, dimeric, or tetrameric forms, insulin was thus directly dissolved in the appropriate buffer (Table 4).

In order to form hexamers in solution, the protocol was a bit more complex. First, the insulin required a zinc content that was appropriate to enable dimers to associate and form hexamers (see section 5.2 for further details). According to the certificate of analysis supplied by Sigma Aldrich (see section 7.16 in Appendices), the batch number SLBP5090V used to make all solutions contained 0.4% of zinc (w/w of insulin), which was in the ideal range to form hexamers. In fact this content matched other studies which estimated that it corresponded approximately to two Zn<sup>2+</sup> cations per insulin hexamer<sup>17,72,75</sup>.

Due to the low solubility of insulin at neutral pH, hexamers could not be formed directly in the 20 mM Tris buffer. Instead the insulin powder had to be dissolved in a small volume (1 mL) of 25 mM HCl buffer (pH 1.6). Aliquots of 25 mM Tris buffer were then added until

<sup>e</sup> <https://www.sigmaaldrich.com/catalog/product/sigma/i0908>

reaching a final composition of 20 mM Tris in the insulin flask. The pH was then adjusted to 7.4 with 2 M NaOH.

All the buffers were prepared in 100 mL volumetric flasks and their pH values were carefully controlled or adjusted to within 0.05 units of the desired value, using a calibrated pH meter (Eutech instruments, model Cyberscan pH 10). For the duration of the data collection, these buffers were kept sealed in a 2-8°C fridge.

In order to avoid changes in the sample composition and a potential re-distribution of oligomer populations, all the insulin solutions had to be made and analysed while fresh, as done by all the other studies that used the same experimental conditions<sup>17,69-73</sup>. At the beginning of each day of analysis, ~10 mg of insulin were weighed and dissolved in the appropriate buffer in a 5 mL volumetric flask. To ensure complete dissolution and homogeneity, the volumetric flasks were slowly inverted 10 times and left to settle for ~ 20 min in the dark at room temperature. This stock solution was then sterile filtered (0.2 µm) and the concentration of insulin (2 mg/mL) was measured by UV-Vis using an extinction coefficient of 1.0 for 1 mg/mL at 276 nm<sup>76</sup>. From this solution, three sample replicates at 2 mg/mL were obtained after filling three cuvettes with ~ 1.2 mL of insulin solution. These samples were left a few minutes in the multicell holder to equilibrate at 25°C prior to collecting the fluorescence data. Meanwhile, some of the stock solution that was left was diluted to obtain three other replicates at 0.2 mg/mL, that were also analysed while fresh. Directly after analysis, the samples were stored in the -70°C freezer.

In total, the data collection took several weeks and 12 randomised/independent (four series of three) replicates per insulin form and concentration (2 and 0.2 mg/mL) were measured using ARMES, normal EEM and UV-Vis spectroscopy.

#### **2.4.2 L-Tyrosine solutions in “insulin buffers”**

In order to investigate the influence of the various insulin buffers (Table 4) on tyrosine emission properties, solutions of L-tyrosine were prepared in a similar manner as for insulin samples. In particular, the aim was to see if the various buffers were causing changes in the shape of the L-Tyrosine ARMES data that would compromise the PARAFAC modelling of the insulin fluorescence data.

To avoid complications and potential band shape changes due to IFEs, these solutions therefore had to have equivalent concentration (or absorbance) values as the most diluted

insulin samples (0.2 mg/mL) that were used to build multivariate models. This corresponded to absorbance values of  $\sim 0.09$ , which was equal to L-tyrosine solutions having a concentration of  $\sim 1.8 \times 10^{-4}$  M using a 0.4 cm pathlength.

The solubility of L-tyrosine in water at 25°C is 0.45 mg/mL<sup>77</sup>. Stock solutions of  $1.5 \times 10^{-3}$  M were prepared by dissolving 0.0068 g of L-tyrosine in 25 mL volumetric flasks, filled to the mark with each of the buffers presented in Table 4. By pipetting 1.2 mL from these stock solutions in 10 mL volumetric flasks and completing to the mark with the appropriate buffer, the final solutions ( $1.8 \times 10^{-4}$  M) were obtained.

For each buffer, the L-tyrosine samples were prepared in triplicate and analysed fresh at 25°C after waiting a few minutes for thermal equilibration.

### 2.4.3 Parameters for the data collection

The parameters used for the data collection were very similar to 2.3.3, except for a few points that are listed below:

- EEM-ARMES and normal EEM data were collected for a  $\lambda_{\text{ex}}/\lambda_{\text{em}}$  range of 240–330/270–400 nm (2 nm steps, scan rate of 1200 nm min<sup>-1</sup>) with 10/10 nm excitation/emission slit widths.
- The PMT voltage was increased to 650 V instead of 600 V to obtain a signal with a satisfying intensity (and higher noise contribution unfortunately).
- EEM-ARMES data were collected using dWGP and samples were measured with four different polarizer settings: VV (vertical-vertical), VH (vertical-horizontal), HH (horizontal-horizontal), and HV (horizontal-vertical). For normal EEM data, the dWGP were removed and instead a 3% attenuation filter<sup>f</sup> was used in the emission channel to avoid saturation of the signal. This was preferred over decreasing the detector voltage or using smaller slits as we wanted to keep EEM-ARMES and EEM data perfectly comparable. The collection of a normal EEM measurement involved a single MDF measurement only (as opposed to ARMES which requires four).

---

<sup>f</sup> This is a type of neutral density filter, which are designed to reduce transmission evenly across all wavelengths

### **3. Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters and validation standards**

This chapter is structured into three parts. First an introduction details the main concepts of fluorescence anisotropy and explains what exactly is the ARMES methodology. The second part focuses on the hardware improvement that was necessary to enable standard fluorescence spectrometers to accurately measure anisotropies between 250 and 300 nm, *i.e.* where proteins naturally absorb light. The last part deals with the validation of the hardware in the UV and visible using appropriate fluorophore standards. Most of the results presented in this chapter were published in *Methods and Applications in Fluorescence* in August 2017<sup>15</sup>.

#### **3.1 Fluorescence polarization and anisotropy**

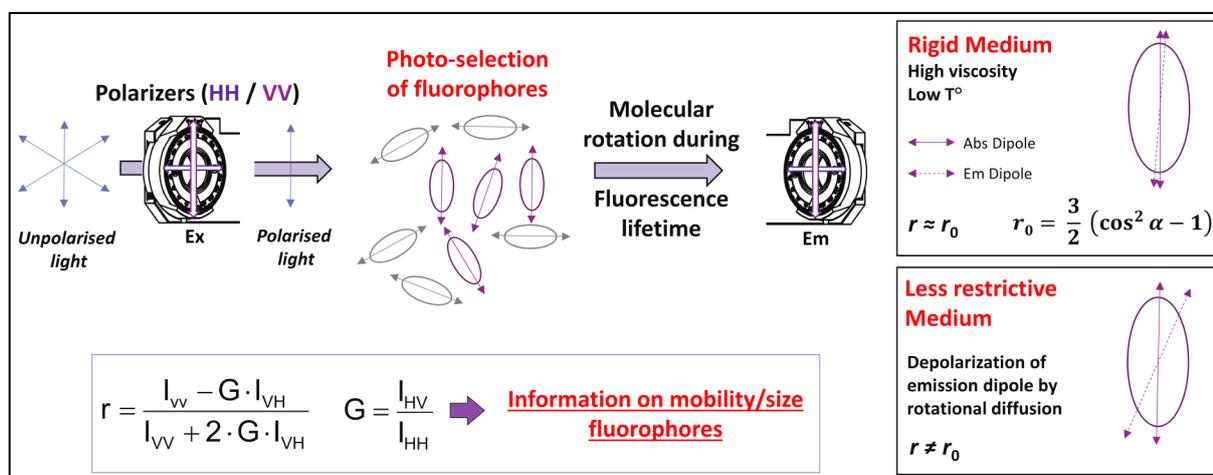
Fluorescence polarization, or anisotropy is based on the fact that fluorophores preferentially absorb photons that have an electric field vector that is aligned parallel to their absorption transition dipole. Upon excitation with polarized light, the fluorophores that are selectively excited from the total fluorophore population are those fulfilling the condition of alignment with electrical field of incident photon (photoselection). It can generally be considered that the closer the angle between the absorption transition dipole and photon's electric field is to 0°, the greater the probability of fluorophore excitation<sup>78-79</sup>. This selective excitation results in partially polarized emission. The relative angle,  $\alpha$ , that exists between the absorption and emission dipole (in absence of depolarization sources) defines the maximum anisotropy or limiting anisotropy  $r_0$  of a fluorophore. This can be measured when the fluorophore is fixed and prevented from moving *e.g.* in a vitrified state<sup>78</sup>.

**Equation 21**

$$r_0 = \frac{1}{5}(3 \cos^2 \alpha - 1)$$

However during the fluorescence lifetime, the orientation of the fluorophore can change due to molecular motion *e.g.* rotation. This was first described by Perrin<sup>80</sup>, who showed that the fluorescence depolarization can occur during the fluorescence lifetime because of a phenomenon called rotational Brownian diffusion. He also showed that the extent of depolarization was dependent on the rate of rotational diffusion, and was thus highly related to

the viscosity of the solvent<sup>81</sup>. The less viscous the solvent, the faster the fluorophore rotates, and this can lead to the complete randomization of the emission dipole orientation.



**Figure 7: Polarization of incident light by a vertically oriented polarizing filter, leading to the photo-selection of a randomly distributed population of fluorophores. Both insets on the right show the influence of the medium rigidity on the extent of depolarization that can occur during the fluorescence lifetime.**

Anisotropy ( $r$ ) and polarization ( $P$ ) are measures of the degree of emission polarization. They are calculated from the measurement of fluorescence intensities when the excitation and emission polarizers are parallel ( $I_{||}$ ) or perpendicular ( $I_{\perp}$ ) to each other:

**Equation 22**

$$r = \frac{I_{||} - I_{\perp}}{I_{||} + 2 \times I_{\perp}}$$

**Equation 23**

$$P = \frac{I_{||} - I_{\perp}}{I_{||} + I_{\perp}}$$

Experimentally, a correction factor ( $G$  factor<sup>§</sup>) needs also to be introduced into these equations. The  $G$ -factor represents the instrument sensitivity ratio between the vertical and horizontal polarized light (Equation 24) and is used to correct for instrument response. In the emission channel, the monochromator grating and to a lesser extent the light detector itself, can be highly sensitive to the orientation of the incoming polarized light. This preference for transmitting one light polarization over another is also wavelength dependant<sup>58,83</sup> and can in some cases cause a very large loss of intensity in the polarization component oriented perpendicularly to the grating's lines. This sudden loss in intensity, called the "Wood's anomaly" and originally thought to be related to a grating deficiency, can appear as shoulders

<sup>§</sup>  $G$  comes originally from grating<sup>58</sup>, can also be called normalization factor is referenced nowadays as "G-factor" after the work of Azumi *et al.*<sup>82</sup>.

in emission spectra and be easily mistaken for real spectroscopic features<sup>79,84</sup>. This is only an extreme example of monochromator dependences on polarization that needs to be corrected by the G-factor:

**Equation 24**

$$G = \frac{I_{HV}}{I_{HH}}$$

**Equation 25**

$$r = \frac{I_{VV} - G \times I_{VH}}{I_{VV} + 2 \times G \times I_{VH}}$$

Where  $I_{\parallel} = I_{VV}$  and  $I_{\perp} = G \times I_{VH}$

The corrected anisotropy is calculated using Equation 25 where  $I_{VH}$  (as an example) represents the fluorescence intensity obtained for a vertical polarized excitation light and horizontal polarized emission light. Measurements with four different polarization settings (HH HV VH VV) are necessary to calculate the anisotropy ( $r$ ).

Anisotropy and polarization are inter-changeable terms and provide almost the same information (Equation 26, Equation 27). The notion of anisotropy was first introduced by Jablonski<sup>85</sup>, and is generally preferred in biophysics or biochemistry instead of polarization because it contains the total intensity in its formula ( $I_{\parallel} + 2I_{\perp}$ ), which normalizes the difference between  $I_{\parallel}$  and  $I_{\perp}$ <sup>78,86</sup>.

**Equation 26**

$$r = \frac{2 \times P}{3 - P}$$

**Equation 27**

$$P = \frac{3 \times r}{2 + r}$$

The observed polarization and anisotropy ( $r$  and  $P$ ) are always smaller than the limiting values (intrinsic)  $r_0$  and  $P_0$ , the latter being measured under vitrified conditions where molecular motion has ceased. Perrin<sup>87-88</sup> gave the relation between  $P$  and  $P_0$  for a spherical molecule, introducing in the equation the effect on polarization of temperature and solvent viscosity:

**Equation 28**

$$\frac{1}{P} - \frac{1}{3} = \left( \frac{1}{P_0} - \frac{1}{3} \right) \left( 1 + \frac{RT}{\eta V} \tau_0 \right)$$

Where  $T$  is the absolute temperature,  $R$  the gas constant,  $\eta$  the medium viscosity,  $V$  the molecular volume of the rotating unit, and  $\tau_0$  the lifetime of the excited state. From this

equation,  $P$  tends to  $P_0$  when  $T/\eta \rightarrow 0$ . By introducing the concept of rotational correlation time<sup>89</sup> ( $\theta$ )<sup>h</sup>, the equation can also be written as:

**Equation 29** 
$$\frac{1}{P} - \frac{1}{3} = \left( \frac{1}{P_0} - \frac{1}{3} \right) \left( 1 + \frac{\tau_0}{\theta} \right)$$

Or using the anisotropy:

**Equation 30** 
$$\frac{1}{r} = \frac{1}{r_0} \left( 1 + \frac{\tau_0}{\theta} \right)$$

If  $\theta \gg \tau_0$ , then the observed anisotropy ( $r$ ) is equal to the fundamental anisotropy ( $r_0$ ). However, if  $\theta \ll \tau_0$ ,  $r$  tends to 0.

Whereas steady state anisotropy gives an “average” value of the anisotropy, time-resolved anisotropy by recording the shape of anisotropy decay curve, can provide precise information about binding mechanisms or any other association reactions that affect the rotational correlation time and diffusional motion of a fluorophore<sup>90</sup>. The measured anisotropy  $r$  in steady state anisotropy can be related to the average of the anisotropy decay  $r(t)$  which occurs during the rotational motion of the molecule<sup>78</sup>, using the following equation:

**Equation 31** 
$$r = \frac{\int_0^{\infty} I(t)r(t)dt}{\int_0^{\infty} I(t)dt}$$

In this equation,  $I(t)$  represents the intensity decay at any time  $t$ . The form of Perrin’s equation described in Equation 30 is only correct if the fluorescence intensity decreases monoexponentially. Time resolved anisotropy has been used to study the motion and flexibility of immunoglobulins in solution, by comparing the anisotropy decay of free IgG or IgE to their antigen-bound form<sup>91-92</sup>. Kawato *et al.* used time resolved anisotropy to reveal the dynamic structure of lipid bilayers, after labeling the hydrocarbon region with a fluorescent probe 1,6-Diphenyl-1,3,5-hexatriene (DPH)<sup>93</sup>. He also investigated the effect of cholesterol on the molecular motion of the labeled hydrocarbon regions<sup>94</sup>, and showed that in the presence of cholesterol, rotational diffusion was hindered, modifying the anisotropy decay of DPH. Time-resolved anisotropy was also widely used to investigate DNA dynamics, studying the change occurring in correlation time after formation of DNA-ethidium bromide (EB) complexes<sup>95-96</sup>, or binding of DNA with HIV-1 integrase<sup>97</sup>.

---

<sup>h</sup> The rotational correlation time, or rotational relaxation time as defined originally by Debye, corresponds to the time required for a molecule to rotate through an angle of one radian. It represents the rate with which rotational diffusion occurs.

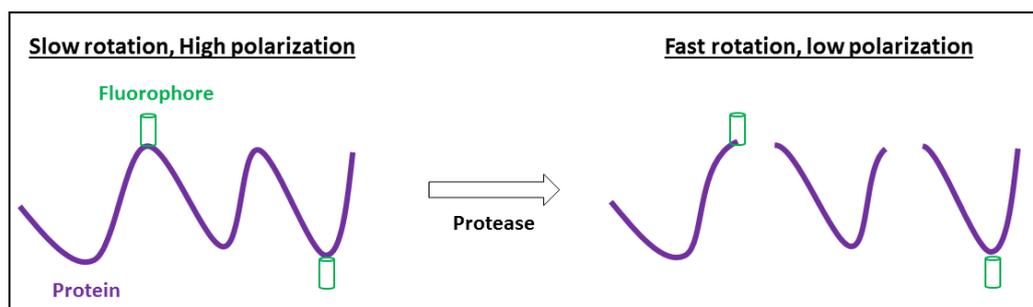
Time-resolved anisotropy decays for complex systems like proteins are expressed generally as the sum of multi-exponential decays<sup>98</sup>:

**Equation 32** 
$$r(t) = \sum_j r_{0j} e^{-\frac{t}{\theta_j}} + r_\infty$$

Where  $r_\infty$  is called “limiting anisotropy”.  $r_\infty$  is obtained at much longer times than the fluorescence lifetime, and is generally equal to 0. However in the case of fluorophores having hindered angular motion, the anisotropy can decay to non-zero values ( $r_\infty > 0$ ).

### 3.1.1 Applications of conventional fluorescence polarization spectroscopy and polarization imaging

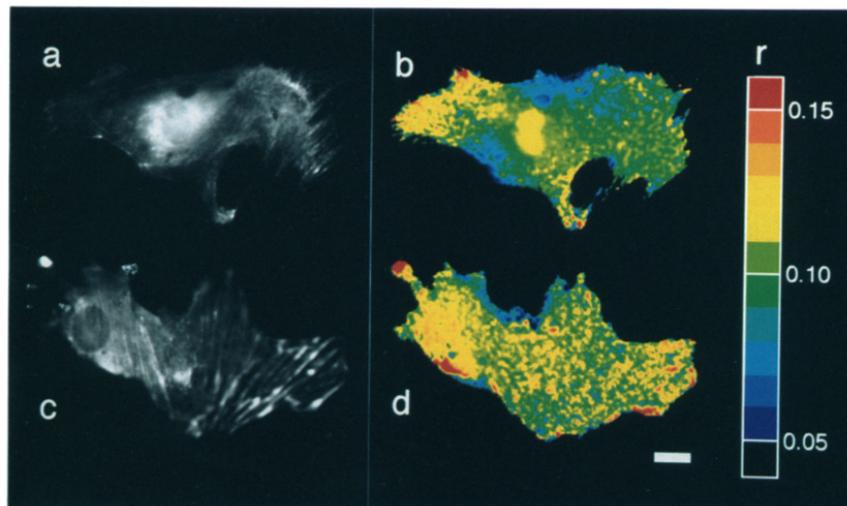
In the early 50’s, Gregorio Weber was the first to develop the use of fluorescence polarization (FP) spectroscopy or nearly equivalent fluorescence anisotropy (FA) for biochemical applications<sup>99-100</sup>. Following his work, this method and its applications became widespread in the clinical, biomedical, and life science fields as highlighted by Jameson *et al.*<sup>86</sup>. FP/FA proved to be useful for studying any molecular interactions or reactions that occur simultaneously with a change in rotational correlation time or local microviscosity. This made these techniques particularly suited to the study of protein interactions/structure<sup>101-102</sup>. Fluorescence polarization immunoassays (FPIs) appeared first and were used to follow the reaction of labeled haptens or antigens<sup>103</sup>, before being applied successfully to high molecular weight macromolecules such as proteins. For example, a Re(I) metal-ligand complex, displaying a highly polarized emission, was used as fluorescent probe to study the binding of anti-HSA with the Re-HSA conjugate<sup>104</sup>. FPIs became widely applied to high-throughput screening applications<sup>105</sup>, designed to identify and quickly assess the activity of enzymes that are essential for many biochemical processes. Proteolysis, which is the hydrolysis of peptide bonds by enzymes called proteases, is another kind of biochemical process particularly suitable for FPI as it involves the cleavage of a labeled protein into lower molecular weight fragments, where the attached fluorophore can rotate faster, causing a decrease in anisotropy.



**Figure 8: Illustration of a protease assay mechanism.**

Protease assays<sup>106-110</sup> as well as kinase assays<sup>111-114</sup> corresponding to the transfer of a phosphate group from a donor to a protein acceptor, are common examples of FPIs. Lea *et al.*<sup>115</sup> highlighted some of the recent applications for small molecule screening, with cases of FPIs successfully designed for G-protein coupled receptors (GPCRs), a prominent therapeutic target and transmembrane receptor family responsible for the cellular response towards the presence of an external molecule<sup>116</sup>. Other applications of FP/FA can be found in the literature, such as determination of free zinc concentration in solution<sup>117</sup>, or for glucose sensing<sup>118</sup> which confirm the versatility and wide usage of this technique.

Polarization imaging (PI) or polarimetry is another active research area in fluorescence, which combines the high resolution of a microscope with the ability of FP/FA to interrogate the local motion of molecules. Axelrod contributed a lot to this field<sup>119</sup> and published one of the first examples of its application in a biological context. He imaged the orientation of a carbocyanine dye in the membrane of erythrocyte ghost cells<sup>120</sup>. Gough *et al.*<sup>121</sup> used PI to study the binding of calmodulin in fibroblasts, and showed that calmodulin bound only in certain regions of the cell (Figure 9), such as leading lamellae and in the retracting tail of highly migrating fibroblasts.



**Figure 9: General pattern of calmodulin binding observed in highly polarized and migrating fibroblasts. (b,d): ~ 20% higher anisotropy observed in regions of tail retraction and leading lamella. Reproduced from Gough *et al.*<sup>121</sup>, with permission from Journal of Cell Biology.**

PI appeared to be a powerful method to provide detailed information about the interior of cells<sup>122-123</sup>, and started to be applied recently to the detection of cancerous cells. Yaroslavsky *et al.*<sup>124</sup> built an automated PI based system to detect and delineate nonmelanoma skin cancers.

Pu *et al.* also used PI combined with time-resolved fluorescence techniques to compare cancerous and non-cancerous prostate tissues<sup>125</sup>, and could identify cancerous tissue from their higher intrinsic fluorescence intensity in the near infrared spectral wing region<sup>1</sup>, as well as their longer anisotropy decay. Both techniques measured differences in the micro-environment of fluorophores in the cancerous tissue compared to normal tissue.

However, conventional FP and PI are both 2-D fluorescence techniques involving a single excitation and emission wavelength, generally chosen carefully to correspond to the specific fluorophore used in each assay. For more complex systems where the required information may emanate from multiple fluorophores, the emission needs to be associated with other dimensions (*i.e.* excitation, emission, lifetime)<sup>127</sup>. New approaches in fluorescence imaging such as multidimensional fluorescence imaging, emerged as powerful techniques to investigate deeply into living cells<sup>128</sup>, to the point to being able to image simultaneously their multiple signaling activities and reveal their relationship<sup>129</sup>. The most powerful of those techniques combine confocal/multiphoton microscopy with fluorescence lifetime imaging (FLIM). This generates a final image of the fluorophores contained in living cells with up to seven dimensions (fluorescence intensity in the XYZ space, emission spectra, time, fluorescence lifetime and anisotropy)<sup>130</sup>. This idea of increasing the measurement dimensionality to enhance our understanding of the signal also emerged in analytical chemistry<sup>131</sup> and fluorescence spectroscopy, with the development of multi-dimensional fluorescence (MDF) spectroscopy.

### **3.1.2 Multidimensional fluorescence (MDF) spectroscopy and the development of ARMES**

A simple, intensity based MDF spectrum, compared to a standard emission spectrum, is basically a data matrix where each element represents the fluorescence intensity for a certain  $\lambda_{\text{ex}} / \lambda_{\text{em}}$  pair. Excitation-Emission Matrix (EEM)<sup>11</sup> and Total Synchronous Fluorescence Spectroscopy (TSFS)<sup>12</sup> are the two main techniques for MDF measurements. In the case of EEM, emission spectra are collected over a range of excitation wavelengths, creating a matrix where every row or column represents an excitation or emission spectrum. TSFS is slightly different in the sense that excitation and emission are scanned simultaneously, by increasing the interval (called delta  $\Delta\lambda = \lambda_{\text{em}} - \lambda_{\text{ex}}$ ) between emission and excitation. Both measurement

---

<sup>1</sup> The near infrared spectral wing region was first introduced by Zhang *et al.* from Alfano's group, when they discovered that biological tissue under far-IR or NIR excitation light, exhibited a strong native signal that could be used to diagnose cancers<sup>126</sup>.

methods have the advantage of investigating a much greater emission space than conventional 2-D emission spectroscopy (see section 4.2 for more details). EEM/TSFS plots provide unique fingerprints of the multiple fluorophores present in complex samples such as proteins or cell culture media<sup>132-133</sup>. The topography of a MDF spectrum is determined by the combination of all the factors that influence the emission process, such as Förster resonance energy transfer (FRET), quenching, solvent effects, and inner filter effects (IFE)<sup>134</sup>.

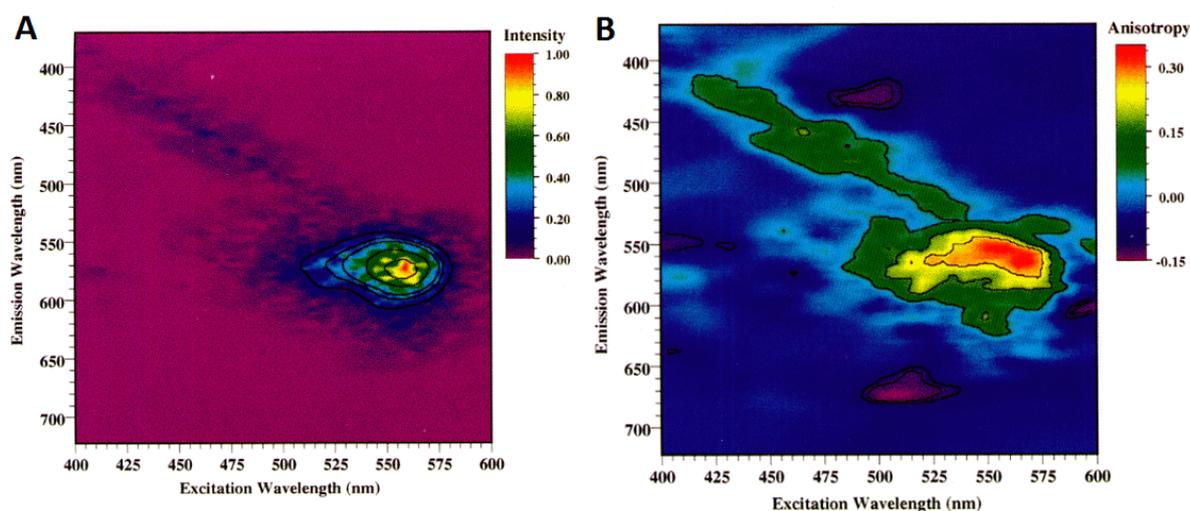
MDF spectroscopy has been shown to be useful in order to assess gross chemical changes in cell culture media<sup>132,135</sup> and to assess protein conformations<sup>136-137</sup>. However due to the extensive spectral overlap existing between fluorophores, MDF spectroscopy only reveals its full potential once combined with factor based methods such as multiway PCA<sup>138-139</sup>, MCR<sup>140-143</sup>, and PARAFAC<sup>144-146</sup>, which help to resolve fluorophore contributions and extract more useful information from the MDF data matrix<sup>147</sup>. Nowadays, the use of multi-way chemometric methods for qualitative and quantitative analysis of MDF data is relatively widespread in academia, with applications in environmental<sup>148-154</sup> and food analysis<sup>155-156</sup>. More recently it has been applied to the biopharma industry, where it has proven to be particularly successful for the simultaneous analysis of multi-fluorophore samples such as hydrosylates, cell culture media and bioreactor broths<sup>132-133,135,157</sup>.

However, even if MDF measurements and conventional 2-D anisotropy are now both well-established techniques, the combination of MDF measurements with anisotropy is rare in the literature. Destrampe *et al.*<sup>158</sup> was one of the first authors to publish and introduce a novel spectrofluorometer<sup>j</sup> for the data acquisition of polarized EEM. They used a dilute mixture of three fluorophores (Coumarin 460, Rose Bengal, Cresyl Violet) in 100% glycerol to show that polarized-EEM was useful for multicomponent resolution<sup>k</sup>. They used a pure solution of rhodamine B and generated a 4<sup>th</sup> dimension in the dataset, by calculating the anisotropy at each point of the emission space. However the anisotropy values obtained were accurate only in a small region localized around the maximum fluorescence intensity (Figure 10).

---

<sup>j</sup> The spectrofluorometer used by Destrampe in 1993, had a 0.35-meter monochromator for excitation and a continuously variable interference filter for selecting the emission wavelength, and allowed the collection of a full EEM spectra in 3.2 min with a detection limit of 500ng/L for fluorescein in ethanol.

<sup>k</sup> In their data analysis, they also found that a more unambiguous resolution was possible if the emission light was measured perpendicularly to the excitation, as it reduced the intensity of the 1<sup>st</sup> order Rayleigh scatter, a source of complication for the separation of fluorophores with similar spectral properties.



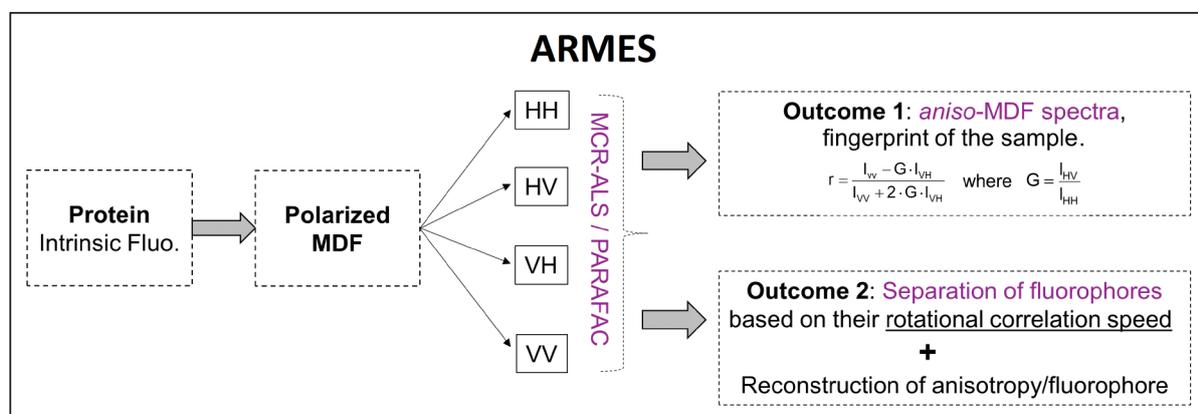
**Figure 10: Anisotropy measurement of rhodamine B at 200 ppb in 100% glycerol. (A) EEM with emission polarizer parallel to the excitation polarizer. (B) Fluorescence anisotropy EEM measurement. Reproduced with permission from Society for Applied Spectroscopy.** <sup>158</sup>

In this study, the inaccurate anisotropy values recovered were due to three factors: first not using a G factor to correct for instrumental response, second the poor extinction ratio for the broad-band polarizers used particularly in the UV region, and third, Rayleigh scatter correction. Nishiyama *et al.*<sup>159</sup> applied the same combination of techniques more than ten years later to the study of spin-coated thin films of polymeric pyridine derivatives (PVPys). In that case the anisotropy “patterns” or *aniso*-MDF spectra generated were corrected by use of G-factor, but the presence of Rayleigh scatter complicated their interpretation due to the overlap with PVPys fluorescence.

Neither of these studies looked at complicated multi-fluorophore systems nor exploited the full potential of this technique. However, despite the inaccurate values obtained and scatter contamination, they did demonstrate that polarized-EEM could potentially increase our understanding of complicated multi-fluorophore systems. Fluorescence anisotropy adds to the MDF dataset information related either to the physical size of the fluorophores or to the rigidity of their surrounding micro-environment<sup>101,160-161</sup>. This new dimension can provide new insights about the origin of the emitting fluorophores<sup>162-163</sup>, and enables the differentiation of components with very similar emission properties in complex mixtures, based on their hydrodynamic volume and rotational correlation speed<sup>164</sup>. Groza *et al.*<sup>13</sup> applied this concept successfully to a complex biogenic sample, in which the protein signal (BSA, big molecule) was clearly discriminated from the overlapping fluorescent background emission from yeastolate (a complex hydrolysate containing many small fluorescent molecules). This enabled

the quantification of protein concentration with a similar level of accuracy to the standard Bradford assay.

The use of factor based chemometrics for the analysis of polarized-MDF was first demonstrated by Groza *et al.*<sup>14</sup>, and led to the development of Anisotropy Resolved Multidimensional Emission Spectroscopy (ARMES) which is outlined in Figure 11. The authors observed in particular that multi-fluorophore proteins (BSA/HSA) generated complex *aniso*-TSFS (excitation wavelength ( $\lambda$ ), wavelength offset ( $\Delta\lambda$ ), anisotropy ( $r$ )) spectra diagnostic of protein structure and stability. This method is the foundation for my research project and is potentially a useful analytical technology platform for the rapid, in-situ, non-contact analysis of multi-fluorophore macromolecules such as proteins.

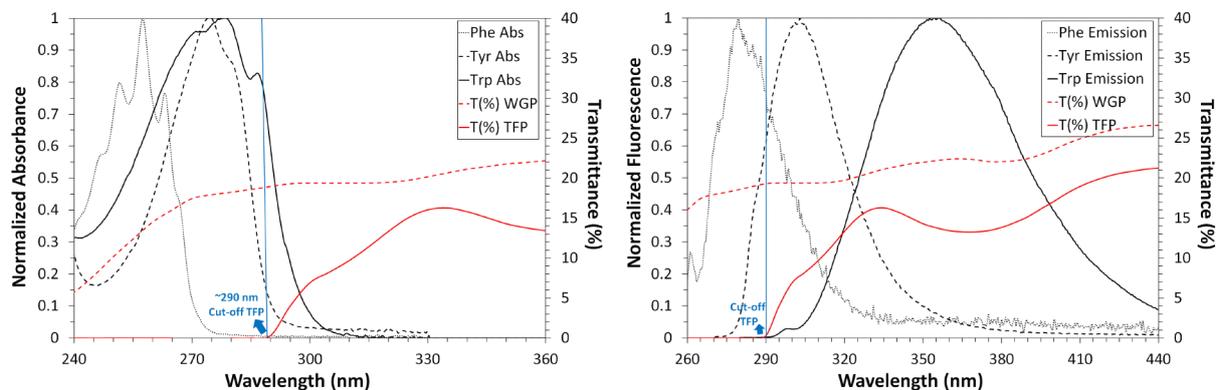


**Figure 11: Schematic representation of ARMES principle and outcomes. Four polarized multidimensional datasets (HH HV VH VV) are collected from intrinsic protein fluorescence. The signal deconvolution is performed by MCR-ALS/PARAFAC and based on the difference of rotational correlation speed between fluorophores. An *aniso*-MDF spectra is also generated from the four original datasets.**

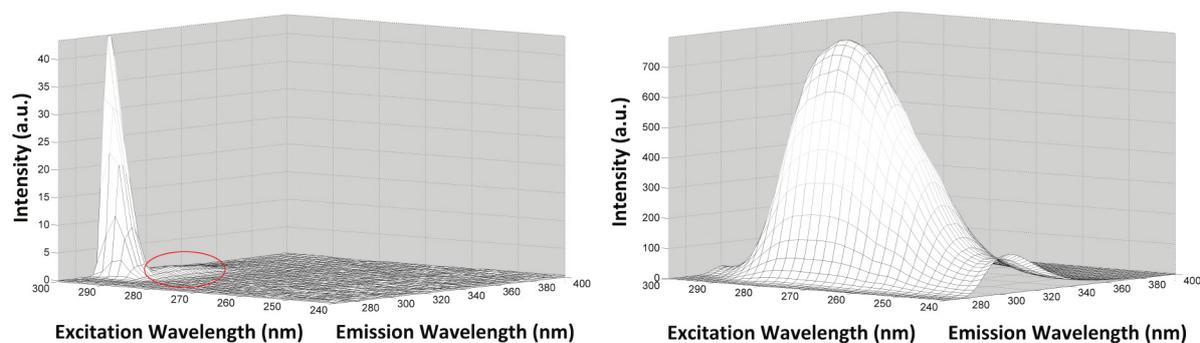
### 3.1.3 Chapter goals

In order to apply ARMES to intrinsic protein fluorescence, it was critical to be able to collect efficiently polarized emission data in the  $\lambda_{ex/em}=250-300$  nm region, as this is where the main amino-acids responsible for the fluorescence of proteins *i.e.* tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe) absorb and emit the light. Groza *et al.* used thin film polarizers (TFP), which strongly absorb below 290 nm and give only weak intensity signals between 290 and 300 nm. TFP also caused the excitation maxima of recovered components to be red-shifted and distorted, compared to non-polarized spectra<sup>14</sup>. The main topic in this chapter was to solve these issues by replacing TFP with wire grid polarizers (WGP), which have much better

transmission in the UV (Figure 12). It was particularly critical for the study of insulin that contains only tyrosine residues (Figure 13).



**Figure 12: Transmittance spectra of WGP and TFP (vertical position) overlay with absorption (left) and emission (right) spectra of Trp Tyr and Phe (data from PhotoChemCad 2.1 software<sup>31-32</sup>).**



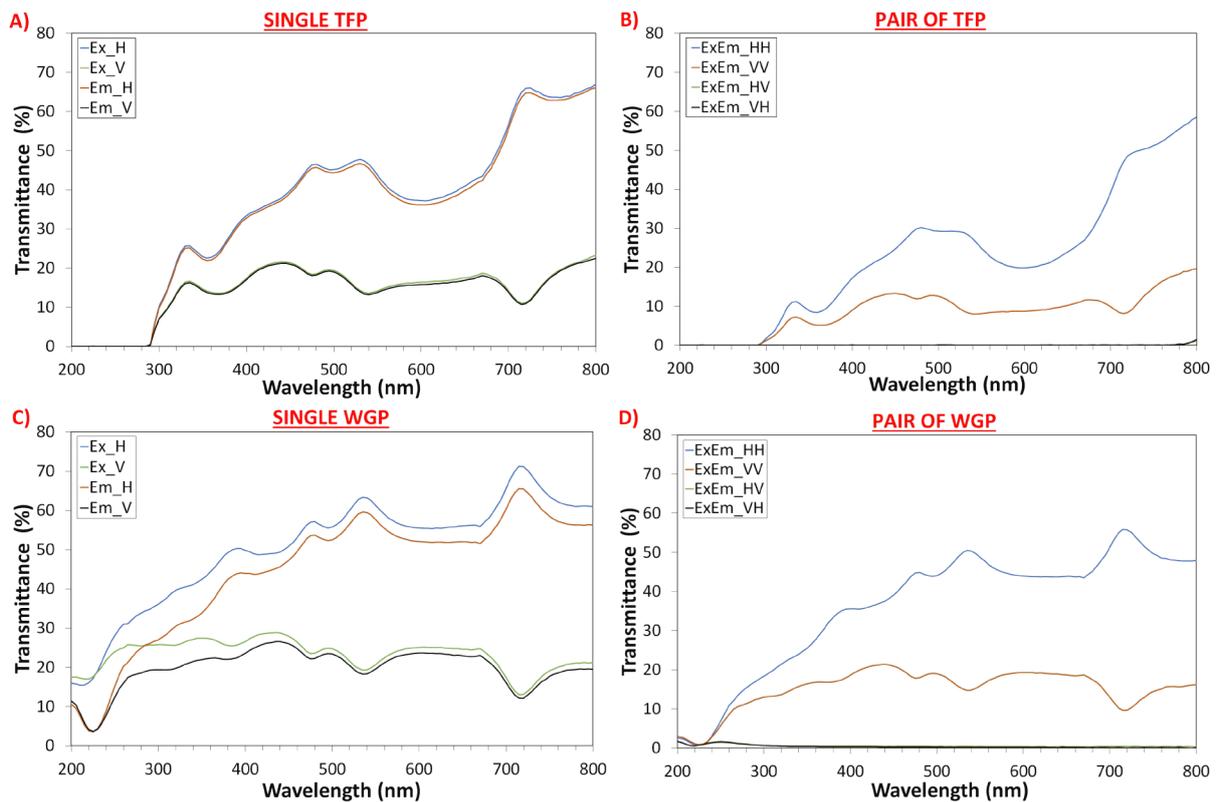
**Figure 13: HH-EEM from 1 mM solution of N-acetyl-L-Tyrosine in PBS buffer (pH 7.2) measured using TFP (left) and WGP (right). The very low fluorescence signal obtained with TFP is highlighted with the red circle (the sharp band is due to the scattered light).**

The other point that is developed in this chapter relates to the necessity of producing a molecule/protocol to validate WGP efficacy and that could be used as a standard method for validating ARMES. However in the liquid state, finding a suitable standard for the UV-vis region was not easy<sup>160</sup>. This chapter explains why perylene, normally used as standard in time resolved anisotropy measurements<sup>165-166</sup>, was a good candidate for MDF anisotropy measurement standard.

### 3.2 Adaption of wire-grid polarizers to spectrofluorometer

In the Cary Eclipse fluorescence spectrophotometer (Agilent Technologies), the polarization of the excitation and emission light is done by TFP which are mounted either on a Manual Polarizer Accessory (part no. 00-100761-00) or on Automated Polarizer Accessory (part no. 8510187100). Here we used ultra-broadband WGP from Thorlabs (part no. WP25M-UB) for a better transmission in the UV. Each of the filters has a vertical (V) and horizontal

(H) setting and can be used along excitation (Ex) or emission (Em) pathlengths. Transmittance spectra of individual filters (Figure 14-A,C) and paired assemblies (Figure 14-B,D) were collected for the four different polarization settings: HH, VV, HV, and VH on a Cary 60 UV-Visible spectrometer, at normal incidence. The face of the filter used to collect the spectra, as well as the filter positions within each individual pair had no influence on transmittance spectra (see section 7.1 in Appendices).



**Figure 14: Transmission spectra of TFP and WGP filter combinations, collected separately first (A, C) as well as in pairs (B, D). Both horizontal (H) and vertical (V) positions were measured. The center of the polarizer was aligned normal to the light beam.**

These plots clearly showed that the TFP were opaque below  $\sim 290$  nm and therefore anisotropy data cannot be generated in this region. WGP on the other hand had much better transmission characteristics. However, the transmittance spectra of individual WGP filters were not identical in contrast to the TFP units, which was not surprising considering that the WGP supplier explicitly stated that filters performance could vary from lot to lot. In some cases, this difference in performance could be attributed to an antireflection coating applied to the protective glass, the quality of which can affect the overall performance of the filters<sup>167</sup>. However it is not the case for these polarizers and the reason for the performance variability observed lies in the difficulty of fabricating high frequency metal gratings on nm size scales

<sup>168</sup>. These small differences between filters should not be a problem as long as polarizer sets are not switched between excitation and emission.

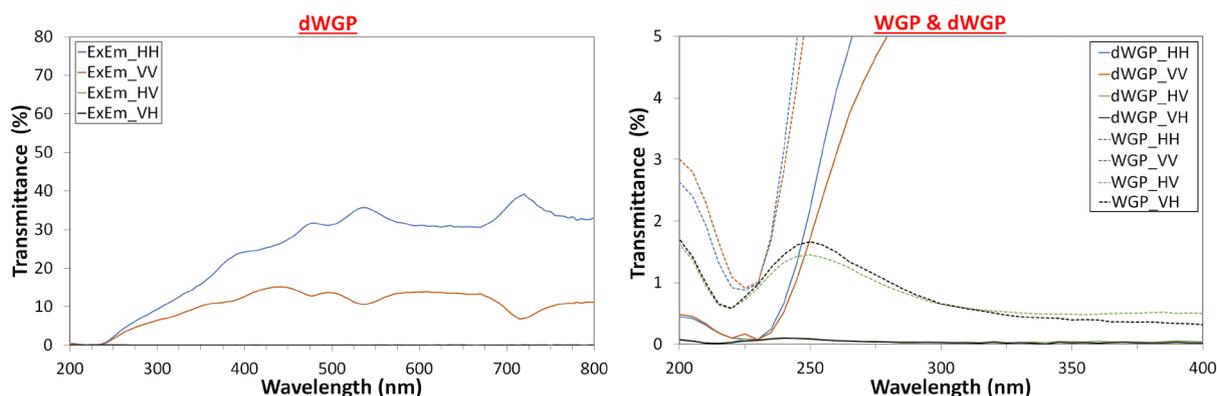
The biggest problem with single WGP filters was their poor extinction ratio, as shown in the HV and VH spectra (Figure 14-D or Figure 15-right below), where appreciable amounts of light were transmitted particularly between 250 and 300 nm which is the primary region of interest for intrinsic protein fluorescence. Poor WGP extinction ratios (~10:1) in this wavelength range was related to wire spacing and thickness<sup>169</sup>. This arose because to attain a high extinction ratio, the wires need to have a spacing and thickness much smaller than the wavelength of the incident light. Between 250–300 nm, the incident wavelength is much closer to the wire spacing explaining the decrease in performance (Table 5).

**Table 5: Performance data for the Ultra broadband WGP (WP25M-UB) used in the study. Data were obtained from manufacturer (ThorLabs).**

WGP	Extinction ratio	Other characteristics
Wavelength range 250 nm – 4 $\mu$ m	[250-300] nm: > 10:1 [300-600] nm: > 100:1 [600-2250] nm: > 1000:1 [2250-4000] nm: > 10 000:1	Fused silica substrate Angle of incidence $\pm$ 20° Thickness = 3.5 mm $\pm$ 0.1 Operating T <sup>0</sup> = -40 to 93°C

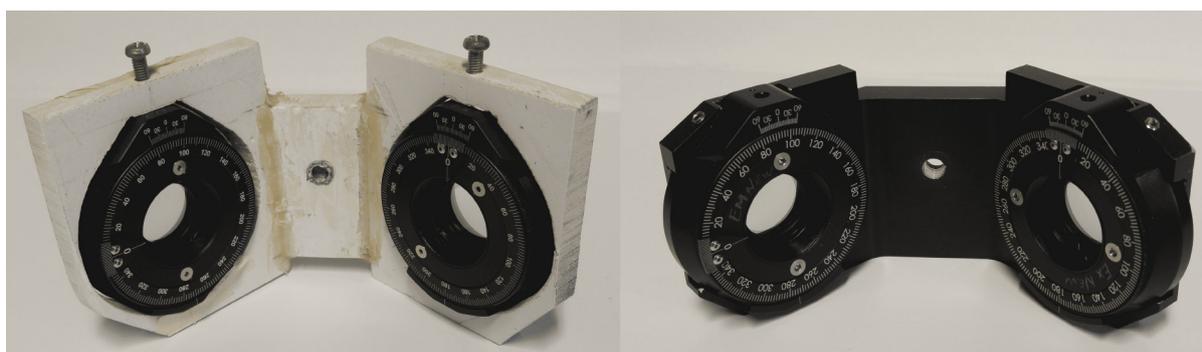
This transmission of non-polarised light adversely affects the accuracy of anisotropy measurements and needs to be corrected. Due to the compact sample compartments of most standard fluorimeters (Cary Eclipse in this case), there was insufficient space for the use of more efficient cube type designs such as Glan-Thompson polarizers<sup>29</sup>. Here, we found that the only other practical alternative was to use pairs of filters on both excitation and emission optical paths (*i.e.* dual-WGP or dWGP), to provide accurate anisotropy measurements. To assemble the dWGP fittings, two filters were very carefully aligned, fixed together and tightly clamped to ensure that the movement could occur during mount rotation.

The improvements are shown in Figure 15, where it is clear that the non-polarized light transmission was drastically improved with a 0.04% average transmittance between 250 and 400 nm (HV and VH orientations) versus 0.8% with the single WGP filter. However, there was a significant (~30%) decrease in the overall transmittance of polarized light. For example, in HH the change is from 17.5% to 13.1% average transmittance (250-400 nm) and for the VV measurement it decreased from 11.1% to 7.9%.



**Figure 15: (left) Transmittance spectra of pairs of dWGP (4 filters used) compared with transmittance spectra of a single WGP (right), in the region of interest.**

The dWGP were then fixed in a homebuilt fitting which was mounted in the sample compartment in a similar fashion to the Agilent manual polarizer accessory. All the measurements in this chapter were done using the first version of the holder (Figure 16, left). A final version was designed afterwards (Figure 16, right), for which the dimensions as well as the autoCAD design are given in Appendices, section 7.2. In any case, we always verified that the pair of filters, once fitted into the holder, remained perfectly aligned by checking that no light was passing for crossed polarizer positions (HV and VH), and were then tightly screwed into the mounting to ensure that no shift could occur during the rotation of the mount wheel. The next phase was to validate that these filters were able to give accurate anisotropy values in the UV region corresponding to protein emission.



**Figure 16: (left) First version (right) final version of the in-house fabricated dWGP filter holder for Cary Eclipse Spectrofluorometer.**

### 3.3 Perylene: a potential ARMES standard

A suitable standard needed to be found in order to validate the use of the dWGP for anisotropy measurements in the 250-400 nm range, and also to provide a standard method for ARMES in order to validate periodically method accuracy, robustness and for calibration transfer. An ideal standard should be<sup>160,170-171</sup>:

### 3. Extended wavelength ARMES measurements: better filters and validation standards

- chemically stable over the required temperature range (10 to 25°C for most of the standards discussed),
- available in high purity,
- have a high anisotropy value in solution,
- have small overlap between excitation and emission spectra, and a high quantum yield to enable the use of dilute concentrations ( $\sim 1\mu\text{M}$ ) where depolarization processes by re-absorption are minimized,
- referenced by other authors for comparison purposes,
- have an anisotropy independent of the wavelength (excitation in particular).

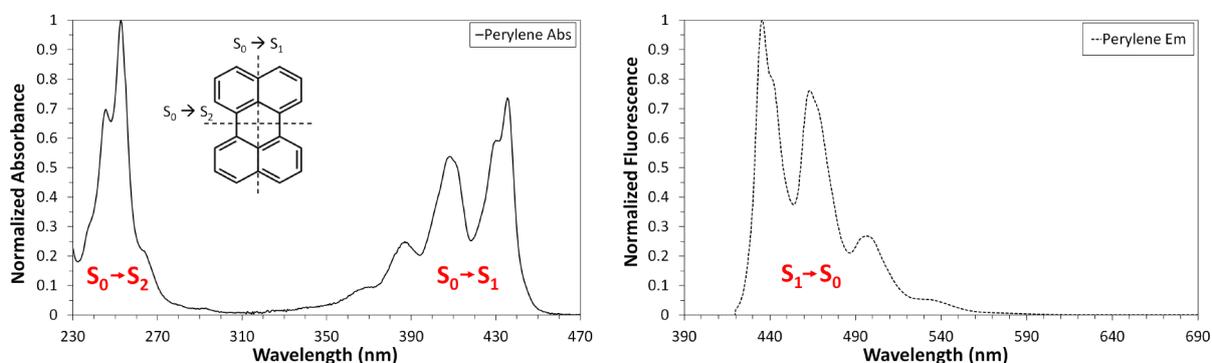
However, when the literature was reviewed, there was no standard available for the deep UV region in the liquid state and almost all the standards discussed<sup>160,171</sup> were only usable for  $\lambda_{\text{ex}} > 400$  nm.

The reason for this lack of standards lies in the desire to selectively reach the lowest energetic excited state  $S_1$ , in order to generate a high anisotropy, independent of the excitation wavelength. Some molecules can have one or several absorption dipoles<sup>172</sup> such as the case for anthracene, pyrene, and naphthalene. The dipole orientation depends on molecular structure and symmetry, and greatly differ for each absorption band. Upon excitation into higher electronic state, excess energy is usually lost by internal conversion, leaving the system in the lowest vibrational level of the  $S_1$  state prior to emission (Kasha's rule<sup>173</sup>). Since anisotropy provides a measure of the angle  $\alpha$  between the absorption and emission dipoles, and that the emission dipole is constant due to Kasha's rule, this means that the average angle  $\alpha$  will vary according to which absorption dipole is engaged, and thus be dependent on the excitation wavelength. Instead of considering it as a rotation of the absorption transition moment with the excitation wavelength, it is better to see it as a change in the average orientation of the photo-selected absorption dipoles. Another way to say this is to consider that when the excitation wavelength changes, there is a simultaneous change in the contribution of the different electronic transitions involved, leading to a varying  $\alpha$  angle and thus a changing measured anisotropy. This property explains for example why the well-known Rhodamine fluorophores (*i.e.* Rhodamine 101), useful as standards because their limiting anisotropy  $r_0$  is close to or equal to the theoretical one-photon fundamental anisotropy of  $2/5$ , cannot be used in UV region (250 - 400 nm) where there are up to seven  $S_0 \rightarrow S_n$  bands<sup>174</sup>. The same issue explains why Fluorescein, Erythrosin B, Phloxine B, or Rose Bengal are also only used as standards for  $\lambda_{\text{ex}} > 500$  nm<sup>171</sup> (where only  $S_0 \rightarrow S_1$  absorption occurs).

It is interesting to note that there are some exceptions to Kasha's rule, such as azulene<sup>175</sup>, cyclopenta[cd]pyrene CPP<sup>176</sup>, and some aromatic thioketones<sup>177</sup> which have an energy gap sufficiently large between  $S_1$  and  $S_2$  to reduce the rate of internal conversion, and allow emission directly from the  $S_2$  state. Itoh<sup>178</sup> summarized this interesting photophysical property in a very good review recently. Another exception to Kasha's rule is observed in molecular crystals<sup>179</sup>, where auto-ionization<sup>1</sup> can occur and acts as a competing path to classical internal conversion. Apart from these few exceptions, most molecules obey Kasha's rule and their emission occurs from the lowest vibrational state of  $S_1$ . It explains why anisotropy is generally independent of emission wavelength (solvent relaxation effects being neglected<sup>180-181</sup>).

During investigations to find a suitable standard for the 250-400 nm region, perylene a highly symmetric polynuclear aromatic hydrocarbon currently used as a standard for time resolved anisotropy<sup>165-166,182</sup> was selected. This molecule attracted attention because it has two absorption bands in the required UV-Vis range (Figure 17):

- a short wavelength band ( $\lambda_{\text{ex/em}} \sim 254 \text{ nm}/440 \text{ nm}$ ) oriented along the short symmetry axis of the molecule, corresponding to  $S_0 \rightarrow S_2$  electronic transition.
- a long wavelength band ( $\lambda_{\text{ex/em}} \sim 434 \text{ nm}/440 \text{ nm}$ ) oriented along the long symmetry axis of the molecule, corresponding to  $S_0 \rightarrow S_1$  transition<sup>183</sup>.



**Figure 17: Absorption (left) and emission (right) spectra of perylene (data from PhotoChemCad 2.1 software) overlaid with the picture of the perylene molecule, represented with its short and long symmetry axis.**

These bands correspond to two electronic states oriented along orthogonal symmetry axes, thus generating very different anisotropies<sup>184,165</sup>. According to Kasha's rule, the emission dipole is oriented along the same axis as the  $S_0 \rightarrow S_1$  absorption dipole, which leads to a theoretical angle  $\alpha$  of  $0^\circ$  and  $90^\circ$  respectively for the  $S_0 \rightarrow S_1$  and  $S_0 \rightarrow S_2$  electronic transitions.

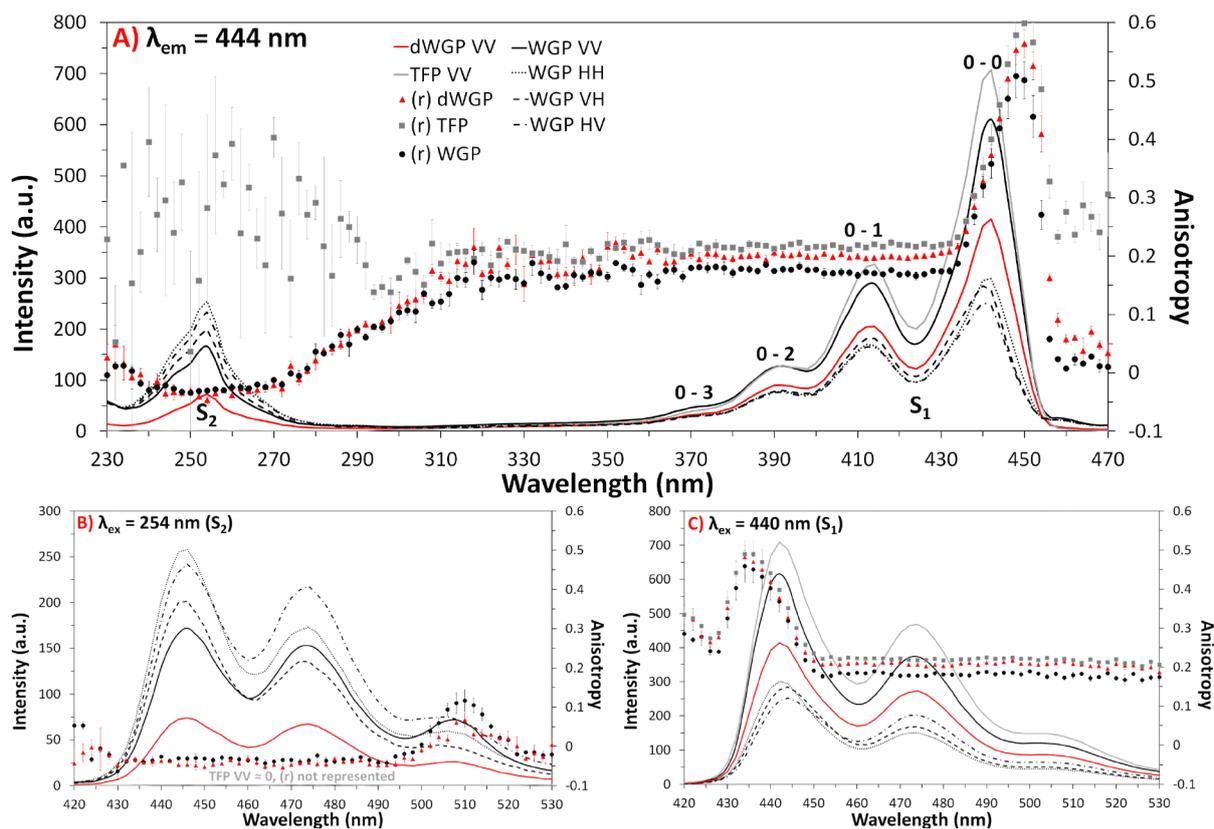
<sup>1</sup> Auto-ionization corresponds to the spontaneous decay of an excited molecule by emission of an electron.

Using Equation 21, a fundamental anisotropy of 0.4 is obtained for  $\lambda_{\text{ex}} > 300$  nm, and  $-0.2$  for  $\lambda_{\text{ex}} < 300$  nm. However, fundamental anisotropy values can only be obtained under cryogenic conditions (typically  $-50^{\circ}\text{C}$  or lower in glycerol) or in polymeric glasses (*e.g.* in solid PMMA<sup>185</sup>) where rotational diffusion processes are absent. The closest measured values to the fundamental value of 0.4 for the low energy band were reported by Johansson<sup>184</sup> and Pantke<sup>186</sup>, who obtained a limiting anisotropy  $r_0=0.37$  below 160K respectively in either propane-1,2-diol/glycerol or ethanol.

In our case, we need a stable standard that can be measured under the same or similar sample conditions as our biogenic samples (*e.g.* insulin) which means that the protocol has to be developed at room temperature, ideally in a safe and common, high purity solvent (*e.g.* water, buffer, or glycerol). This kind of medium being less rigid at room temperature than under cryogenic conditions, depolarization processes will be present and reduce the anisotropy. Despite this fact, perylene still appears to be a promising standard considering that we can collect simultaneously two different anisotropies over a wide wavelength range. The literature is relatively sparse in regard to reference steady state anisotropy measurements of both perylene bands at room temperature, therefore most of the work presented here will be compared to Barkley's 1981 study<sup>165</sup>.

### **3.4 Comparison of filter efficiency using perylene: *conventional 2-D anisotropy***

The first issue with using perylene was to assess the efficiency of the dWGP filter combinations, by comparing the accuracy of the measurements obtained by WGP and TFP to the literature. For this experiment, perylene was dissolved in a viscous solvent (100% glycerol) at a concentration of  $\sim 1\mu\text{M}$ . Because of the solvent viscosity and the orthogonality of the absorption dipoles, the highest fluorescence intensities were expected from the VV and HH polarization settings. More precisely, if the maximum of one band was measured for VV, then the same polarization setting would lead to the minimum for the other band. Below are shown the polarized excitation (Figure 18 A) and emission spectra (Figure 18 B,C) collected for a solution of  $1\mu\text{M}$  perylene in glycerol using TFP, WGP, and dWGP configurations.



**Figure 18: Overlay of polarized excitation (A /  $\lambda_{\text{em}} = 444\text{ nm}$ ) and emission spectra (B /  $\lambda_{\text{ex}} = 254\text{ nm}$ , C /  $\lambda_{\text{ex}} = 440\text{ nm}$ ) collected using WGP, dWGP and TFP, with calculated anisotropies. The spectra and anisotropy represent the average values obtained for 3 replicate measurements of a  $1\mu\text{M}$  solution of perylene in 100% glycerol. For clarity purpose, only the VV spectra are shown for TFP and dWGP.**

Figure 18-A confirms that both excited states have their respective intensity maxima with orthogonal polarization settings:  $S_2$  maximum intensity with HH and  $S_1$  maximum with VV. The intensity obtained using the dual WGP is weaker compared to a single WGP and to the TFP, as expected (Figure 15). The figure also shows the unsuitability of TFP for UV use with no emission being measured for the high-energy band, leading to random spurious anisotropy values. Another major source of error in anisotropy, observed in both excitation and emission spectra, is due to the presence of Rayleigh scatter. This type of scatter often appears as 1<sup>st</sup> order ( $\lambda_{\text{ex}} = \lambda_{\text{em}}$ )<sup>187</sup> and 2<sup>nd</sup> order ( $\lambda_{\text{ex}} = 2 \times \lambda_{\text{em}}$ ) artefacts in EEM due to the grating diffraction effect<sup>188</sup>.

Rayleigh scatter is elastic light scattering from molecules and small particles. It is a simplification for small particles<sup>189</sup> of the broader Mie theory<sup>190</sup>, applicable for spherical particles of a size similar to the incident light wavelength. Tyndall scattering<sup>191-192</sup>, induced by colloids or particles in fine suspension, is sometimes included under the term of Rayleigh scatter<sup>193-194</sup> due to their very similar nature. Because of the almost pure polarized nature of the scattered light<sup>158</sup>, even a small degree of scatter will drastically distort the observed

anisotropy<sup>29</sup>. For perylene, the 1<sup>st</sup> order Rayleigh scatter band overlaps the main S<sub>1</sub> emission band and causes large erroneous anisotropy values (Figure 18 C). This overlap, which could be mistaken for a small blue shift in the emission, is even stronger for the VV and HH emission profiles, which is explained again by the highly polarized nature of scattered light. It can also be seen in the emission band of the S<sub>2</sub> excited state (Figure 18 B), where the intensity of the 2<sup>nd</sup> order Rayleigh scatter ( $\lambda_{em} = 2\lambda_{ex}$ ) appears larger in VV and HH compared to VH and HV.

Even though the presence of Rayleigh scatter and shot noise limits the working range for perylene, one can clearly see that specific spectral regions ( $\lambda_{ex} = 240\text{-}270$  nm and  $\lambda_{ex} = 370\text{-}430$  nm) have stable, constant anisotropy values (Figure 18). The different polarizers were evaluated (Table 6) by comparing the average anisotropy ( $r$ ) value obtained across these two wavelength ranges, and their relative standard deviation (RSD).

**Table 6: Anisotropy and standard deviation (STD) values obtained in the short ( $\lambda_{ex} = 244\text{-}264$  nm) and long ( $\lambda_{ex} = 390\text{-}430$  nm) wavelength region for a 1  $\mu\text{M}$  solution of perylene in glycerol, measured in triplicate using WGP, dWGP and TFP.**

Polarizers	Average $r$ value for $\lambda_{ex} = 244\text{-}264$ nm (S <sub>2</sub> )	STD (n = 3)	Average $r$ value for $\lambda_{ex} = 390\text{-}430$ nm (S <sub>1</sub> )	STD (n = 3)
WGP	-0.026	$\pm 0.004$	0.173	$\pm 0.004$
dWGP	-0.035	$\pm 0.006$	0.200	$\pm 0.003$
TFP <sup>m</sup>	Not measurable	N/A	0.217	$\pm 0.003$

Compared to the fundamental anisotropy of both excited states (0.4 for S<sub>1</sub> and -0.2 for S<sub>2</sub>), the values obtained at 25°C were much lower due to the rotational diffusion present at ambient temperatures in the viscous solvent. In general, the anisotropies obtained with the dual WGP were slightly closer to the fundamental values than for a single WGP. Even if the difference was small ( $|\Delta r| = 0.009$  for short wavelength region), it corresponds to a real variation in their accuracy. Barkley *et al.*<sup>165</sup> obtained steady state anisotropy values of -0.031 and  $0.197 \pm 0.002$  for a 2  $\mu\text{M}$  solution of perylene in glycerol at 25°C, which were almost identical to the values obtained using the dWGP. TFP measurements on the other hand were unsuitable for short wavelength range, but were more efficient than the dWGP for the long wavelength range with  $r = 0.217 \pm 0.003$ . This value was also reported by Johansson<sup>184</sup>, who

<sup>m</sup> TFP could not measure anisotropy for the S<sub>2</sub> band.

studied the temperature dependence of fluorescence anisotropy and obtained an anisotropy value  $r \approx 0.22 \pm 0.002$  for perylene in glycerol at  $\approx 300\text{K}$ .

Those small differences being observed were mostly related to the different polarizer extinction ratios, as the measurements were otherwise made in conditions as close as possible identical<sup>n</sup>. Johansson used a SPEX Fluorolog 112 spectrometer, equipped with Glan-Thompson polarizers that have very high extinction ratio<sup>29</sup> of  $10^6$  whereas Barkley used a Perkin-Elmer MPF-4 spectrofluorometer equipped with Polaroid polarizer on the emission, and a Polaroid ( $\lambda_{\text{ex}} > 290 \text{ nm}$ ) or Polacoat polarizer ( $\lambda_{\text{ex}} < 300 \text{ nm}$ ) for the excitation. Both are dichroic polarizers<sup>195</sup> that generally yield a much smaller extinction ratio<sup>196-197</sup> compared to Glan-Thompson polarizers ( $\sim 10^3$  vs  $\sim 10^6$ ), which explains the lower  $r$  values obtained by Barkley in the long wavelength range.

This difference in polarizer efficiency also explains the slightly higher  $r$  values obtained with TFP compared to WGP/dWGP. WGP are usually only comparable to TFP above  $700 \text{ nm}$ <sup>198</sup>. In the case of linear polarizers<sup>o</sup>, the extinction ratio ( $\rho$ ) is normally obtained using a source of plane-polarized light, by calculating the ratio (Equation 33) between the minimum transmittance observed when the polarizer is perpendicular to the plane of polarized beam ( $T_2$ ), to the maximum transmittance observed when they are parallel ( $T_1$ )<sup>199-200</sup>. It is important to notice that  $\rho$  can also be found from the ratio between the maximum and the minimum transmittance, which is how values of  $10^3$  or  $10^6$  were obtained.

**Equation 33**  $\rho = \frac{T_2}{T_1}$

If unpolarized light is being used as a source instead of plane-polarized light, then the transmittance ( $T$ ) for a single polarizer is different and defined as:

**Equation 34**  $T = \frac{1}{2}(T_1 + T_2)$

For a pair of linear polarizers placed in an unpolarized light beam, the transmittance changes again as polarizers can be aligned parallel to each other, or in crossed positions. If the polarizers are parallel, the transmittance is defined as:

**Equation 35**  $T_{\parallel} = \frac{1}{2}(T_1^2 + T_2^2)$

---

<sup>n</sup>  $25^\circ\text{C}$ , diluted solutions to avoid reabsorption effects  $A \leq 0.1$  for the long wavelength absorption band.

<sup>o</sup> Linear can be used to represent the capacity of any polarizer to transform an incident unpolarised beam, into a light where the electric vector evolves mainly along one plane. The polarizers used in this study are all linear.

For a pair of crossed polarizers, the transmittance is defined as:

$$\text{Equation 36} \quad T_{\perp} = T_1 T_2$$

Considering these equations, and implementing the assumption that  $T_2^2 \ll T_1^2$ , a good estimation of the extinction ratio can be obtained for a pair of polarizers illuminated with unpolarized light using the following equation<sup>201</sup>:

$$\text{Equation 37} \quad \rho \approx \frac{T_{\perp}}{2 \times T_{\parallel}}$$

The transmittance spectra (Figure 14, Figure 15) collected were obtained using unpolarized light. Using Equation 37, the extinction ratio can be estimated for the pairs of TFP and dWGP, by replacing  $T_{\parallel}$  with the transmittance spectra obtained for HH (> VV) and  $T_{\perp}$  by the HV spectra corresponding to their crossed position.

**Table 7: Estimation of the extinction ratio ( $\rho$ ) for the pairs of TFP and dWGP in the 390-430 nm range.  $1/\rho$  is also given as some authors define the extinction ratio as the ratio between maximum and minimum transmittance.**

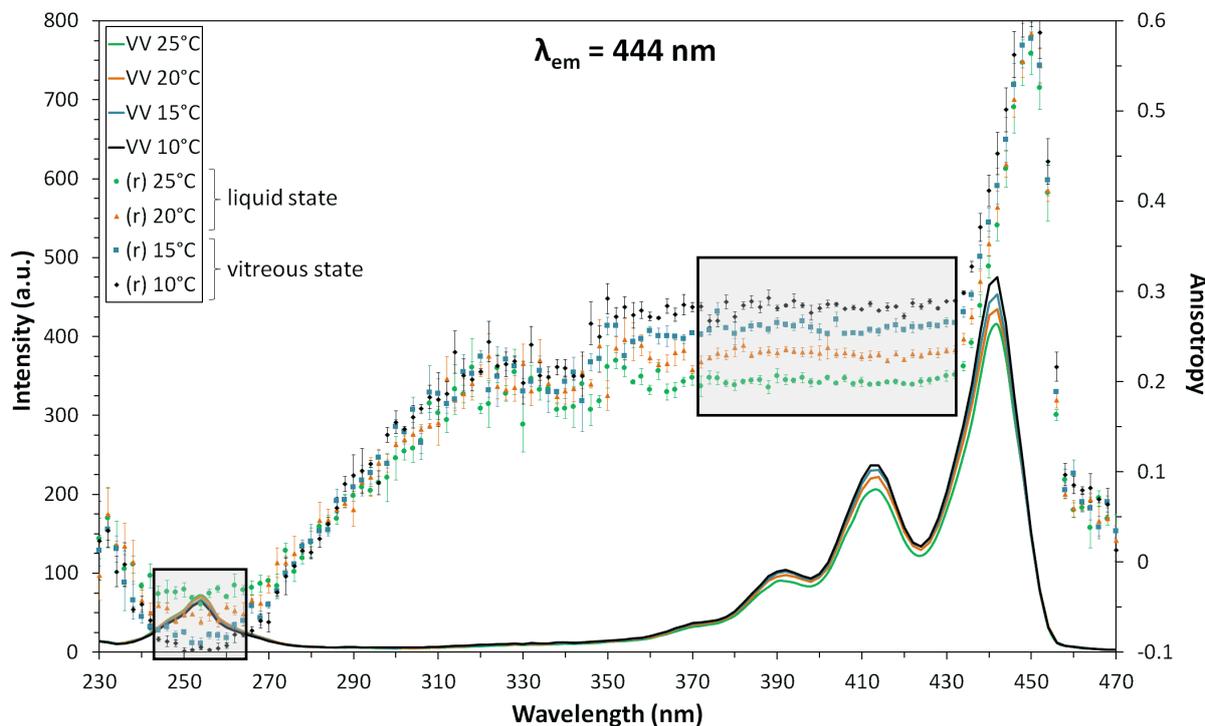
Polarizers	$\rho$ ( $\lambda = 390\text{-}430$ nm)	$1/\rho$ ( $\lambda = 390\text{-}430$ nm)
TFP	0.00072	~ 1400
dWGP	0.00101	~ 1000

As one can see, TFP have a higher extinction ratio in the 390-430 nm range, which explains the higher anisotropy collected using TFP compared to dWGP ( $r = 0.217$  vs  $0.200$ ). In terms of efficiency, Glan-Thompson polarizers (GTP) would be best, unfortunately in many conventional spectrofluorometers, bulky prism based GTP<sup>202</sup> cannot be used due to the limited space in the sample compartment. The dWGP appeared to be the best alternative, as they transmit deep UV light and have proven to give accurate anisotropy values in both UV and visible range.

### 3.5 Temperature dependence study to confirm dWGP accuracy

To further validate the accuracy of the dWGP configuration for accurate anisotropy measurements, we undertook a temperature dependence study, by collecting the same solution of perylene ( $1\mu\text{M}$  in 100% glycerol) at 25, 20, 15, and  $10^{\circ}\text{C}$ . Samples are in the liquid state at

20°C and 25°C, and in the rubbery or “vitreous” state at 15°C and 10°C as discussed by Johansson<sup>184</sup> in his work on perylene<sup>p</sup>.



**Figure 19: Overlay of VV polarized excitation spectra ( $\lambda_{em} = 444$  nm) with calculated anisotropy for perylene ( $1\mu\text{M}$  solution in 100% glycerol), using dWGP. The spectra and anisotropy represent the average values from three replicate measurements. The important regions for the anisotropy values are highlighted in the boxes.**

As the solutions were cooled down from 25 to 10°C, the intensity of bands in the VV spectra increase and decrease respectively for the  $S_1$  and  $S_2$  states respectively. This change in intensity is due to the rigidity of the medium increasing, slowing the rotational diffusion of perylene. Indeed, pure glycerol is known to have a viscosity increasing by 10 orders of magnitude between its melting temperature and the glass transition temperature<sup>206,210</sup>. The averaged anisotropy values across wavelength ranges free from Rayleigh scatter are displayed in Table 8, and compared to similar values reported by Barkley *et al.*<sup>165</sup> in 1981.

<sup>p</sup> Lane<sup>203</sup> in 1925 stipulated that the freezing temperature of pure glycerol was 17°C, value of 18°C also reported by another source<sup>204</sup>. However, the glycerol we used was a viscous liquid even at 10°C. This is due to the capacity of glycerol to exist in a “supercooled” or “rubbery” state, below its melting point<sup>205-207</sup>. Glycerol during cooling does not crystallize but rather becomes a supercooled liquid, which can be vitrified around  $\sim 190\text{K}$  (called the glass-transition temperature  $T_g$ )<sup>208,209</sup>. Due to this characteristic, glycerol is often used as a model for studying physico-chemical properties in a supercooled state or close to glassy state<sup>210-214</sup>. This propensity to not crystallize below freezing temperature is related to the high viscosity of glycerol, and its three hydroxyl groups that can form a complex intermolecular hydrogen bond network<sup>206,215</sup>. Some studies showed that in some cases, a completely dehydrated glycerol can undergo crystallisation<sup>216</sup>. But in general it is a very unusual and unstable process for glycerol, which is very depend on water content, temperature history of the sample, and also on the impurities present that can act as seeds for crystallisation<sup>205</sup>.

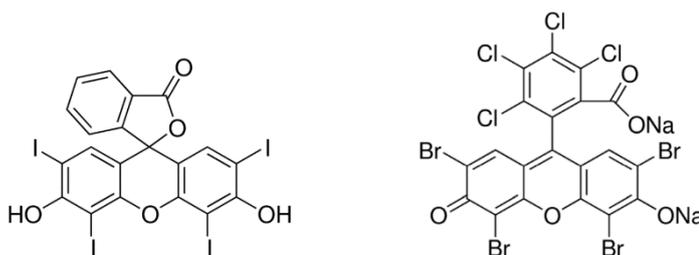
**Table 8: Average of anisotropy values obtained for 1 $\mu$ M solution of perylene in glycerol using dWGP at different temperatures, in the short ( $\lambda_{\text{ex}} = 244\text{-}264$  nm) and long ( $\lambda_{\text{ex}} = 390\text{-}430$  nm) wavelength regions; Intensity ratios of  $I_{\text{HH}}$  ( $\lambda_{\text{ex}} = 254$  nm) to  $I_{\text{HH-noise}}$  (average  $\lambda_{\text{ex}} = 280\text{-}300$  nm), and  $I_{\text{VV}}$  ( $\lambda_{\text{ex}} = 414$  nm) to  $I_{\text{VV-noise}}$  (average  $\lambda_{\text{ex}} = 280\text{-}300$  nm). Data were collected in triplicate.**

	<b>r <math>\pm</math> STD (RSD)</b> $\lambda_{\text{ex}} = 244\text{-}264$ nm	<b>Barkley</b> $\lambda_{\text{ex}} = 256$ nm	$\frac{I_{\text{HH-254}}}{I_{\text{HH-noise}}}$	<b>r <math>\pm</math> STD (RSD)</b> $\lambda_{\text{ex}} = 390\text{-}430$ nm	<b>Barkley</b> $\lambda_{\text{ex}} = 430$ nm	$\frac{I_{\text{VV-414}}}{I_{\text{VV-noise}}}$
25°C	-0.035 $\pm$ 0.006 (17.14%)	-0.031 $\pm$ 0.002	18.9	0.200 $\pm$ 0.003 (1.50%)	0.197 $\pm$ 0.002	37.6
20°C	-0.057 $\pm$ 0.006 (10.53%)	-0.052 $\pm$ 0.002	19.2	0.231 $\pm$ 0.003 (1.29%)	0.231 $\pm$ 0.002	38.1
15°C	-0.079 $\pm$ 0.008 (10.13%)	-0.076 $\pm$ 0.002	19.5	0.260 $\pm$ 0.005 (1.92%)	0.260 $\pm$ 0.002	38.6
10°C	-0.091 $\pm$ 0.007 (7.69%)	-0.097 $\pm$ 0.002	20.0	0.283 $\pm$ 0.004 (1.41%)	0.285 $\pm$ 0.002	39.1

The values obtained here matched Barkley results, with low, stable RSD values for the  $S_1$  state. The RSD gradually decreases for  $S_2$  with temperature, which was due to larger absolute anisotropy values being obtained, leading to less relative error between measurements. This was also verified by checking the intensity ratio between the band of interest and the noise ( $I_{\text{HH-254}} / I_{\text{HH-noise}}$ ), which increases when temperature decreases. These results confirm the accuracy of the dWGP anisotropy measurements, and a higher integration time helps to decrease even more the RSD (Appendix 7.1).

### 3.6 Validation of dWGP accuracy in the 450 to 640 nm range

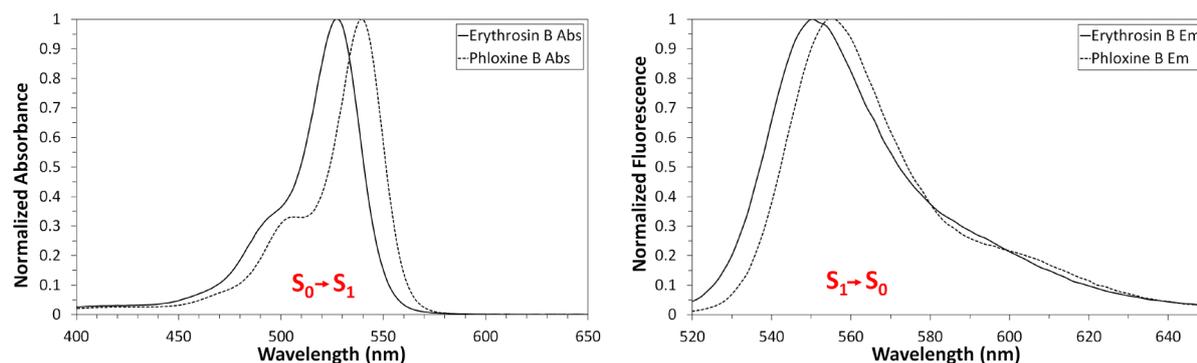
In order to complete the validation of dWGP accuracy, we had to prove that they gave satisfying result in the longer wavelength region. We chose to use Erythrosin B (EB) and Phloxine B (PB), proposed as steady-state anisotropy standards by Thompson *et al.*<sup>171</sup>. Those compounds are both halogenated derivatives of fluorescein (high solubility in water) and are part of the xanthene dye family, also known as triplet forming dyes<sup>217</sup>.



**Figure 20: Structures of Erythrosin B (left) and Phloxine B (right) fluorophores.**

Those molecules are interesting as anisotropy standards because they exhibit a significant anisotropy even in aqueous solutions. The iodine (EB) and bromine (PB) atoms enhance spin-orbit coupling, favoring forbidden intersystem crossing from  $S_1$  to  $T_1$  (also called “heavy atom effect”<sup>218-219</sup>). This causes an effective quenching of fluorescence, decreases the fluorescence lifetime thus reducing the randomization of dipole orientation by Brownian motion<sup>220</sup>.

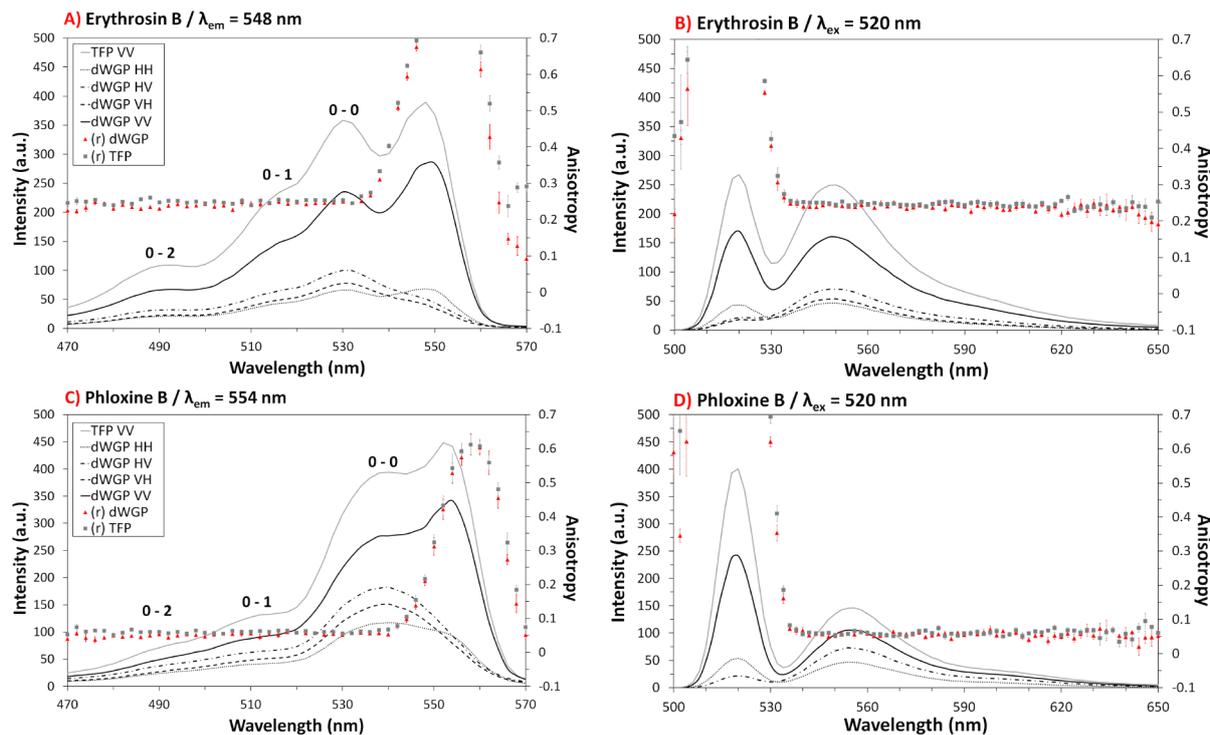
One advantage in relation to the application of ARMES to proteins is that these fluorophores are soluble under similar environmental conditions (*e.g.* HEPES buffer, pH 7.2 at 25°C). They also have very similar absorbance and emission spectra (Figure 21), but with completely different anisotropy values. For  $\lambda_{\text{ex}} > \sim 480$  nm (where only  $S_0 \rightarrow S_1$  absorption dipole is engaged and the fluorescence intensity not too low), Thompson *et al.* obtained a constant anisotropy  $r = 0.243 \pm 0.002$  for EB and  $0.055 \pm 0.002$  for PB at 25°C in 50 mM HEPES buffer. This makes a mixture of these standards the perfect MDF case study to judge if *e.g.* ARMES can accurately recover both fluorophores spectra and reconstruct their anisotropy (this point will be studied in the next chapter).



**Figure 21: Absorption (left) and emission (right) spectra of 10  $\mu\text{M}$  EB and 10  $\mu\text{M}$  PB collected in 50 mM HEPES buffer, pH 7.2 at 25°C. The emission spectra were recorded for  $\lambda_{\text{ex}} = 510$  nm. The absorption and emission spectra displayed correspond to the region of constant anisotropy.**

The anisotropy of a 6  $\mu\text{M}$  solution of EB, and 0.6  $\mu\text{M}$  solution of PB (see section 2.3.1 in material and methods) was measured using dWGP and TFP. The concentrations were chosen in order to limit reabsorption effects that would depolarize the fluorescence emission, but also in order to obtain similar fluorescence intensities for each fluorophore.

### 3. Extended wavelength ARMES measurements: better filters and validation standards



**Figure 22: Overlay of polarized excitation ((A)  $\lambda_{em}=548$  nm, (C)  $\lambda_{em}=554$  nm) and emission spectra ((B)  $\lambda_{ex}=520$  nm, (D)  $\lambda_{ex}=520$  nm) collected using dWGP and TFP for EB and PB, with calculated anisotropies. The spectra and anisotropy are average values obtained from three measurements. For clarity, only VV spectra are shown for TFP.**

As one can see, the red and grey dots are almost overlapping for both filters, in the region free from Rayleigh scatter. We averaged the anisotropy value across this wavelength range ( $\lambda_{ex} = 500-530$  nm) and calculated the standard deviation. The results are compared to the values obtained by Thompson *et al.* (Table 9).

**Table 9: Anisotropy and STD values obtained in 500-530 nm region for a 6  $\mu$ M solution of Erythrosin B and 0.6  $\mu$ M solution of Phloxine B in HEPES buffer (25°C), measured in triplicate using dWGP and TFP.**

Polarizers	Average r value for $\lambda_{ex} = 500-530$ nm. Erythrosin B	Average r value for $\lambda_{ex} = 500-530$ nm. Phloxine B
dWGP	$0.243 \pm 0.002$	$0.054 \pm 0.002$
TFP	$0.248 \pm 0.002$	$0.060 \pm 0.002$
Thompson <i>et al.</i>	$0.243 \pm 0.002$	$0.055 \pm 0.002$

Thompson *et al.* measured the anisotropy using two systems: a ThermoSpectronics AB-2 with film polarizers and a SLM 8000 with Glan Thompson polarizers. Over 10 replicate measurements, the author obtained values with both systems that agreed within the STD of the

measurements. This was also the case with the values we obtained using dWGP and TFP, which proves the efficiency of the dWGP even in the longer wavelength region.

### 3.7 Conclusions

Replacing standard TFP polarizers with dWGP provided much superior UV performance with adequate extinction ratios between 250 and 400 nm. The dWGP were easy to assemble using off-the-shelf components and fitted easily within the spectrometer sample chamber. Overall, the cost of this homebuilt dWGP was less than 4.5k€, of which 80% was due to the filters. dWGP hardware performance was validated using perylene in glycerol and solutions of EB and PB in HEPES buffer and the results were in excellent agreement with the literature<sup>165,171</sup>. This confirmed the accuracy of dWGP anisotropy measurements, over most of the important, short ( $\lambda_{\text{ex}} < 300$  nm) and long ( $\lambda_{\text{ex}} > 300$  nm) wavelength regions required for protein analysis by intrinsic and extrinsic fluorescence.

Perylene in glycerol at temperatures between 25 and 10 °C was an ideal standard for ARMES measurements because it enabled the simultaneous collection of clearly discriminated positive and negative anisotropy values at different wavelengths. This characteristic was critical for a standard suitable for use with multidimensional measurements and chemometric data analysis.

The next chapter explores how to use chemometrics to analyse ARMES data from perylene standard samples. The method is then extended to a mixture of EB and PB, a much more difficult case in terms of chemometric resolution due to the similarity between their spectra. The aim is to find the right chemometric method that enables ARMES to accurately recover each of the fluorophore spectra and subsequent anisotropies from various dyes mixtures. This is a critical step before implementing the method for protein analysis.

## 4. Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR)

This chapter is divided into three parts. The first section introduces some key concepts about the main multivariate methods used in this thesis, starting with a particular focus on the bilinear MCR algorithm before extending to the concept of trilinearity and the PARAFAC method. The second part explains in more details the EEM and TSFS measurements structures. Finally, the last part shows how ARMES can be used to accurately recover the anisotropy of each fluorophore from various dye mixtures. Most of this work was published in *Analytica Chimica Acta* in February 2018<sup>16</sup>.

### 4.1 Introduction

We saw in the previous chapter how the polarization set-up and a standard test method were developed and validated. The next step is to investigate the accuracy of ARMES, which involves the use of MDF measurements rather than conventional 2-D fluorescence. By applying chemometrics to fluorophore mixtures, the goal here is to show how it is possible to recover accurate and unique components from which the precise anisotropy of each individual fluorophore can be measured. First, MDF measurements along with MCR and PARAFAC decomposition methods will be discussed. Having defined these methods, we will show the importance of being able to remove accurately the Rayleigh scatter from EEM measurements, that otherwise contaminate the *aniso*-MDF spectra and bias the modeling of fluorophore emission<sup>221</sup>. To solve this, an algorithm developed by Rinnan *et al.*<sup>222</sup> was modified slightly to also address the problem of second order Rayleigh scatter. The work presented in this chapter is important, as we address the critical issue of standardizing the chemometric methods required to accurately extract spectral and anisotropy information from mixtures of fluorophores. The sample systems studied must be relatively simple to provide a robust way to validate the spectral measurement and chemometric data analysis elements of ARMES. The first standard sample set is composed of various perylene concentrations in glycerol. The method is then extended to a more complicated case, where compounds having very similar MDF spectra (EB and PB) and are mixed using various concentrations. This more closely replicates the situation encountered in proteins where Trp and Tyr emission are heavily overlapped.

### 4.1.1 MCR decomposition

Multivariate Curve Resolution (MCR) is in principle a soft mathematical curve resolution technique that can resolve signals from an unknown mixture, without any *a priori* information about its nature/composition<sup>143,223</sup>. This technique is a factor-based method, of which Principle Components Analysis (PCA) is probably the most well-known example.

PCA is a method widely used in analytical chemistry but also in many other fields, as it offers the possibility to better interpret complex data which have many inter-related variables. The principle of PCA is to reduce the dimensionality of the data into a smaller number of uncorrelated principal components (PCs), while minimising the loss in information<sup>224-226</sup>. The loadings are linear combinations of the original variables, constructed so that they are orthogonal to each other and account successively towards maximising the variance<sup>224,227</sup>. They form the axis of this lower dimensional space, the scores representing the coordinates of the samples in this new space.

PCA components can however be hard to interpret as they are linear combinations of pure analytes or physical effects affecting the spectra. In contrast, MCR decomposition searches for pure components that have a real chemical or physical meaning<sup>142,228-230</sup>. The primary objective of undertaking MCR is to extract meaningful concentration and spectral profiles, for each component/species contributing to the measured spectrum.

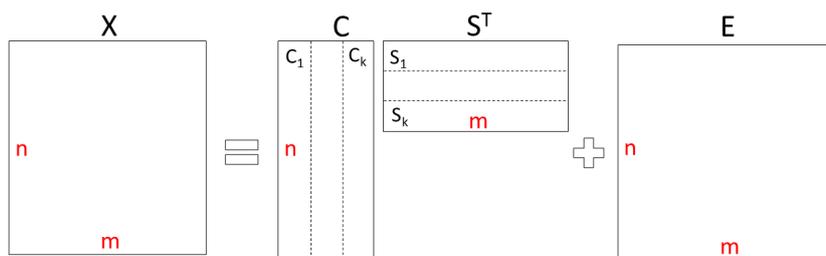
For a two-way data matrix  $X$  of dimension  $n \times m$ , containing the data of the unknown mixture, bilinear MCR resolution is given by:

**Equation 38**  **$X = CS^T + E$**

Where  $C$  ( $n \times k$ ) contains the concentration profiles (quantitative information),  $k$  being the number of measurable profiles present in the unknown mixture, and  $S$  ( $k \times m$ ) the matrix of spectral contributions (qualitative information).  $E$  ( $n \times m$ ) is an error matrix, of the same dimensions as  $X$ , and represents the part of  $X$  not explained by MCR (residuals). One criterion for good resolution is to have  $E$  as small as possible and not containing any valuable chemical information.

An illustration of the MCR decomposition of a two-way data matrix is shown in Figure 23 below.

4. Accurate anisotropy recovery from fluorophore mixtures using MCR

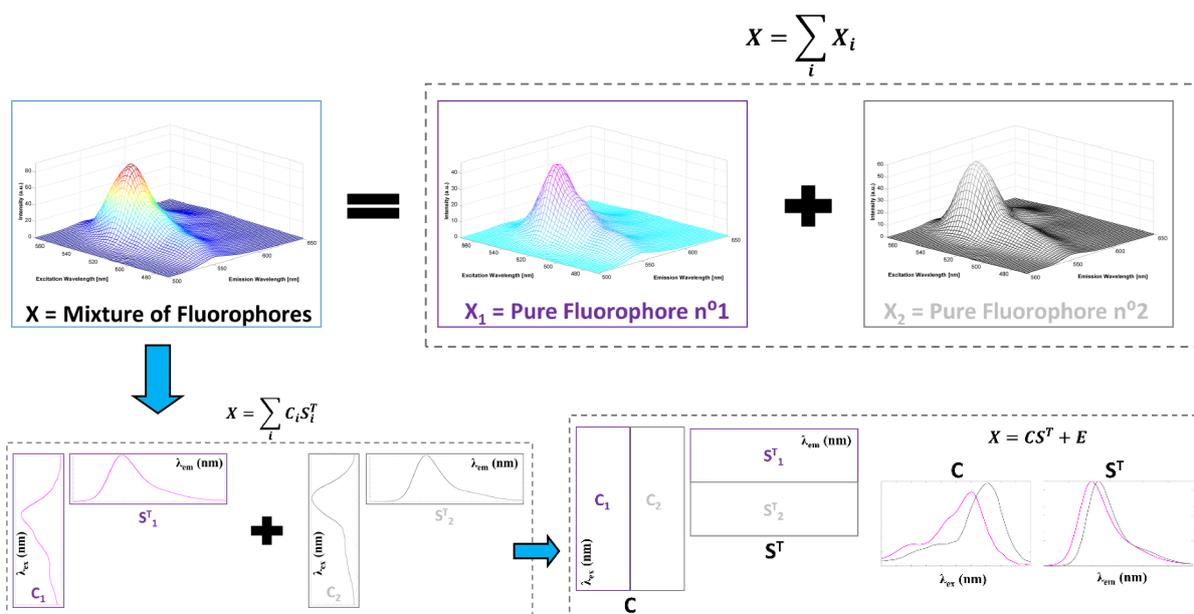


**Figure 23: Illustration of two-way data matrix (X) decomposition by MCR.**

Each pure component  $X_i$  corresponds to the product of  $C_i \times S_i$ , and the initial matrix  $X$  can be divided into the sum of these individual components as follows:

**Equation 39**  $X = \sum_{i=1}^k c_i s_i^T + E = \sum_{i=1}^k X_i + E$

MCR was developed originally as a mathematical method to resolve the shapes of two overlapping functions<sup>231</sup>, and it rapidly found applications in resolving data having an “evolutionary character” such as HPLC data<sup>232-234</sup>. This example became widely used to explain or compare the principle of MCR with other methods<sup>235-236</sup>, and some good illustrations can be found in the literature<sup>237</sup> that show clearly the different and equivalent ways to express MCR decomposition. For multidimensional fluorescence data, the principle is exactly the same except that the components that are extracted correspond to pure fluorophore signals (in theory) or group of fluorophores<sup>133,141</sup> instead of eluted molecules (for HPLC<sup>232</sup>), or absorbance spectra in UV-Vis<sup>238-239</sup>.



**Figure 24: Bilinear model obtained from MCR for a fluorescence data set, plotting (top) the sum of pure fluorophores, (bottom left) the sum of pure concentration profiles and spectra and (bottom right) the product of matrices of pure concentration profiles and spectra.**

### 4.1.2 MCR: a bilinear model

MCR as well as PCA, are bilinear modelling methods<sup>240-241</sup> for which a general equation is given in Equation 38. Bilinear instead of linear is employed when the product of matrices is used (or vector based objects), and a bilinear model assumes that each element of X can be re-obtained by a linear combination of all the elements of C and S<sup>T</sup>.

Spectroscopic or chromatographic data obey (to a certain extent) the well-known Beer-Lambert Law<sup>242</sup>, which can also be expressed as a bilinear combination of pure contributions ( $\epsilon_k$ ,  $C_k$ ) as shown in Equation 40. MCR or other bilinear models are thus particularly suited for analysing data such as UV-Vis data<sup>238-239</sup>, as they share the same underlying structure<sup>237</sup>. As shown in Figure 24, fluorescence data may also be suitable for MCR analysis but only under certain restricted conditions. Without mentioning the case of TSFS measurement for which the complex data structure induces non-bilinearity<sup>243</sup> (see section 4.4.3.2 for more details), EEM can also deviate from total bilinear or trilinear structure which complicates the application of curve resolution methods. Such deviations can be observed if EEM data suffer severe IFE or scatter issues<sup>65</sup>, but also if there is presence of FRET which is we think the key parameter that really complicates the analysis of protein EEM data. This is mentioned in a little more detail in section 4.1.8 and is further developed in chapter 5. Nevertheless MCR provides for any dataset a bilinear approximation/description of the data variation<sup>244</sup>.

#### Equation 40

$$\mathbf{A} = \boldsymbol{\epsilon} \times \mathbf{I} \times \mathbf{C} = \sum_{k=1}^k \boldsymbol{\epsilon}_k \times \mathbf{I} \times \mathbf{C}_k$$

This equation shows the linear relationship between the absorbance of a solution (A), and the concentration of the species present in solution ( $C_k$ , for k species), their molecular attenuation coefficient ( $\epsilon_k$ ) and the optical pathlength (I). This law is additive, same as MCR models for which the resolution is generally facilitated if the data presents a sequential order in at least in one dimension (*e.g.* elution times in HPLC<sup>232</sup> or component spectra in UV-Vis<sup>238</sup>). EEM, at least for well-behaved samples, is an example of “second-order” data as the sequential order is obtained in two dimensions (excitation and emission spectra).

With the progress made during recent years, MCR is now applied successfully to image analysis where no sequential order is needed, along any dimension<sup>245-246</sup>. This field is in constant expansion and interesting applications already emerged. Zhang *et al.*<sup>247</sup> for example were able to analyse the different constituents of dark and white chocolate, using MCR to resolve Raman and IR hyperspectral images. Amigo *et al.*<sup>248</sup> used MCR and other chemometric

methods to monitor qualitatively and quantitatively the blending process of two components (ibuprofen and starch), by resolving NIR hyperspectral images.

### 4.1.3 Iterative and non-iterative approaches

MCR resolution can be divided in two categories: iterative and non-iterative approaches. In most non-iterative methods, the pure component profiles are recovered by combining information obtained from subspaces of the dataset. These subspaces are carefully chosen due to the presence or absence of a particular component. Generally, only one matrix (C or S) is recovered, and the corresponding S or C is obtained through a single least-square step. The oldest and most well-known methods used are Heuristic Evolving Latent Projections (HELP)<sup>249</sup>, Window Factor Analysis (WFA)<sup>250</sup> and Subwindow Factor Analysis (SFA)<sup>251</sup>. These methods were developed when MCR was limited to sequential structured data, as the order obtained along the concentration direction permitted one to define accurately the window of existence<sup>9</sup> for each of the components<sup>143</sup>. For a dataset fulfilling this requirement, correct/unique solutions are obtained. However, due to the increasing complexity and size of spectral and imaging data now being produced, these methods are no longer suitable (sequential order being often absent).

Iterative methods are more commonly used now due to their greater flexibility in dealing with complex data structures. This family of methods begins by estimating or guessing initial estimates for C and  $S^T$ , before refining these profiles under a certain set of constraints through repeated alternative least square optimization steps. This process stops when convergence is reached, or if the maximum number of iteration is exceeded<sup>252</sup>. The most well-known methods in that category are Iterative Target Transformation Factor Analysis (ITTFA)<sup>253-254</sup>, and Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS)<sup>228,255-256</sup>. ITTFA starts by calculating a PCA model of the unknown dataset X:

**Equation 41**  $\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$

There is a strong analogy between the scores (T) and loadings ( $P^T$ ) matrices of PCA model and C and  $S^T$  from MCR resolution. The only difference lies in the constraint of orthogonality imposed on T and  $P^T$  in PCA analyses<sup>224</sup>. As PCA and MCR decompose the same space, T and  $P^T$  are highly related to C and  $S^T$  respectively<sup>257</sup> and a target transformation matrix (R)<sup>258</sup> can be created to transform C in T such that:

---

<sup>9</sup> The window of existence, or concentration window, correspond to a subspace of the original data matrix where compounds are present or absent.

**Equation 42**

$$\mathbf{C} = \mathbf{TR}$$

An initial estimate of the concentration profile of the  $i^{\text{th}}$  component ( $c_i^0$ ) is then used to calculate the  $i^{\text{th}}$  column of R ( $r_i$ ) using the following equation:

**Equation 43**

$$\mathbf{r}_i = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{c}_i$$

By substituting the value of  $r_i^0$  into Equation 42, a new  $c_i^1$  is obtained. This process is repeated  $k$  times until a criterion of convergence is obtained such as:

**Equation 44**

$$\|\mathbf{c}_i^k - \mathbf{c}_i^{k-1}\| < \text{constant}$$

Whereas ITTFA performs first the optimization of C to obtain S from the model through a least square step, MCR-ALS optimizes both at the same time in each iterative cycle. The aim of ALS optimization is to solve, under a chosen set of constraints, the following equations<sup>257</sup>:

**Equation 45**

$$\min_{\hat{\mathbf{C}}} \|\hat{\mathbf{X}}_{\text{PCA}} - \hat{\mathbf{C}} \hat{\mathbf{S}}^T\|$$

**Equation 46**

$$\min_{\hat{\mathbf{S}}^T} \|\hat{\mathbf{X}}_{\text{PCA}} - \hat{\mathbf{C}} \hat{\mathbf{S}}^T\|$$

In these equations,  $\hat{\mathbf{X}}_{\text{PCA}}$  represents the PCA-reproduced data of the original unknown matrix X.  $\hat{\mathbf{C}}$  and  $\hat{\mathbf{S}}^T$  are the least-squares approximations of C and  $\mathbf{S}^T$  respectively. The idea is then to minimize the norm of residuals between PCA and the ALS reproduced data, by keeping  $\hat{\mathbf{C}}$  (Equation 45) or  $\hat{\mathbf{S}}^T$  (Equation 46) constant.

The least squares solution of Equation 45 can then be written as:

**Equation 47**

$$\hat{\mathbf{S}}^T = (\hat{\mathbf{C}}^T \hat{\mathbf{C}})^{-1} \hat{\mathbf{C}}^T \mathbf{X}_{\text{PCA}}$$

The equivalent solution for Equation 46 gives:

**Equation 48**

$$\hat{\mathbf{C}} = \mathbf{X}_{\text{PCA}} \hat{\mathbf{S}} (\hat{\mathbf{S}}^T \hat{\mathbf{S}})^{-1}$$

For each iteration, C is used to obtain S, and then S is used to obtain a new estimate of C by solving, repetitively, Equation 45 and Equation 46 in a sequential manner (*i.e.* “alternative”). This presupposes that an initial estimate of C had to be provided to start the iteration process.

Once the optimization step is complete, it is possible to assess the quality of the MCR-ALS model by calculating two parameters, called  $R^2$  (percentage of variance explained) and LOF (lack of fit):

**Equation 49**

$$R^2 = \frac{\sum_{ij} x_{ij} - \sum_{ij} e_{ij}}{\sum_{ij} x_{ij}}$$

**Equation 50**

$$\text{LOF (\%)} = 100 \times \sqrt{\frac{\sum_{ij} e_{ij}}{\sum_{ij} x_{ij}}}$$

In these equations,  $x_{ij}$  and  $e_{ij}$  represent the elements from  $X$  and  $E$ . These parameters allow the comparison of one model to another, which is interesting in cases where different sets of initial estimates or constraints are tested.

#### 4.1.4 Importance of initial estimates for MCR decomposition

The initial estimates that are used in MCR modelling are essential, and they need to be chosen “wisely” in order to “guide” the correct resolution. Either concentration or pure spectra profiles can be used, as long as they are not random and obey the same constraints as those used in the model<sup>143,259</sup>. If some *a priori* knowledge of the chemical species is available (*i.e.* their pure spectra), then this can be used. However, in most cases, this information is not available and one needs to develop another way to obtain sensible estimates. The most well-known methods are called Evolving Factor Analysis (EFA)<sup>260</sup> and Simple-to-use self-modelling analysis (SIMPLISMA)<sup>261</sup>. The difference between these methods is once again based on requirement for an ordered structure along the concentration direction.

EFA is a local rank method, and as such its goal is to find and describe each component in a system, by performing multiple rank analysis on limited portions of the dataset which are called windows. This method works by performing PCA on submatrices, where the size increases gradually, to obtain components that emerge along with the evolving process that is being studied. EFA needs ordered structure in the concentration direction, and was originally applied to HPLC diode-array data<sup>262</sup>. It can provide good initial estimates for MCR if the system studied fulfils that condition<sup>†</sup>.

SIMPLISMA is a type of pure variable selection method, and as such does not require any order along concentration direction<sup>237</sup>. It only requires data that are strictly positive, which is a condition fulfilled by fluorescence and UV-Vis data. SIMPLISMA looks for the most dissimilar rows and columns in the original dataset, and selects them to build the pure components. This method is more versatile than EFA as it works even if sequential order is

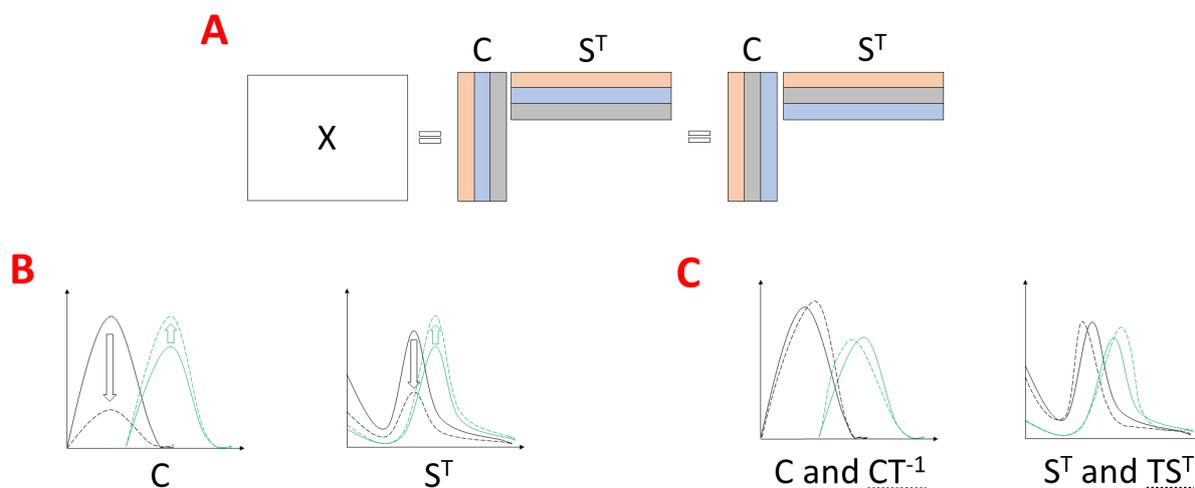
---

<sup>†</sup> A variant of EFA, based on fixed-size window (FSMW-EFA)<sup>263</sup> has also been developed, but focuses more on the local-rank information rather than describing the evolution of the compounds in the process.

missing, and is often used to initialize MCR algorithm for resolving multidimensional fluorescence measurements<sup>264-266</sup>.

### 4.1.5 Ambiguities in MCR

The big issue associated with MCR decomposition are the ambiguities present in the solution. The ambiguities are expressed by the fact that the  $C/S^T$  solutions resolved in MCR are often non-unique. A range of  $C/S^T$  combinations exist, which can be used to model the original data matrix with same quality of fit. As suggested by Tauler<sup>240</sup> and de Juan<sup>237</sup>, there are three kinds of ambiguities observed in MCR resolution (see Figure 25):



**Figure 25: This illustration represents the: A) Permutation B) Intensity C) Rotational ambiguities that are encountered during MCR-ALS deconvolution<sup>240</sup>.**

**Permutation ambiguities:** during modelling, the pure components resolved in both  $C$  and  $S^T$  (Figure 25 A) do not respect a strict order. A component resolved in 2<sup>nd</sup> position for both  $C/S^T$  could be 3<sup>rd</sup> in next model, which do not change the MCR result as the sum of all components is identical. The solution to permutation ambiguities is to check manually that for each  $i^{\text{th}}$  component,  $c_i$  correspond to  $s_i^T$ . If that is not the case, a Procrustes rotation or a manual permutation needs to be performed to assemble the correct pairs together.

**Intensity ambiguities:** these are related to all the  $C/S^T$  pairs that fit the original matrix  $X$  equally well, but with different scaling ratios (Figure 25 B). This property corresponds mathematically to the following equation:

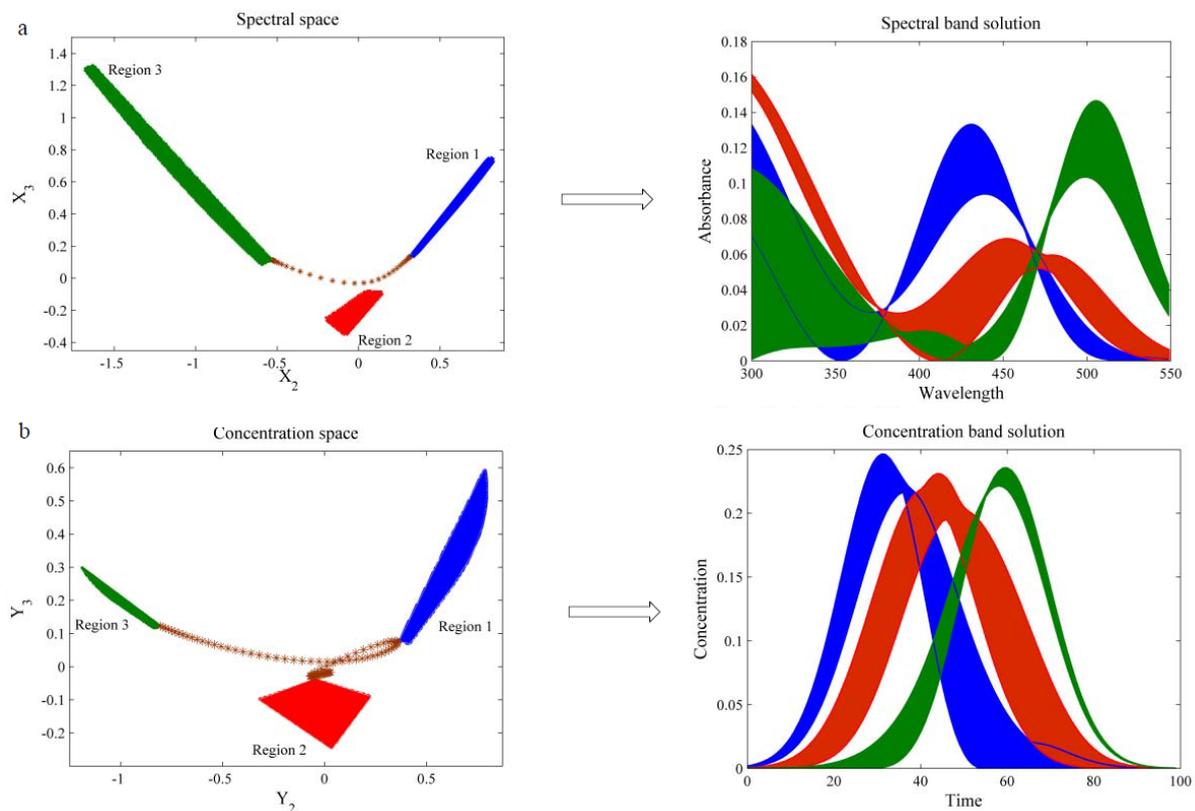
$$\text{Equation 51} \quad \mathbf{X} = \sum_{i=1}^k \mathbf{c}_i \mathbf{s}_i^T + \mathbf{E} = \sum_{i=1}^k (\mathbf{c}_i z_i) \left( \mathbf{s}_i^T \frac{1}{z_i} \right) + \mathbf{E}$$

In this equation,  $z_i$  is a scaling factor which can be suppressed by applying a normalization constraint<sup>237</sup> during the ALS optimization on either  $C$  or  $S^T$ .

**Rotational ambiguities:** a difference in shape of the components in both  $C$  and  $S^T$  can also produce ambiguities. Compared to intensity ambiguities that are an issue only for quantitative analysis, rotational ambiguities are more complex and affect both qualitative and quantitative analysis<sup>228</sup>:

$$\text{Equation 52} \quad \mathbf{X} = \mathbf{CS}^T + \mathbf{E} = (\mathbf{CT}^{-1})(\mathbf{TS}^T) + \mathbf{E} = \mathbf{C}_{\text{new}}\mathbf{S}_{\text{new}}^T + \mathbf{E}$$

In this equation,  $T$  can be any transformation matrix, which generates an infinity of possible solutions to the deconvolution. The concept of rotational ambiguities and its impact on the deconvolution of spectroscopic data can be hard to comprehend. Some geometrical methods, limited at the moment to a mixture of maximum four components, have been developed and have the advantage to show clearly the area of feasible solutions (Figure 26) for MCR decomposition<sup>267-268</sup>.



**Figure 26: Obtained feasible solutions (a: spectral solutions, b: concentration solutions) under non-negativity constraint for HPLC dataset of 3 components mixture. In both cases (a, b) left panels show the obtained areas of feasible solutions and right panels display the translated profiles of feasible regions. Reproduced from de Beyramysoltan *et al.*<sup>269</sup>, with permission of Elsevier.**

In this example, Beyramysoltan *et al.*<sup>269</sup> used a HPLC dataset (completely noise free) to show all the range of solutions for a model constrained only by non-negativity. By translating

the area to spectral information, Figure 26 clarifies why rotational ambiguities are a big issue in MCR analysis. This problem needs to be addressed if one wants to trust the solution obtained, as the number of possible combinations that equally fit the original matrix is infinite. The situation encountered with EEM measurements is exactly the same, if we replace the elution time obtained in the concentration space by excitation wavelength.

One method is particularly suitable to check for the extent of rotational ambiguities for multi-component data, called MCR-BANDS<sup>223,270</sup>. This method calculates, under a certain set of constraints, the maximum and minimum value of a scaling vector ( $f_n$ ) defined as:

**Equation 53** 
$$\mathbf{f}_n = \frac{\|\mathbf{c}_i \mathbf{s}_i^T\|}{\|\mathbf{C} \mathbf{S}^T\|}$$

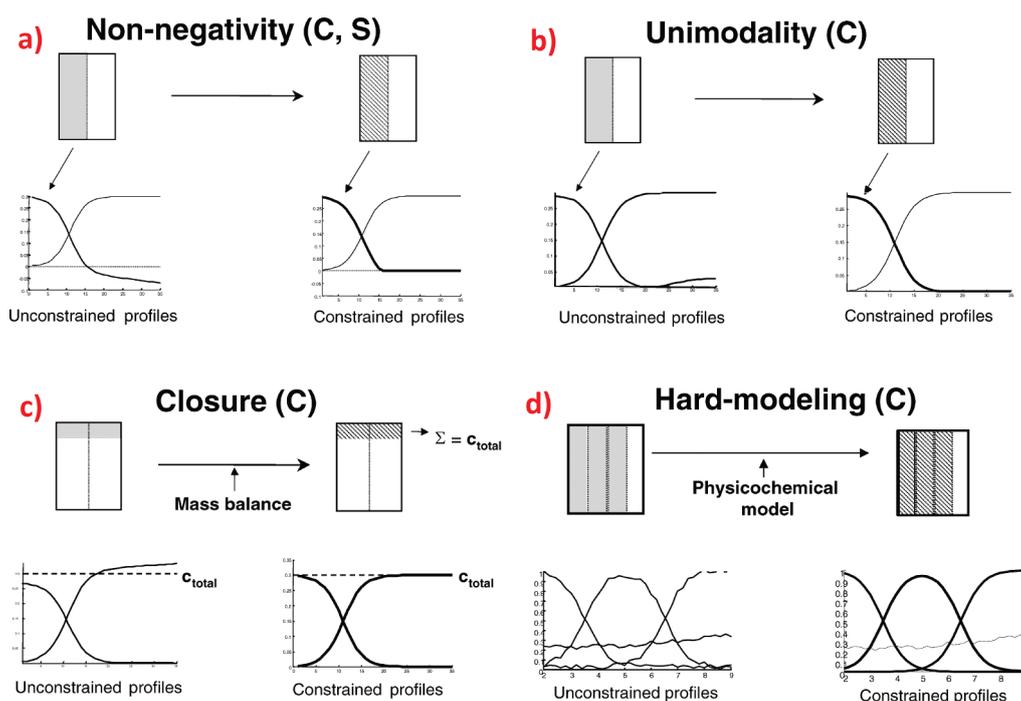
$f_n$  is the ratio of the frobenius<sup>s</sup> norm of the  $i^{\text{th}}$  pure component in the total mixture and represents the relative contribution of that  $i^{\text{th}}$  component in the whole recovered signal. This method calculates under a selected set of constraints, an initial  $f_n$  value for each component ( $f_n^{\text{init}}$ ). A non-linear optimization is then used in order to generate a  $f_n^{\text{min}}$  and  $f_n^{\text{max}}$  value. The idea is that the more rotational ambiguities there are, the more the relative contribution of each component will vary during the optimization due to its uncertainty. By comparing the degree of variation in the contribution of each component ( $f_i^{\text{max}} - f_i^{\text{min}}$ ), one can obtain a good estimate of the extent of rotational ambiguity for each component<sup>t</sup>.

#### 4.1.6 Constraints and advantages of MCR-ALS

Once the degree of ambiguity has been estimated, the goal is to try to reduce them in order to get, if possible, a unique solution. This is where the ALS variant of MCR is probably one of the most useful algorithms for curve resolution to date, as it offers the possibility to reduce ambiguity by adding constraints during the optimization of  $\mathbf{C}$  and  $\mathbf{S}^T$ <sup>143,228,271</sup>. A constraint can be defined as a physico-chemical or mathematical property, that is fulfilled by the entire system studied or by some of its pure components<sup>272</sup> (Figure 27).

<sup>s</sup> The frobenius norm of a matrix represents the square root of the sum of the absolute squares of all its elements.

<sup>t</sup> For a component resolved without any ambiguity,  $f_i^{\text{max}} - f_i^{\text{min}} = 0$ .



**Figure 27: Common constraints used in iterative MCR deconvolution. Reproduced from de Juan et al.<sup>143</sup>, with permission of Taylor & Francis Group.**

The most common physico-chemical constraints (Figure 27 a,b,c) are non-negativity, unimodality and closure. Non-negativity<sup>273</sup> is applied to a dataset whose values should always be positive. This constraint particularly suits UV-Vis and fluorescence data and can be applied to both concentration and pure spectra profiles. In the case of unimodality<sup>274</sup>, only the concentration profiles are constrained to show a single maximum per profile. This is suitable for systems following the gradual appearance or disappearance of a specific compound<sup>257</sup>, acid-base titrations being a typical example<sup>275</sup>. Closure is also applied to concentration profiles, but is specific to closed reaction systems where mass is conserved. Considering all species present during the reaction, it forces the sum of their concentrations to be equal to a constant value at each stage of the process. An example of such a system where closure can be applied is the monitoring of protein folding, using circular dichroism<sup>276</sup> and fluorescence spectroscopy<sup>277</sup>. Another physico-chemical constraint not shown in Figure 27, consists of using during the optimization, pure response profiles (C or  $S^T$ ) obtained from standards or reference materials prior to modelling<sup>237,278</sup>.

Hard modelling (Figure 27 d) differs a lot from the constraints presented above, and refers to the implementation of a physicochemical model during the MCR optimization step that forces the concentration profiles extracted from spectroscopic data to fit a suitable kinetic or thermodynamic law<sup>279</sup>. This hybrid combination of hard constraint with soft-modelling method like MCR is a very powerful tool as it allows the resolution of complex systems (*i.e.*

kinetic<sup>280</sup>, photodegradation<sup>281</sup>, or binding<sup>282</sup> studies), where soft or hard modelling alone could not succeed<sup>283</sup>. Diewok *et al.*<sup>284</sup> added for example an acid-base titration law to the soft MCR-ALS modelling of FTIR data of pH modulated aqueous solutions, to resolve unambiguously analyte profiles that presented highly correlated concentration profiles. Manouchehri *et al.*<sup>282</sup> investigated the binding of BSA with vitamin B6 (VB6), using UV-Vis, fluorescence and FTIR measurements. While simultaneous analysis of fluorescence and UV data by MCR allowed the authors to resolve the contribution of pure compounds and complex species (BSA-VP6) and get an indication about the protein-ligand stoichiometry, hard modelling with mass balance equation was necessary to fit the concentration profiles more accurately and obtain a good estimate of the binding constant between BSA-VB6.

Apart from physico-chemical constraints, mathematical constraints are also available. These constraints are powerful and are associated with the concept of local rank in a matrix, which determines, by performing repeated PCA in fixed or moving windows (*i.e.* EFA), the number and distribution of components within the data matrix. These constraints are generally applied on concentration profiles and can identify the presence or absence of components in various spectral regions. The term selectivity is employed if only one component is present, otherwise the correct terminology is local rank constraint<sup>228,256</sup>. Also, when MCR is extended to multi-way data with analysis of several samples (see below), a similar constraint called “correspondence among species”<sup>240</sup> can be used to inform the model about the presence/absence of species in each sample. The application of this kind of constraint is very important and can significantly improve the quality of MCR resolution, because ambiguities are seriously decreased or sometimes completely removed<sup>228</sup>.

#### 4.1.7 Extension of MCR to multi-way data

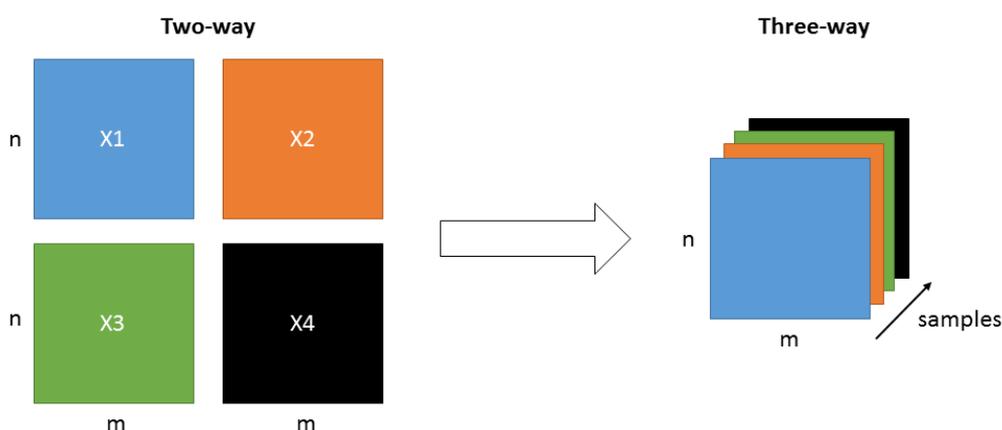
MCR was originally intended for the resolution of two-way datasets  $X (n \times m)^u$ , however three-way datasets or even higher order datasets are now very common<sup>237,285</sup>. The main advantage is to gather and combine in the same analysis, different information that describe the same system, making the resolution of complex cases much easier and less ambiguous with MCR<sup>279</sup>. Monitoring protein structural changes due to thermal or chemical denaturation is an example of complex process that cannot be resolved by a simple 2-D measurement. However, Navea *et al.* proved that combining several different spectroscopic techniques, such as mid-IR and near-IR<sup>286</sup>, or 2-D fluorescence with circular dichroism (CD) and UV-Vis

---

<sup>u</sup> The collection of “n” UV-Vis spectra is a typical example of two-way data, as the absorbance is measured “n” times along “m” wavelengths, creating a “n×m” matrix (X) of numbers.

spectroscopies<sup>277,287</sup> in the same MCR model, enabled the resolution of structural transitions and intermediate species which otherwise were undetectable. In fact, extending MCR to multi-way data enables the breaking of rank deficiency usually encountered in these complex systems<sup>279</sup>. The concept of rank deficiency is explained in more details in the next chapter (section 5.8) when we encountered this issue during the analysis of insulin aggregation by ARMES.

Nevertheless, the conclusion is that the extensive description of complex processes in chemistry via spectroscopic methods requires richer data structures. This is obtained by merging various two-way X matrices<sup>283</sup>, generated either from different analytical instruments (UV-Vis, fluorescence, FTIR, Raman, etc.), or from the same measurement but applied to different process runs (varying in pH, temperature, concentration, etc.)<sup>288</sup>. Those X matrices (called “samples” in chemometrics), once stacked together, form what is called a three-way data structure (Figure 28).



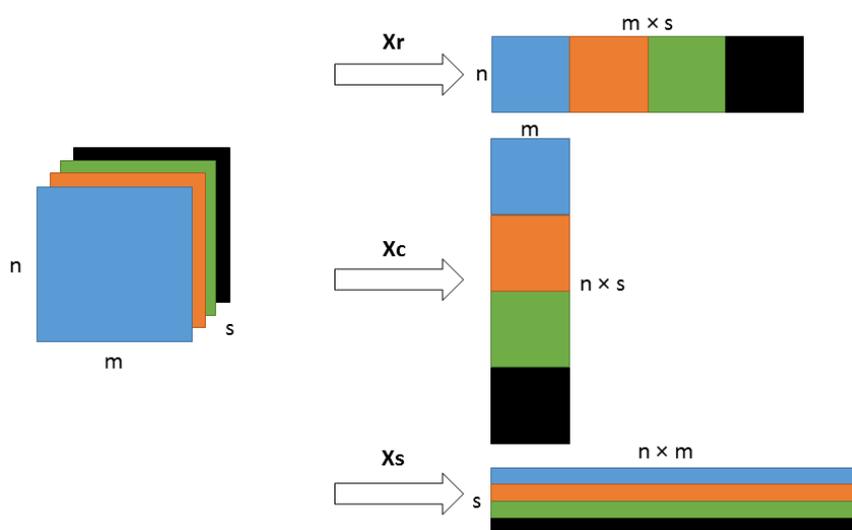
**Figure 28: Schematic representation of two-way and three-way data.**

In this example, four samples (X1-X4) of “n” measurements were collected along “m” wavelengths to generate a final cuboid of dimension ( $n \times m \times \text{samples}$ ). MDF measurements are a typical example of three-way data<sup>289</sup>, where “n” represents the different excitation wavelengths used to collect the fluorescence intensity along “m” emission wavelengths. Each MDF measurement (or “sample”) is represented then by a single X matrix.

However, to deal with a three-way or even higher order datasets, a transformation is necessary because MCR-ALS only deals with two-way datasets (bilinear model). The solution lies in what is called the “unfolding” of the data cube, to generate a bilinear, augmented two-way matrix<sup>256</sup> (Figure 28). The augmented matrix (also called multiset structure) can then be resolved by MCR-ALS exactly as a single second-order data matrix. A cube of data sized ( $n \times$

$m \times s$ ) can be unfolded along each of its directions (also called modes), giving a row-wise augmented (RWA) matrix  $X_r$  ( $n \times ms$ ), column-wise augmented (CWA) matrix  $X_c$  ( $m \times ns$ ), or tube-wise augmented matrix  $X_t$  ( $s \times nm$ ) also called sample-wise augmentation (SWA). The different types of augmentation are illustrated in Figure 29.

According to the bilinear expression of MCR (Equation 38), in RWA (also called data fusion) the concentration ( $C = n$ ) is fixed while the spectra ( $S^T = m \times s$ ) are varying across the samples. This is particularly suited to situations where the same chemical system is monitored by various analytical instruments, as all the methods measure the same chemistry *i.e.* the “concentration” ( $n$ ) is shared in all the measurements/slices<sup>279,290</sup>. Investigating protein structural transitions via multiple spectroscopies as mentioned above was an example of RWA<sup>276-277,287</sup>.



**Figure 29: Schematic representation of different ways to unfold a data cube. Each “sample” or two-way  $X$  matrix that forms the data cube is characterised by a colour, and each  $X$  matrix could represent a single MDF measurement.**

In CWA ( $X_c$ ), it is the concentration ( $C = n \times s$ ) that varies while the spectra ( $S^T = m$ ) are fixed. This is applied to situations where the same spectroscopic technique (*e.g.* fluorescence) is used to monitor the changes in a system subject to parametric changes (*i.e.* change in temperature, ionic strength, pH). The rationale is that this time the chemistry is varying *i.e.* the “concentration” of the various component evolves, but the pure spectra of the components ( $m$ ) are found in all the slices. This is why CWA is the method of choice when a chemical system is followed by *e.g.* EEM measurements. In this case the fixed spectral mode  $S^T$  is generally replaced by the emission, while the augmented mode ( $n \times s$ ) is formed by the product of (excitation  $\times$  samples). One part of the “loadings” or pure spectral information

characterising the resolved components is thus obtained in  $S^T$  (emission). The other part (excitation) is contained in the pure concentration information, also called “scores” in chemometrics, that reveal how the different components are evolving in the process studied. Because one concentration-weighted-excitation spectrum is resolved per component and sample, summing all the variables for each spectrum gives the pure scores information. In terms of loadings, we chose to show in this chapter only the normalized-averaged excitation and normalized emission spectra (see section 4.4.3.3 for more details).

SWA ( $X_s$ ), as opposed to the two previous augmentations, is almost never used in MCR. This type of augmentation arranges the data so that all the spectral information is in  $S^T$  ( $n \times m$ ) and is fixed, while  $C$  ( $= s$ ) varies across the samples. The samples share therefore the same  $X$  matrix (or MDF spectrum) meaning that SWA do not enable shape variations across samples. This type of augmentation was useful for us in a very special case related to the complex data structure of TSFS measurements, but more details are given in section 4.4.3.2.

#### 4.1.8 Concept of trilinearity and multilinear constraints

The rank of a matrix corresponds to the quantity of vectors (eigenvectors of principal components) that are necessary to span the row and column space of that matrix<sup>257</sup>. The concept of trilinearity is highly related to this concept of rank. More precisely, a data cube can be considered trilinear if the rank of the augmented matrix is the same, regardless of the mode used in the augmentation. Therefore, the data cube is “non-trilinear” if the rank changes.

If  $R_x$ ,  $C_x$ , and  $T_x$  represent respectively the resolved data cube  $X$  after row-wise, column-wise, and tube-wise augmentation, a trilinear system means that an equal number of pure profiles “ $p$ ” are retrieved for  $R_x$ ,  $C_x$ , and  $T_x$ . In that case, each element of  $X$  ( $x_{ijk}$ ) can be obtained using the following equation<sup>257</sup>:

**Equation 54** 
$$x_{ijk} = \sum_{f=1}^p (r_{x_{if}})(c_{x_{jf}})(t_{x_{kf}})$$

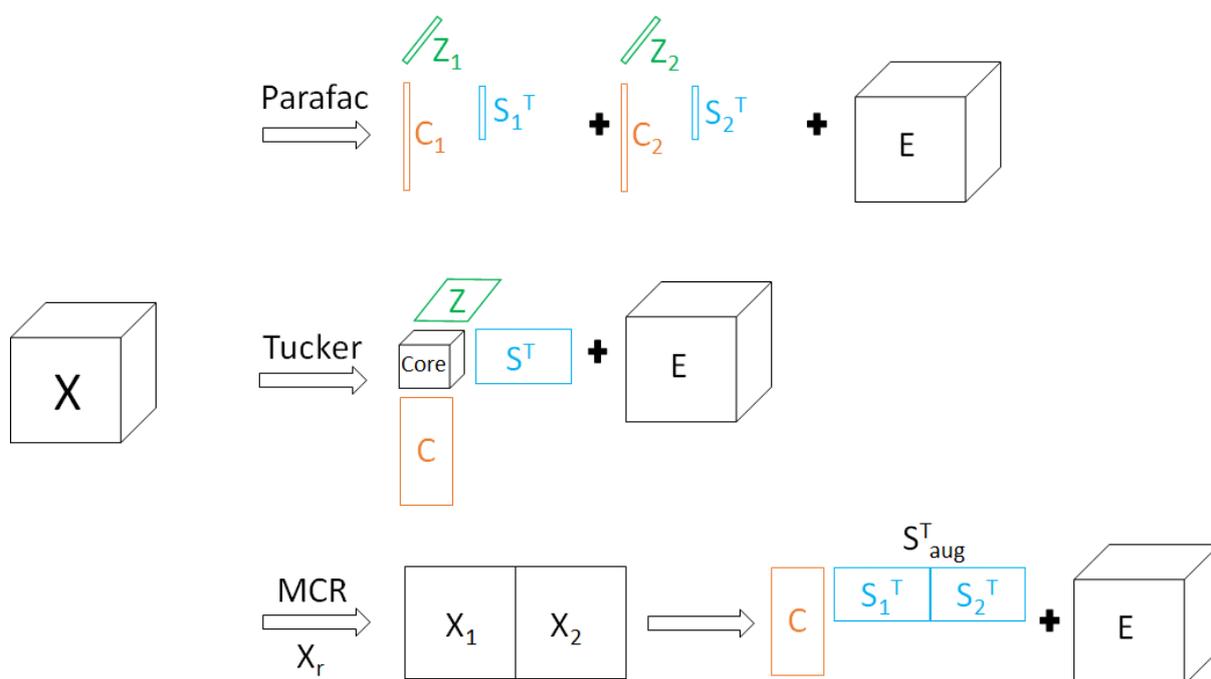
This equation represents the principle behind the PARAFAC model<sup>146</sup>, a deconvolution method capable of resolving trilinear datasets.

However, if the data cube  $X$  is non trilinear, then the number of pure profiles is different for  $R_x$ ,  $C_x$  and  $T_x$ . Naming them respectively “ $rp$ ”, “ $cp$ ”, and “ $tp$ ”,  $x_{ijk}$  can be obtained using the modified equation:

**Equation 55** 
$$x_{ijk} = \sum_{f=1}^{rp} \sum_{g=1}^{cp} \sum_{h=1}^{tp} (r_{x_{if}})(c_{x_{jg}})(t_{x_{kh}})c_{fgh}$$

This equation is more complex and involves the appearance of a core matrix<sup>v</sup>  $C$ . In fact, it forms the principle behind a more flexible and general form of multilinear models called Tucker<sup>3291-292</sup> or N-way PCA, capable of handling non-trilinear data cubes. In this case, the core matrix is different from the identity matrix, as otherwise Equation 55 becomes Equation 54 showing why PARAFAC is a restricted version of Tucker<sup>3146</sup>.

PARAFAC and Tucker, as multilinear models, have an advantage compared to MCR in being able to cope directly with a data cube (Figure 30). Augmentation is not needed which means that for *e.g.* EEM data, the scores and the excitation/emission loadings are all directly obtained in separated modes.



**Figure 30: Schematic representation of PARAFAC, TUCKER, and MCR principles, on a two component mixture.**

However PARAFAC is not suited for the analysis of non-trilinear systems and Tucker models are generally hard to interpret without *a priori* knowledge and are subject to rotational ambiguities<sup>293</sup> (see section 5.8 for a more in depth discussion). With the analysis of multi-way datasets using MCR-ALS and the development of a new range of mathematical constraints called multilinear constraints<sup>294</sup>, it is now possible to reproduce the inner trilinear structure<sup>271</sup> of PARAFAC in MCR-ALS, by implementing trilinearity as an optional constraint. Trilinearity is very powerful and probably the most important of all multilinear constraints, because if the

<sup>v</sup> In the core matrix, each element  $c_{igh}$  are weighting factors in possible produces of  $rx \times cx \times tx$ .

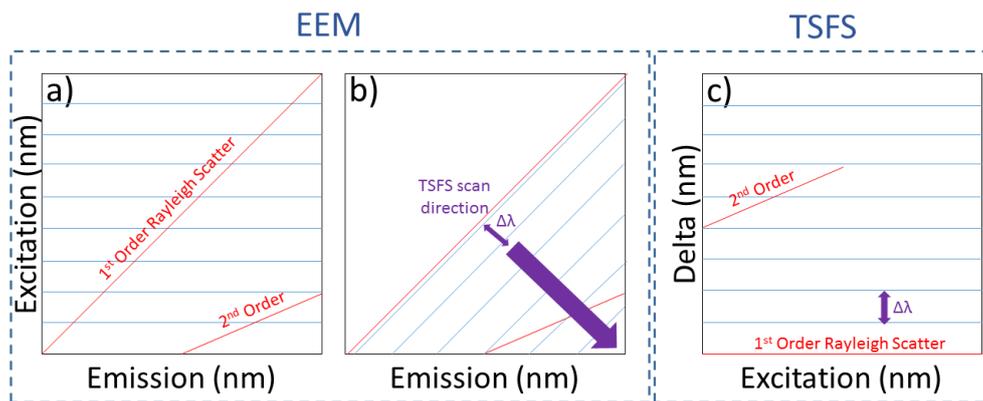
dataset fulfils the condition required for its use, this constraint ensures the resolution of pure and unambiguous profiles<sup>271,228</sup> (like in PARAFAC). The only issue is that the conditions to use this constraint are very demanding<sup>294</sup>, and ideally one should base its use on scientific principles.

The trilinear constraint in MCR-ALS forces the resolved pure profiles to behave the same in all C or S<sup>T</sup> matrices. In fact, because this constraint can be implemented separately on each species<sup>255</sup>, it allows a certain degree of flexibility in cases of components varying from pure bilinearity to complete trilinearity<sup>295</sup>. Within the trilinear case, the profiles can be a bit more constrained by synchronising them, which will correct for some shifts and force the profiles to appear at the same positions among all samples.

A typical example of a non-fully trilinear system are LC-DAD datasets, where the elution peaks can have slightly different shapes and retention times from run to run<sup>236,256</sup>. However, an example where trilinearity can sometimes be achieved is in fluorescence spectroscopy using EEM, because both excitation and emission are highly reproducible spectral orders<sup>294</sup> (the same spectral shape is obtained, at the same wavelength range, if measurement is done under the same physical conditions). This is true except for when the spectra are affected by either IFE or FRET processes. IFE<sup>59</sup> breaks the linearity between fluorescence intensity and concentration, but also drastically distorts the shape of EEM. Because this distorted shape can vary from sample to sample, the reproducibility of excitation and emission spectral orders are then highly reduced, this in turn breaks the trilinearity of EEM data<sup>289,296-298</sup>. For the same reason, some deviation from a total trilinear behaviour can also be observed in the presence of FRET<sup>299-300</sup>, which is a subject that is further developed in the next chapter (section 5.8). This non-radiative energy transfer<sup>301</sup> is particularly significant in proteins where fluorophores and chromophores are in very close proximity.

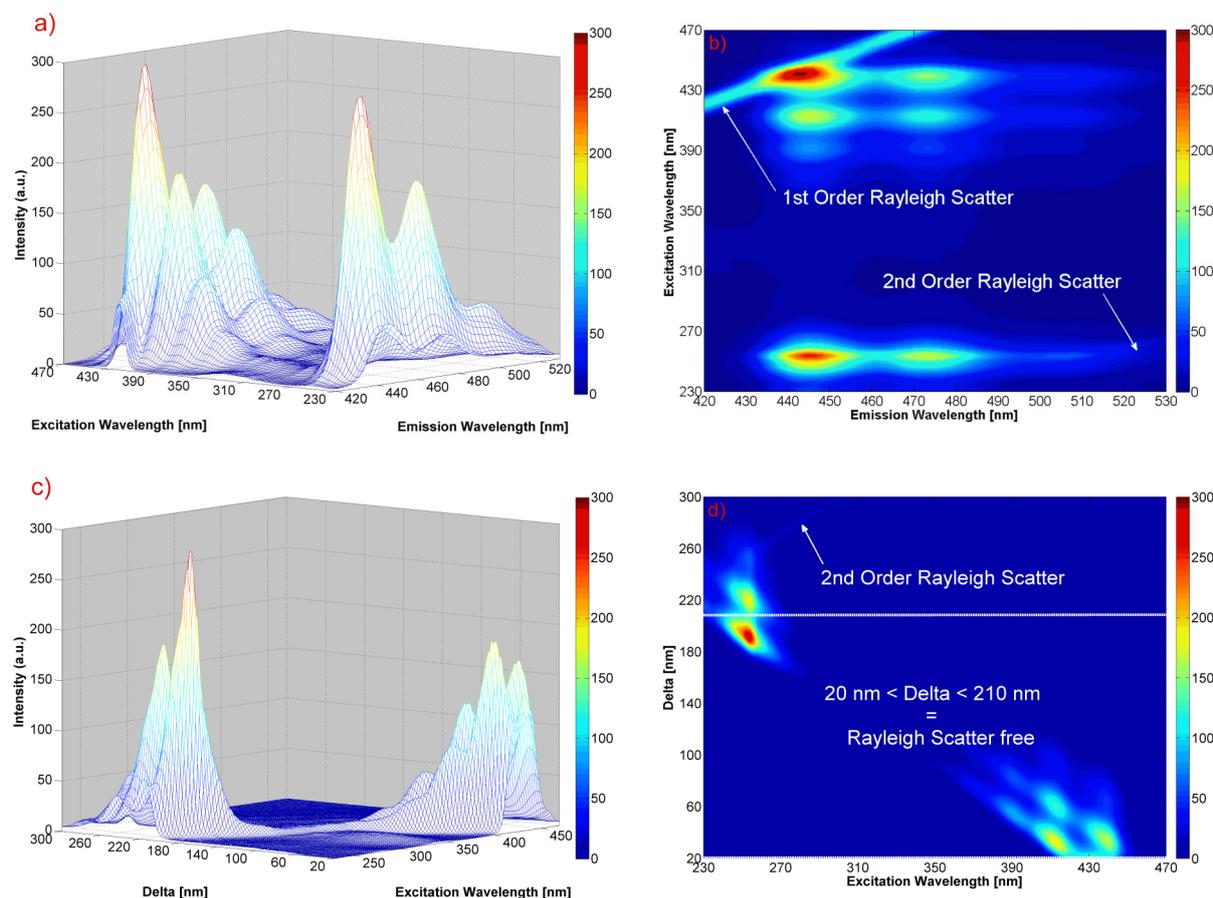
## 4.2 The EEM and TSFS measurements

A MDF spectrum is a data matrix where each element represents the fluorescence intensity at a specific  $\lambda_{\text{ex}} / \lambda_{\text{em}}$  wavelength pair. EEM and TSFS are the two main measurement methods for MDF measurements. In the case of EEM, emission spectra are collected for a range of excitation wavelengths, creating a matrix where every row or column represents an excitation or emission spectrum. TSFS is slightly different in the sense that excitation and emission are scanned simultaneously, by increasing the interval (called delta  $\Delta\lambda = \lambda_{\text{em}} - \lambda_{\text{ex}}$ ) between emission and excitation (Figure 31).



**Figure 31: Representation of a) EEM landscape b) TSFS scanning mode in EEM landscape c) TSFS landscape. In red are shown the 1<sup>st</sup> and 2<sup>nd</sup> order Rayleigh scatter.**

For an EEM landscape, the emission is generally plotted along the x-axis and excitation on the y-axis, while for TSFS the excitation is on x-axis and  $\Delta\lambda$  on y-axis (Figure 32).



**Figure 32: EEM (a,b) and TSFS (c,d) spectra of a 1  $\mu$ M solution of perylene in 100% glycerol<sup>w</sup>, shown either as landscape plots (a,c) or contour plots (b,d). Some of the key differences, such as the absence of 1<sup>st</sup> order Rayleigh scatter in TSFS are highlighted in the latter.**

<sup>w</sup> HH polarized spectra, collected using the same excitation range ( $\lambda_{\text{ex}} = 230\text{--}470$  nm) for both EEM/TSFS with  $\lambda_{\text{em}} = 420\text{--}530$  nm/ $\Delta\lambda = 20\text{--}300$  nm respectively. The data interval was 2 nm and the slits fixed to 5 nm on both excitation and emission pathlengths.

TSFS and EEM have their specific advantages and disadvantages. Apart from the non-trilinear behaviour of TSFS versus EEM data (see section 4.4.1 for a more in-depth discussion), TSFS is faster to collect (nearly by a factor of two), in contrast to EEM because it only measures the spectral region where  $\lambda_{em} \geq \lambda_{ex}$  ( $\Delta\lambda$  being greater than or equal to 0<sup>117</sup>). Indeed, for most conventional one photon excitation-based fluorescence there is no emission in the anti-Stokes region ( $\lambda_{ex} > \lambda_{em}$ ). The other advantage of TSFS is that Rayleigh scatter can be easily avoided. The 1<sup>st</sup> order Rayleigh scatter for example appears in EEM at  $\lambda_{ex} = \lambda_{em}$  which is a diagonal on the landscape, but appears as horizontal line on TSFS as this is equivalent to  $\Delta\lambda = 0$  (Figure 31). Whereas in EEM, the 1<sup>st</sup> order scatter can be avoided for each excitation wavelength by adding roughly 20 nm to the emission due to the scatter width ( $\lambda_{em} \geq \lambda_{ex} + 20$  nm), in TSFS because  $\Delta\lambda = \lambda_{em} - \lambda_{ex}$ , the scatter is eliminated by simply starting the data collection at  $\Delta\lambda \geq 20$  nm (Figure 32 c,d). Similarly, the 2<sup>nd</sup> order Rayleigh scatter in EEM is avoided if  $\lambda_{em} \leq 2 \times \lambda_{ex} - 20$  nm. In TSFS this is equivalent to  $\Delta\lambda \leq \lambda_{ex} - 20$  nm. The 2<sup>nd</sup> order Rayleigh scatter is therefore not a horizontal line in TSFS landscapes as the offset values vary according to  $\lambda_{ex}$ . In this case, if that data collection starts from e.g.  $\lambda_{ex} = 230$  nm, the 2<sup>nd</sup> order Rayleigh scatter can also be easily eliminated from TSFS landscape by fixing the upper limit at  $\Delta\lambda = 210$  nm (Figure 32 d).

### 4.3 Accurate Rayleigh scatter removal in EEM.

EEM measurements are usually contaminated by the presence of Rayleigh scatter. In the case of perylene (Figure 32 a,b), the 1<sup>st</sup> order Rayleigh band overlaps with the main vibronic band of the S<sub>1</sub> state (S<sub>1,0</sub> → S<sub>1,1</sub>), while the 2<sup>nd</sup> order Rayleigh band overlaps with the S<sub>2</sub> state. We saw in the previous chapter that perylene in 100% glycerol was an ideal 2-D anisotropy standard for ARMES, but its working range had to be reduced due to the presence of Rayleigh scatter. Because ARMES is a multidimensional measurement, perylene's adequacy can be improved by rather considering it as a multidimensional anisotropy standard.

ARMES generates a multidimensional data matrix ( $\lambda_{ex} \times \lambda_{em} \times r$ ) that we designate *aniso*-MDF spectra. These spectra are appropriate for a multidimensional anisotropy standard, but *aniso*-MDF spectra require setting a threshold to remove low intensity regions that otherwise generate erroneous anisotropy values (see section 7.4 in Appendices for more information).

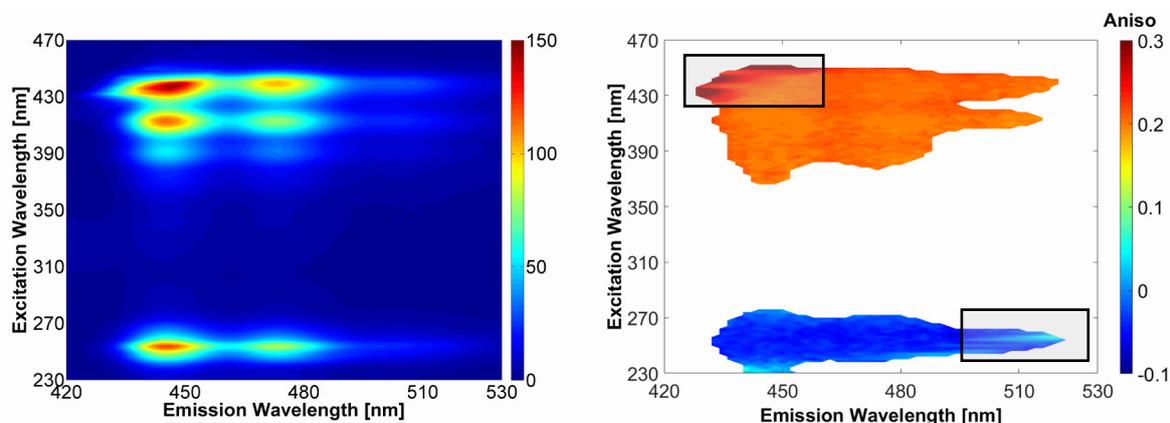
For perylene, it was defined using the weakest intensity polarization measurement (HH for S<sub>2</sub> and VV for S<sub>1</sub>). The resulting *aniso*-MDF spectra appeared however highly contaminated by the Rayleigh scatter (Figure A- 3, right), with abnormally high *r* values generated<sup>29,158</sup>. The

presence of 1<sup>st</sup> and 2<sup>nd</sup> order Rayleigh scatter needs thus to be efficiently corrected, which is even more critical for ARMES considering that the scattered light contamination is known to ruin the decomposition of EEM landscape<sup>222,289,302-303</sup>.

### 4.3.1 Rinnan's method and threshold level

Several mathematical approaches have been proposed to try to minimize or correct for the Rayleigh scatter in EEM. Thygesen *et al.*<sup>304</sup> proposed to insert zero values outside of the data area, whereas Bro proposed to insert NaN values combined with a non-negativity constraint<sup>146</sup> during PARAFAC modelling. Jiji *et al.*<sup>302</sup> used another method based on the weight of data points, in order to reduce the significance of the scattering band during trilinear decomposition. Wentzell *et al.*<sup>305</sup> used the same kind of strategy but chose to eliminate the scatter by using weighted PCA on the unfolded EEM matrix, before refolding the data and decomposing the EEMs.

Probably the most common method used to correct for Rayleigh or Raman scatter in MDF-EEM data is based on the use of interpolation<sup>306</sup>, which replaces contaminated parts of the EEM with new data calculated from nearest neighbours, using polynomial functions. However in the case of Perylene (Figure 32), we found that in the regions of overlap with the emission bands, interpolation was not able to completely remove the scatter (Figure 33).

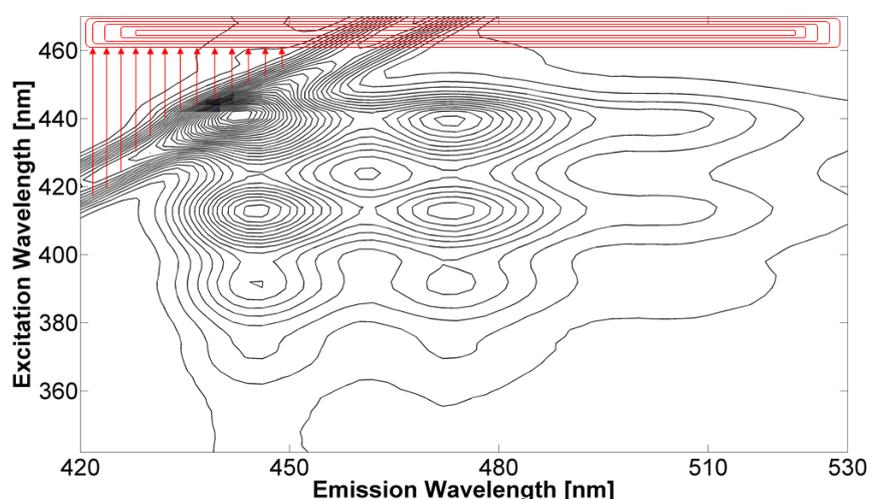


**Figure 33: (Left) HH-EEM of 1  $\mu$ M perylene in glycerol (using dWGP, 25°C, in triplicate) after interpolation correction. (Right) Corresponding *aniso*-MDF spectrum (10% threshold), where the residual Rayleigh scatter is clearly visible and highlighted in the boxes.**

One way to overcome this problem of overlap is to model the Rayleigh scatter as a separate set of components, and then remove it from the original data. This is in essence the method developed by Rinnan *et al.*<sup>222</sup>. One of the main reasons we selected this method is because scattered light intensity can also be related to particle size in solution<sup>307-308</sup>. Therefore

it could be very informative for monitoring protein aggregation to be able to extract this information as a separate set of components for different analysis (see section 5.9.1).

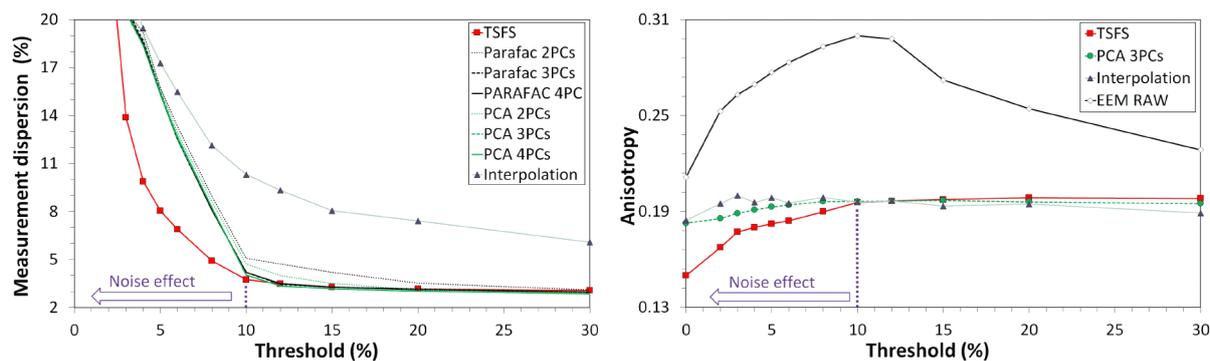
The principle of this method is to transform Rayleigh scatter from a non-bilinear structure to a low rank bilinear structure, which can then be modelled by PCA or PARAFAC. This is done by shifting each contaminated emission spectrum (Figure 34), to form a line on the EEM landscape (bilinear element) rather than a diagonal (non-bilinear element). Once all emission spectra are shifted, the signal is truncated in order to keep only the part related to the scatter, which can then be modelled by PCA or PARAFAC. The modelled scatter is then re-shifted back in the original matrix space, and subtracted from the raw signal<sup>222</sup>.



**Figure 34: Illustration of the principle behind Rinnan’s method, showing the shifting of perylene  $S_1$  emission spectra (red arrows) contaminated by 1<sup>st</sup> order Rayleigh scatter to form a low rank bilinear structure (red rectangles).**

To assess the efficiency of this method, the  $r$  dispersion as well as the average  $r$  value (see section 7.4 in Appendices for more information) obtained across the entire  $S_1$  region were calculated for different threshold values and compared to the results obtained using interpolation.

The TSFS dataset acted as a reference (Rayleigh free). The closer the  $r$  dispersion values were to the TSFS values, the more efficient the correction. Both PCA and PARAFAC options were studied, and because the Rayleigh band is a low-rank bilinear structure<sup>222</sup>, two to four components were used to model the 1<sup>st</sup> order Rayleigh scatter.

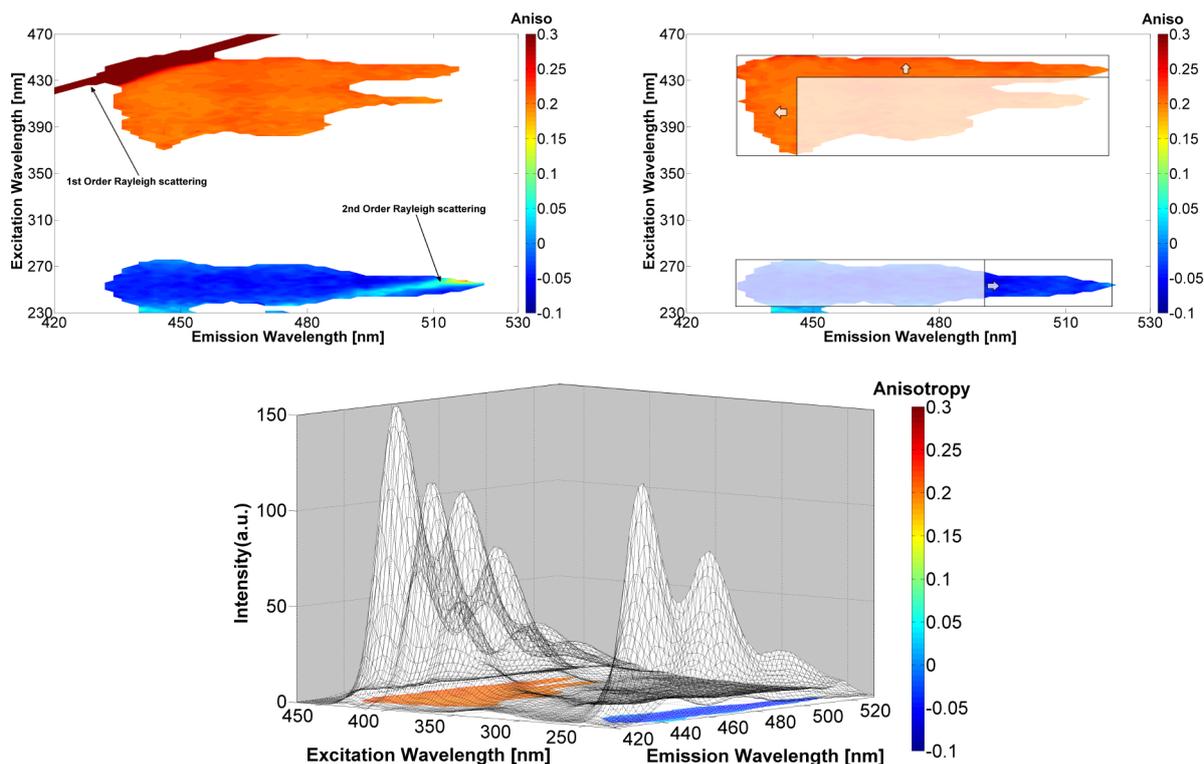


**Figure 35. Efficiency comparison between modelling (PCA, PARAFAC) and interpolation methods for correcting 1<sup>st</sup> order Rayleigh scatter by calculating (left)  $r$  dispersion vs threshold and (right)  $r$  average vs threshold. TSFS data are used as reference (Rayleigh free). Calculations were done on the  $S_1$  band from a 1  $\mu\text{M}$  solution of perylene in 100% glycerol (collected using dWGP, in triplicate, 25°C).**

Rinnan's method seemed to be more efficient in correcting for the presence of 1<sup>st</sup> order Rayleigh scatter than the interpolation method, because the  $r$  dispersion values were much more similar to the TSFS values (Figure 35, left). In terms of the modelling method used, it is difficult to say whether PCA or PARAFAC was best, but it appears that PCA was marginally better. This can probably be explained by a non-ideal alignment during the shifting of the emission spectra, generating a small distortion in bilinearity, thus reducing the efficiency of PARAFAC. It also appeared that three components were enough to model the Rayleigh scatter, as no improvement was seen for four components in terms of  $r$  dispersion. This was expected due to the low-rank nature of shifted scatter<sup>222</sup>. The second conclusion from the data (Figure 35), was that a threshold of 10% is best suited to build the *aniso*-MDF spectra, because the influence of noise increases dramatically at lower thresholds.

### 4.3.2 Analysis of corrected *aniso*-MDF spectra: increase of perylene's working range as multidimensional anisotropy standard

Using Rinnan's method and a three component PCA model, a corrected *aniso*-EEM was produced for 1  $\mu\text{M}$  perylene samples in 100% glycerol (Figure 36). By modifying slightly the algorithm to also correct for the 2<sup>nd</sup> order Rayleigh scatter (details on code modifications are given in the Appendices, section 7.5), it appeared that both 1<sup>st</sup> and 2<sup>nd</sup> order were entirely removed, and that correct  $r$  values were retrieved in the regions of overlap with scatter, proving again that only pure scatter was removed from raw data.



**Figure 36. (Top) *Aniso*-EEM plots of 1  $\mu\text{M}$  solution of perylene in glycerol, shown before (left) and after (right) Rayleigh scatter correction by modelling. The boxes show the increase in the working range of perylene. These plots were obtained with a 10% threshold of the HH and VV scan in the  $S_2$  and  $S_1$  regions respectively. (Bottom) Overlay of the Rayleigh corrected HH-EEM/*aniso*-EEM of perylene. Data collected using dWGP, in triplicate at 25  $^\circ\text{C}$ .**

The working range for Perylene as a multidimensional anisotropy standard was consequently increased, for  $S_1$  from  $\lambda_{\text{ex/em}} = 370\text{-}430/450\text{-}520$  nm, to  $\lambda_{\text{ex/em}} = 370\text{-}460/430\text{-}520$ , and for  $S_2$  from  $\lambda_{\text{ex/em}} = 240\text{-}270/430\text{-}490$  nm to  $\lambda_{\text{ex/em}} = 240\text{-}270/430\text{-}520$  nm. We thus proved that an efficient method was available to remove the Rayleigh scatter from ARMES-EEM data. The next step is to find the right chemometric method that enables ARMES to accurately recover each of the fluorophore spectra and the component anisotropy from various dyes mixtures.

## 4.4 Anisotropy recovery from fluorophore mixtures using MCR-ALS

### 4.4.1 Problem statement

A major problem with MDF spectroscopy is that extensive spectral overlap makes spectral interpretation difficult. As shown in the introduction, Multivariate Curve Resolution (MCR)<sup>141,143,237</sup> and Parallel Factor analysis (PARAFAC)<sup>146,309-310</sup> can be used to identify the

spectral contributions of individual constituents from MDF data. However, these methods work well only when MDF data is not, or minimally affected by scatter, IFE, and FRET. Rayleigh and Raman scatter also cause problems as they are non-bilinear elements, negatively impacting the deconvolution of EEM<sup>222,289,302-303</sup>. While scatter can be corrected with a variety of methods, it would be much more advantageous to avoid the strongest 1<sup>st</sup> order Rayleigh using TSFS measurements.

Unfortunately, one drawback with TSFS data modelling is its different data structure which arises from the dependence of the emission measured at the offset ( $\Delta\lambda$ ) with  $\lambda_{\text{ex}}$ <sup>311</sup>. Whereas in EEM (of well-behaved samples like a single ideal fluorophore) the intensity of the fluorescence signal changes as a function of both the excitation or emission wavelength, the spectral profile tends to remain the same in both dimensions (*i.e.* the shape of the excitation and emission spectra do not vary). EEM is thus considered to be trilinear because both excitation and emission are highly reproducible spectral orders<sup>294</sup>. In TSFS, a change in  $\Delta\lambda$  induces a change in spectral profile, which means that TSFS does not behave trilinearly<sup>312</sup>, which has a negative impact on MCR or PARAFAC modelling. The situation becomes more complex if IFE<sup>59</sup> or FRET<sup>299-300</sup> are present. Both effects break the linearity between fluorescence intensity and concentration, and can also drastically distort EEM profiles. Because this distortion can vary from sample to sample, the reproducibility of excitation and emission spectral orders is greatly reduced, which breaks the trilinearity of EEM data<sup>289,296-298</sup>. In TSFS data, IFE also affects spectral profiles and can result in misinterpreted data<sup>313</sup>.

In the first ARMES studies<sup>13-14</sup> using TSFS measurements and polymer thin film polarizers (TFP), albumin proteins generated complex *aniso*-TSFS spectra that were diagnostic of both protein identity and structure. However, as explained in the previous chapter, TFP do not transmit below  $\sim 290$  nm, which did not permit the collection of the complete intrinsic protein fluorescence originating from Tyrosine and Tryptophan. TFP also caused the excitation maxima of the recovered components to be both red-shifted and reshaped, compared to non-polarized spectra. While this distortion compensated for the lack of trilinearity in TSFS data, it also distorted the recovered excitation profiles which often appeared to have dual bands, an artefact caused by the TSFS data structure<sup>14</sup>.

We showed that replacing TFP with dWGP that have better UV transmittance enabled the acquisition of data in the important tyrosine/tryptophan region<sup>15</sup>. However, the use of UV

transmitting filters also resulted in the elimination of the spectral reshaping previously observed in TSFS measurements<sup>14</sup> which has consequences in terms of data structure.

Before applying ARMES to the analysis of proteins, it was first necessary to investigate and standardize the chemometric methodology used. In particular, one has to determine if it was possible to extract using curve resolution methods (here MCR) accurate excitation and emission, scores and profiles, and therefore anisotropy values for individual fluorophores using either TSFS or EEM data. To do this we also had to determine the optimal data augmentation procedure to ensure recovery of reliable spectral profiles and scores from MCR. A second facet of the study was to determine if TSFS measurements made using a dual wire grid polarizer (DWGP) could still be analyzed as demonstrated previously<sup>14</sup> which required an investigation into how to deal with trilinear (EEM) and non-trilinear (TSFS) data. We also had to ascertain what were the optimal constraints to use, to produce accurate solutions without rotational ambiguities. This was important in some measurement situations where excess light scatter cannot be avoided and thus TSFS will be more suitable measurement, for example in turbid media like bioreactors.

We decided to re-investigate the two systems used in the last chapter *i.e.* perylene in glycerol, and a mixture of Erythrosin B and Phloxine B. The latter is more challenging as both molecules have overlapping emission spectra of the same intensity. However their different anisotropies make them ideal to test the accuracy of the method. These two sample sets had minimal IFE and FRET, which ensured data trilinearity and minimized unwanted depolarization effects.

#### 4.4.2 Materials and data analysis

The reader is invited to refer to sections 2.3.2 and 2.3.3 for details about the sample preparation and the data collection respectively. All concentrations were chosen to minimise IFE. Also there was no particular reason for selecting a closed system ( $\sum C_i = 100\%$ ) for the EB:PB mixtures. However, this is a worst-case example for MCR because the different concentration profiles are linearly dependent (if one decreases, the other one increases, or both decrease or increase at same time), which can create rank deficiency. One, therefore, can evaluate the quality of the MCR models if they can deal with this issue.

In terms of data analysis, the chemometric analyses were performed using MATLAB (ver. 7.0.1) using either PLS\_Toolbox4.0® or MCR\_ALS 2.0 toolbox<sup>314</sup> for implementation of trilinear/correspondence-among-species constraints. In this chapter, MCR-ALS was

selected for data analysis to ensure continuity with previous studies<sup>13-14</sup> although PARAFAC<sup>146,309-310</sup> could also be applied to this type of data. TSFS data were organized by  $\lambda_{\text{ex}}$  (mode 1),  $\Delta\lambda$  (mode 2), and sample (mode 3), same as EEM where mode 2 was  $\lambda_{\text{em}}$ .

Spectral decomposition was performed on augmented matrix datasets and unless otherwise stated, for TSFS the “concentration” mode corresponded to “ $\lambda_{\text{ex}} \times \text{sample}$ ” and the “spectral” mode to  $\Delta\lambda$ . This was the same for EEM except that the “spectral” mode was  $\lambda_{\text{em}}$ . Spectral deconvolution was performed using MCR-ALS<sup>255</sup> and estimation of rotational ambiguities were performed using MCR-BANDS<sup>270</sup>. The number of factors (or rank analysis) were assessed using singular value decomposition (SVD)<sup>x</sup>. Separate models were built for each polarization setting once initial estimates had been obtained using SIMPLISMA<sup>316</sup>. A non-negativity constraint was used on both concentration and pure spectra profiles, as well as normalization of the spectral profiles to equal length (Euclidean) to avoid intensity ambiguities between components. In all cases, models converged with fewer than 100 iterations (due to the simplicity of the dataset). A similarity index (SimI), was used to rapidly compare EEM/TSFS spectra or *aniso*-MDF spectra<sup>317</sup>. SimI between two matrices  $X_1$  and  $X_2$  (*e.g.* EEM or TSFS of dimensions  $I \times J$ , discrete data points) was calculated using the following equation:

**Equation 56** 
$$\text{SimI} = 1 - \lambda \frac{\sqrt{\sum_{i=1}^I \sum_{j=1}^J p_{x_1-x_2}^2}}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J p_{x_1+x_2}^2}}$$

$\lambda$  was a penalty parameter, used to set a detectable limit of variance between  $X_1$  and  $X_2$ . In these calculations it was set to 4 which corresponds to 5% variance in the multidimensional fluorescence data and was proven quite adequate.  $p_{x_1-x_2}$  and  $p_{x_1+x_2}$  are elements of  $(X_1-X_2)$  and  $(X_1 + X_2)$  respectively. The closer SimI value was to one, the more alike the two matrices  $X_1$  and  $X_2$  were.

### 4.4.3 Results and Discussion

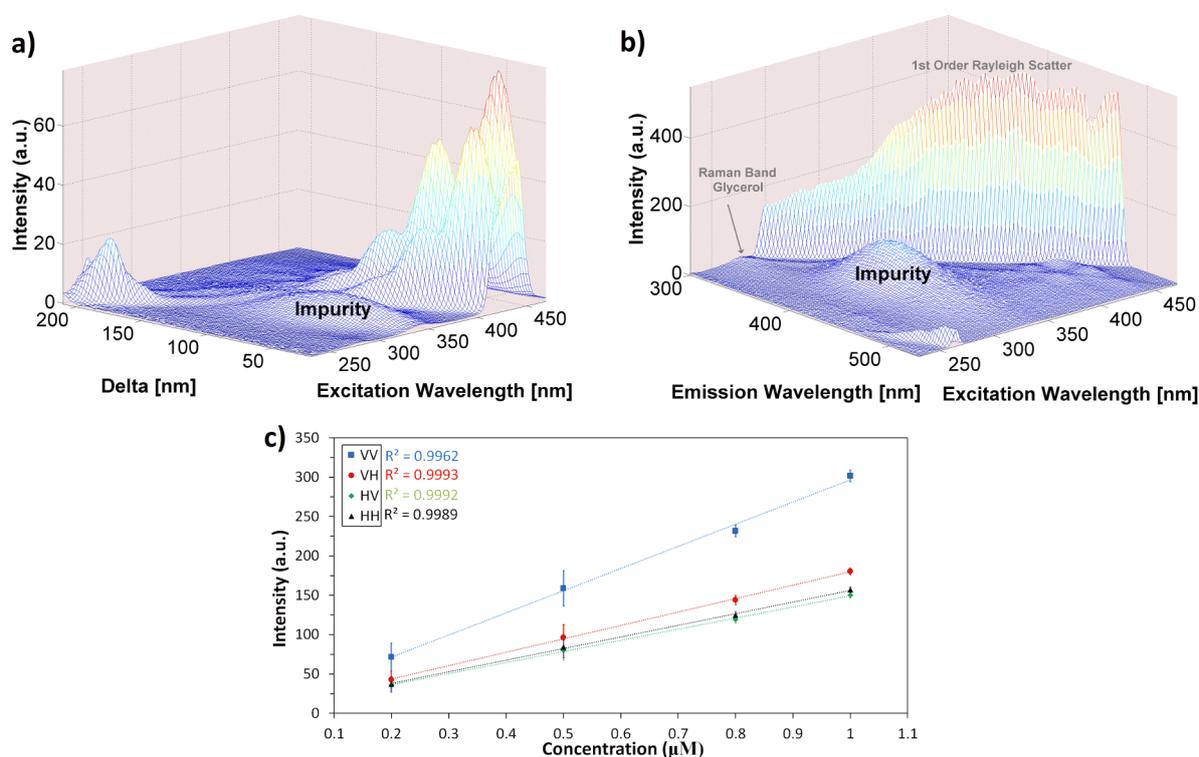
#### 4.4.3.1 Perylene data: presence of a fluorescent impurity

A simple standard sample set (perylene, varying concentration in glycerol<sup>15</sup>) was used as the initial test to establish if either TSFS and EEM, could accurately extract component anisotropy using MCR modelling.

---

<sup>x</sup> SVD is a widely used method in exploratory data analysis<sup>315</sup>. It decomposes a matrix  $X$  ( $m \times n$ ) such as:  $X = USV^T$ .  $U$  ( $m \times n$ ) and  $V$  ( $n \times n$ ) are unitary matrices with orthonormal columns.  $S$  ( $n \times n$ ) is a diagonal matrix of positive elements for which the first  $k$  elements enable to obtain a matrix  $X_k$  that is the closest approximation of  $X$  (minimizing the sum of the squares of the difference of the elements of  $X$  and  $X_k$ ). The  $k$  elements are called the singular values. SVD is thus a useful method to estimate the rank of a matrix  $X$ <sup>315</sup>.

These samples contained a weak fluorescent impurity (Figure 37) which originated from the solvent. This was confirmed by overlapping the EEM spectrum of pure glycerol with perylene (Figure A- 5, Appendices). Glycerol is known to contain organic fluorescent impurities<sup>318-320</sup>. While the use of more highly purified solvents and blank subtraction is recommended to remove the influence of minor solvent impurities and Raman scatter, it may not always be feasible. However, we felt that it was also important to demonstrate that chemometric modelling could also recover weak fluorescing components such as impurities (which could originate from fluorophore or glassware contamination). Especially, as weak fluorescing components can also be encountered during protein analysis.



**Figure 37: (a) VV-TSFS spectrum of a 0.2 μM Perylene solution, showing the impurity signal at  $\lambda_{\text{ex}}/\Delta\lambda = 328/70$  nm; (b) VV-EEM spectrum of glycerol (blank) collected using a higher detector gain (PMT voltage of 800V instead of 600V); (c) Fluorescence intensity vs concentration of perylene in 100% glycerol, for each polarization setting. Error bars were obtained over three replicates.**

In this TSFS data, 1<sup>st</sup> order Rayleigh scattering was intrinsically eliminated but Raman scattering from the solvent was present, even though it was too weak to be observed with the experimental conditions used. The Raman spectrum of glycerol (mainly O-H stretching mode at  $\sim 3400$  cm<sup>-1</sup>) was only measurable from blank measurements in the VV orientation (highly polarized scatter) using a higher gain on the detector (PMT voltage of 800V instead of 600V).

But even here, the Raman band of glycerol was < 7% of the intensity of the impurity signal (Figure 37), which means that it could be eliminated as a source of error.

Rayleigh scatter was therefore the major non-bilinear element to be considered here for the EEM data of perylene because the 1<sup>st</sup> order band overlapped the S<sub>1</sub> emission and 2<sup>nd</sup> order the S<sub>2</sub> emission. We saw in section 4.3 that due to this extensive overlap, classical interpolation could not correct sufficiently well for anisotropy studies. We therefore used the modelling method instead<sup>222</sup>. If the Rayleigh was not removed then MCR modelling produces very erratic results (Figure A- 6, Appendices).

#### 4.4.3.2 MCR Analysis of polarized TSFS perylene data

The TSFS trilinear-like behaviour observed by Groza *et al.*<sup>14</sup> which was due to TFP induced spectral reshaping, enabled the use of column wise augmentation (CWA),  $\Delta\lambda \times (\lambda_{\text{ex}} \times \text{samples})$ , for MCR-ALS analysis. This however contradicted the rank dependency normally observed in this type of augmentation<sup>311</sup>. This reshaping was eliminated once the dWGP were used since the full emission spectra was now obtained. Modelling this dWGP-TSFS perylene dataset (for all four polarization configurations) using CWA required five components which was obviously too many (scores and loadings are shown Figure A- 8/figure A- 9 in Appendices). Instead of fitting/generating a pure perylene spectrum, the components appeared as combinations/rearrangements of the different vibronic bands associated with each electronic transition (Figure A- 7) and this effect was directly related to the non-bilinear behaviour of TSFS data<sup>312</sup>.

Even though all the CWA-MCR-TSFS models explained >90% of the variance (Table A- 2) and components were recovered without any rotational ambiguities (Table A- 3) the Lack of Fit (LOF) was too high, which is symptomatic of non-bilinear data trying to be fitted with a bilinear model. LOF<sup>255</sup> corresponds to the difference between the input data (X) and the data reproduced by MCR-ALS:  $C \times S^T$ . In this equation,  $x_{ij}$  correspond to an element of X and  $e_{ij}$  of the model residuals.

**Equation 57** **lack of fit (%) = 100**  $\sqrt{\frac{\sum_{i,j} e^2_{ij}}{\sum_{i,j} x^2_{ij}}}$

Increasing component number (to 9 or 13 suggested by SVD local minima) decreased LOF to ~ 2% and increased the explained variance to >99% in all modes. However, this did not make chemical sense, because the recovered components (not shown) all overlapped. In addition, the anisotropies calculated from five pure profiles extracted from each MCR model were not constant along  $\Delta\lambda$  (for example component 1 and 5, Figure A- 10). Therefore, we can

conclude that CWA was not satisfactory. Row-wise augmentation would be expected to give similar level resolution, as  $\lambda_{\text{ex}} \times (\Delta\lambda \times \text{samples})$  with concentration and spectral modes that were co-dependant.

**Table 10: Perylene in Glycerol MCR best model results. Percentage of variance captured by each MCR model component (Fit %X). Each model converged with ~ 100 iterations, and were built using variety of constraints 1). normalization, 2). non-negativity, 3). unimodality, and 5). total trilinearity. For SWA-TSFS and EEM models, non-negativity was applied on both concentration and spectral modes, and normalization was applied on spectral modes. Unimodality for SWA-TSFS model was applied to spectral profiles and total trilinearity for EEM model was applied to augmented mode (concentration mode).**

<b>SWA-TSFS MCR model (constraints 1,2,3)</b>				
<b>Comp</b>	<b>HH (%X)</b>	<b>HV (%X)</b>	<b>VH (%X)</b>	<b>VV (%X)</b>
<b>1</b>	99.82%	99.86%	99.89%	99.86%
<b>2</b>	0.12%	0.10%	0.07%	0.12%
<b>Sum</b>	99.94%	99.96%	99.96%	99.98%
<b>Lack of Fit</b>	2.40%	1.95%	2.02%	1.34%
<b>EEM MCR model (constraints 1,2,5)</b>				
<b>Comp</b>	<b>HH (%X)</b>	<b>HV (%X)</b>	<b>VH (%X)</b>	<b>VV (%X)</b>
<b>1</b>	99.77%	99.78%	99.84%	99.80%
<b>2</b>	0.19%	0.17%	0.12%	0.15%
<b>Sum</b>	99.96%	99.95%	99.96%	99.95%
<b>Lack of Fit</b>	2.09%	2.12%	1.93%	2.26%

The correct way to unfold a TSFS dataset for MCR-ALS is along the sample direction (samples  $\times (\lambda_{\text{ex}} \times \Delta\lambda)$ ) as only this allowed generation of an independent and equal rank for both directions of the unfolded matrix (samples and  $\lambda_{\text{ex}} \times \Delta\lambda$ )<sup>311</sup>. This rank should also equal to the real number of fluorophores in solution if there was no FRET or IFE effects present. Using this sample wise augmentation (SWA) approach, the unfolded TSFS matrix behaved bilinearly and MCR-ALS modelling of this SWA-TSFS perylene data only required two components (Table 10). A first attempt using only normalization and non-negativity constraints (Table A- 6), yielded one component for perylene (98% explained variance), and a second for the fluorescence impurity (~2% variance). However, component two appeared to be a composite of impurity emission (centred at  $\lambda_{\text{ex}}/\lambda_{\text{em}}$  328/398 nm) mixed with perylene emission (Figure A- 15). To improve resolution, stricter constraints were applied by adding unimodality

to component two, using the average implementation and a tolerance of 1.1, because the second component behaved unimodally (Figure 37 a). Consequently, LOF increased from ~1.3% to ~1.9%, however, compared to CWA, the results were indicative of a true, bilinear structure in the augmented TSFS dataset.

Resolution was improved (Figure A- 16/Figure A- 17) and the first component (perylene) now explained >99.8% of the variance, and the second component <0.2% (similar to EEM, Table 10). The second component was however only partially resolved (inset Figure A- 16) and not equally resolved in all samples (as shown by the decreasing scores in Figure A- 17). It seems likely that some of the impurity emission is still present in the 1<sup>st</sup> component (Figure A- 19). This model, however, enabled the recovery of accurate perylene anisotropy values (Table A- 7, Figure A- 18) which was important because TSFS may be the only practical ARMES measurement method for highly scattering samples. However, the main issue with this model was the presence of rotational ambiguities in the solution (especially for impurity, Table 11), which cannot be reduced (by application of trilinear constraint for example) showing the limitations of SWA-TSFS.

**Table 11: Estimation of component rotational ambiguities by MCR-BANDS using various constraints: 1). Normalization, 2). Non-negativity, 3). Unimodality, and 5). Total trilinearity, for TSFS and EEM MCR models.**

<b>SWA-TSFS MCR model.</b>								
<b>1,2,3</b>	<b>Component 1 (~ 99.8%)</b>				<b>Component 2 (~ 0.1%)</b>			
	<b>HH</b>	<b>HV</b>	<b>VH</b>	<b>VV</b>	<b>HH</b>	<b>HV</b>	<b>VH</b>	<b>VV</b>
<b>f<sub>n</sub> initial</b>	0.998	0.998	0.999	0.998	0.035	0.031	0.026	0.035
<b>f<sub>n</sub> max</b>	0.998	0.998	0.999	0.998	0.145	0.136	0.132	0.172
<b>f<sub>n</sub> min</b>	0.937	0.896	0.950	0.884	0.035	0.136	0.026	0.035
<b>f<sub>n</sub> max-f<sub>n</sub> min</b>	0.061	0.102	0.049	0.114	0.110	0.105	0.106	0.137
<b>EEM MCR model</b>								
<b>1,2,5</b>	<b>Component 1 (~ 99.8%)</b>				<b>Component 2 (~ 0.2%)</b>			
	<b>HH</b>	<b>HV</b>	<b>VH</b>	<b>VV</b>	<b>HH</b>	<b>HV</b>	<b>VH</b>	<b>VV</b>
<b>f<sub>n</sub> initial</b>	0.992	0.991	0.994	0.995	0.044	0.041	0.034	0.038
<b>f<sub>n</sub> max</b>	0.992	0.991	0.994	0.995	0.044	0.041	0.034	0.038
<b>f<sub>n</sub> min</b>	0.992	0.991	0.994	0.995	0.044	0.041	0.034	0.038
<b>f<sub>n</sub> max-f<sub>n</sub> min</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The loss of trilinearity because of the WGP increased UV transmittance meant that CWA was not applicable to TSFS data. SWA unfolding on the other hand was mathematically correct for MCR-ALS<sup>312</sup>, and enabled the separation of perylene and impurity once the correct constraints were applied. Even if the 2<sup>nd</sup> component only partially recovered the impurity, we were able to accurately reconstruct anisotropy values of Perylene for both excited states. Although SWA permitted bilinear decomposition of augmented TSFS data, the data was still not trilinear. This meant that neither PARAFAC nor trilinear constraints could be applied, and one therefore must deal with presence of rotational ambiguities in the model solutions.

#### 4.4.3.3 MCR-ALS models on EEM data.

For EEM data, a key consideration was implementation of a trilinearity constraint during MCR modelling to ensure accurate and unambiguous recovery of all components. We saw in the introduction (section 4.1.8) that the development of mathematical multilinear constraints<sup>294</sup> made it possible to reproduce the inner trilinear structure<sup>271</sup> of PARAFAC in MCR-ALS, by implementing trilinearity as an optional constraint. The trilinear constraint in MCR-ALS forces the resolved pure profiles to behave the same in all C or S<sup>T</sup> matrices. Trilinearity constraints are powerful as they ensure resolution of pure and unambiguous profiles<sup>271,228</sup>, but fluorescence dataset requirements for its use are demanding as they must be free of scatter, IFE, and FRET<sup>294</sup>. The Rayleigh corrected polarized-EEM datasets of perylene in glycerol fulfils the trilinearity criteria as shown by the plots of maximum fluorescence intensity against concentration obtained for each polarized dataset (Figure 37 c) which are all highly linear ( $r^2 > 0.99$ ). This is further confirmed by the anisotropy measurements (Table 13) which were constant (and agreed with the literature<sup>165</sup>) at all concentrations, indicating the absence of depolarization processes such as IFE or FRET.

We also verified that the dataset was not unnaturally forced to fit the trilinear constraint by following how the MCR model fit evolved as we progressed from the use of a simple bilinear constraint, to the use of a partial trilinear (shape) constraint, and finally a total trilinear (shape + synchronisation) constraint were applied (Table 12). If the percentage of variance explained  $R^2$  and the lack of fit LOF were constant, then one can reasonably consider that the data structure was suitable for using a trilinear constraint. Here all the polarised perylene/glycerol datasets were found to be suitable. It is useful to note that PARAFAC would have been a good alternative to MCR with the trilinear constraint for modelling this type of data. However, here we also needed to compare bilinear with trilinear models, and MCR offers the advantage to switch between bilinear, partial trilinear, and total trilinear cases.

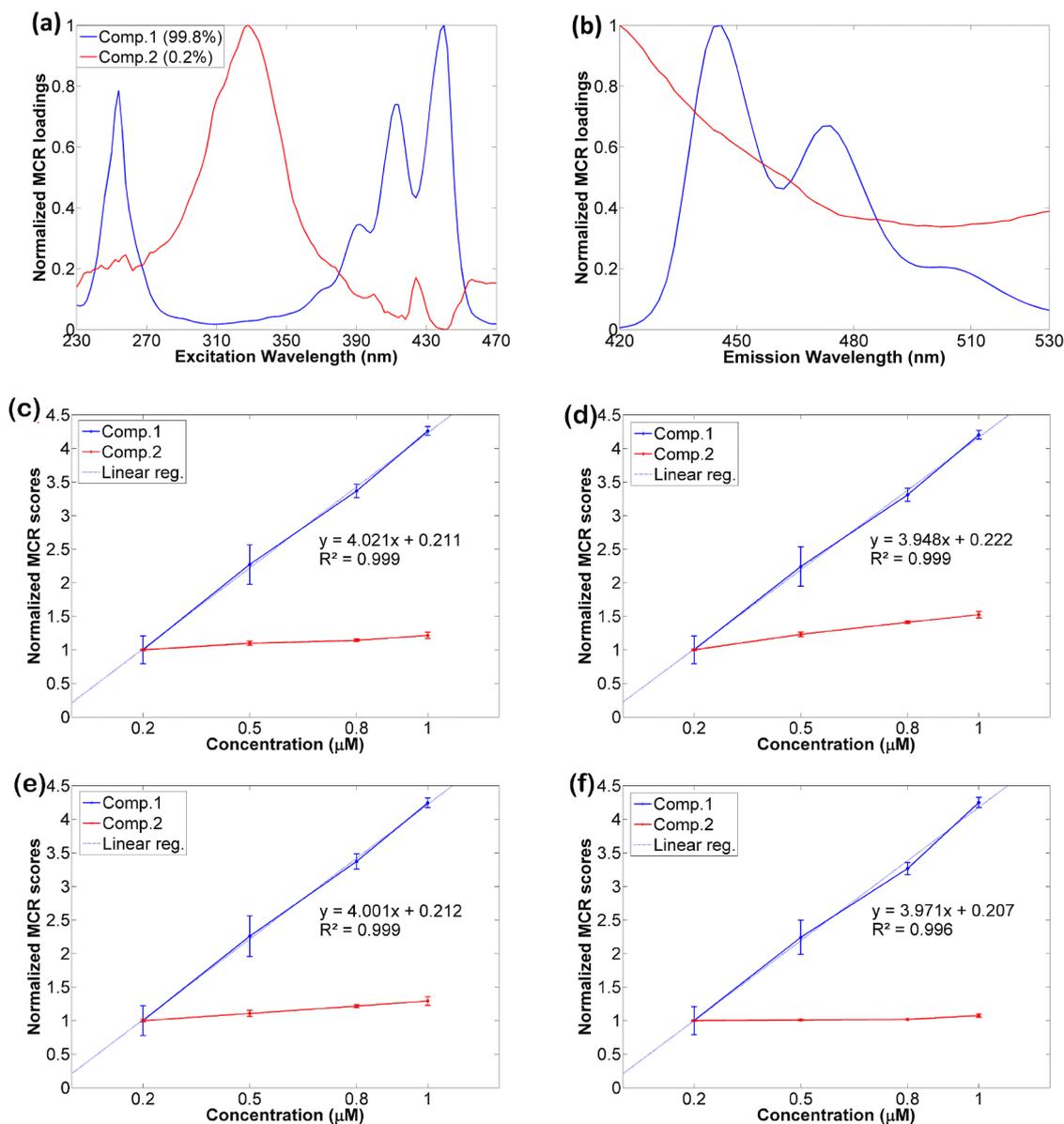
**Table 12: Evolution of R<sup>2</sup> and LOF of MCR models applied to CWA-EEM polarized data, when: 1). Normalization, 2). Non-negativity, 3). Bilinear, 4). Partially trilinear, and 5). Total trilinear constraints were applied.**

	R <sup>2</sup> (total variance explained)				LOF (lack of fit)			
	HH	HV	VH	VV	HH	HV	VH	VV
<b>1,2,3</b>	99.96%	99.95%	99.97%	99.95%	1.96%	1.95%	1.83%	2.19%
<b>1,2,4</b>	99.96%	99.95%	99.96%	99.65%	2.10%	2.12%	1.95%	2.28%
<b>1,2,5</b>	99.96%	99.95%	99.96%	99.95%	2.09%	2.12%	1.93%	2.26%

**Table 13: Real and modelled (from EEM-MCR, constraints 1,2,5) averaged anisotropy values calculated for each concentration, for  $\lambda_{ex/em}$  (S<sub>2</sub>) = 240-270/430-520 nm and  $\lambda_{ex/em}$  (S<sub>1</sub>) = 370-460/430-520 nm.**

	Real anisotropy $\pm$ STD		Modelled anisotropy $\pm$ STD	
	S <sub>2</sub>	S <sub>1</sub>	S <sub>2</sub>	S <sub>1</sub>
<b>0.2 <math>\mu</math>M</b>	-0.034 $\pm$ 0.010	0.204 $\pm$ 0.015	-0.031 $\pm$ 0.012	0.201 $\pm$ 0.014
<b>0.5 <math>\mu</math>M</b>	-0.026 $\pm$ 0.004	0.202 $\pm$ 0.007	-0.030 $\pm$ 0.005	0.202 $\pm$ 0.007
<b>0.8 <math>\mu</math>M</b>	-0.030 $\pm$ 0.004	0.192 $\pm$ 0.004	-0.037 $\pm$ 0.002	0.194 $\pm$ 0.003
<b>1 <math>\mu</math>M</b>	-0.035 $\pm$ 0.004	0.204 $\pm$ 0.002	-0.027 $\pm$ 0.002	0.205 $\pm$ 0.002
<b>Av/std</b>	-0.031 $\pm$ 0.004	0.201 $\pm$ 0.005	-0.031 $\pm$ 0.004	0.201 $\pm$ 0.005

As expected, MCR modelling of EEM data required two components (Table 10) with explained variance for the second impurity component (0.2%) slightly higher than SWA-TSFS where impurity was only partially recovered (Figure A- 16). This recovered second component was purer, entirely recovered as shown by the normalized-averaged excitation and normalized emission of the spectral loadings (Figure 38 a/b for HH, others are shown in Figure A- 12 in Appendices). The use of a total trilinear constraint during ALS optimization significantly improved impurity resolution. This can be shown by comparison with the results from a model run only with non-negativity and normalization constraints (Figure A- 11). In addition, the trilinearity constraint also ensured that both components were resolved without any ambiguities (Table 11) whereas the use of only non-negativity and normalization constraints resulted in some rotational ambiguities, particularly with the weak second component (Table A- 4). This was also a big improvement compared to SWA-TSFS model.



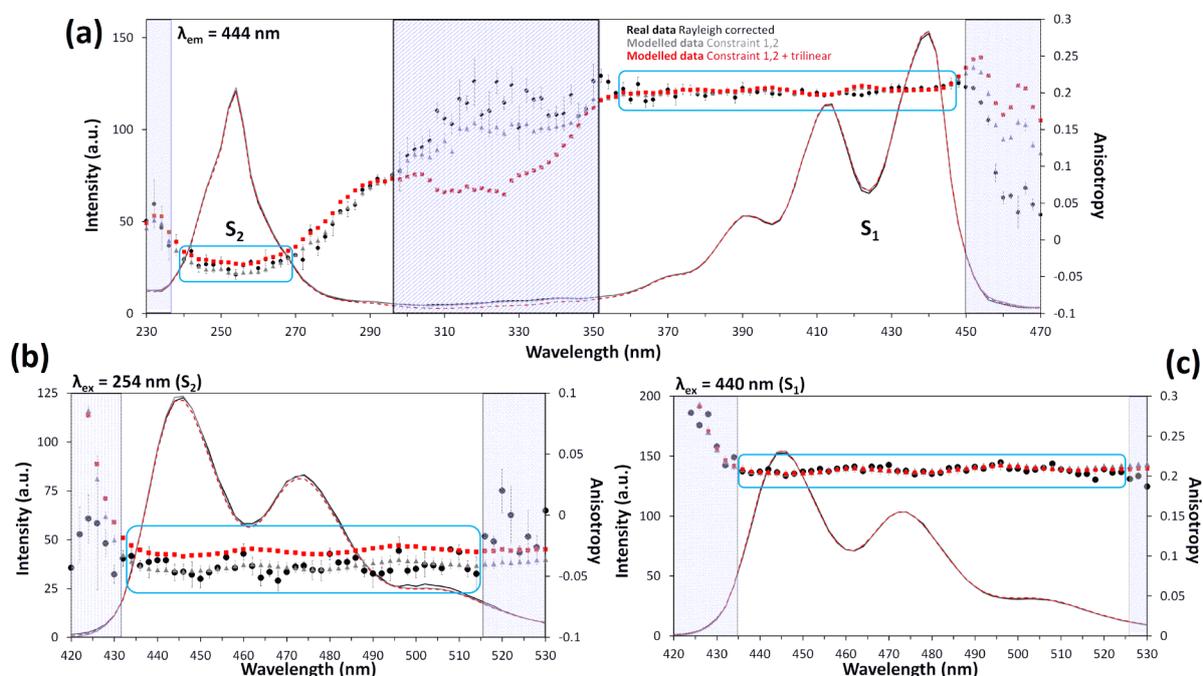
**Figure 38:** (top, a/b) Normalized excitation (averaged across all samples, left) and emission (right) profiles obtained for each pure component after MCR modelling of the HH polarized EEM datasets. Note: emission axis in (b) was truncated to better show solely the perylene emission.; (bottom) MCR scores (scaled to 1st sample) obtained for HH (c), HV (d), VH (e), and VV (f) polarized data. Models obtained using: 1). normalization, 2). non-negativity constraints and 5) total trilinear constraints.

The normalised MCR scores (Figure 38 c-f) which represent the relative concentration of each extracted fluorophore show that the first component scores increase very linearly ( $R^2 > 0.99$  for all) with perylene concentration. The scores for component 2 are not constant with perylene concentration and tend to increase particularly in HV and VH. Due to the very low fluorophore concentrations for both components, there is low probability of this being FRET. This is also confirmed by Figure A-13 (Appendices) showing that the fluorescence intensity of the impurity was  $\sim$  constant with perylene concentration, for each polarization setting. The

observed increase was probably due to incomplete resolution in the excitation spectra ( $< 270$  nm and  $> 390$  nm) where there was small, yet significant overlap with perylene (Figure A- 12) particularly in the orthogonal HV and VH configurations.

#### 4.4.3.4 Perylene anisotropy modelling

Finally, using the spectral profiles for each component generated from the best four polarized-EEM models, we calculated the anisotropy and generated an overlay plot which compared the real anisotropy to the modelled anisotropy (Figure 39). This showed that the modelled and real anisotropy values were in excellent agreement along both excitation and emission directions for both excited states. Once averaged across the entire  $S_1$  and  $S_2$  bands (Table 13), the real and modelled anisotropy values were identical within experimental error.



**Figure 39: Overlay of HH-EEM excitation (A,  $\lambda_{em} = 444$  nm) and emission (B,  $\lambda_{ex} = 254$  nm & C,  $\lambda_{ex} = 440$  nm) spectra (all corrected for 1<sup>st</sup> and 2<sup>nd</sup> order Rayleigh scatter) from perylene (1  $\mu$ M) in glycerol with corresponding anisotropy values: black = real data; grey = modelled using constraints 1 & 2; red = modelled using 1, 2, & trilinear constraints. The blue boxed areas show the regions of most reliable anisotropy data. The shaded areas represent areas of low fluorescence intensity and thus unreliable anisotropy data.**

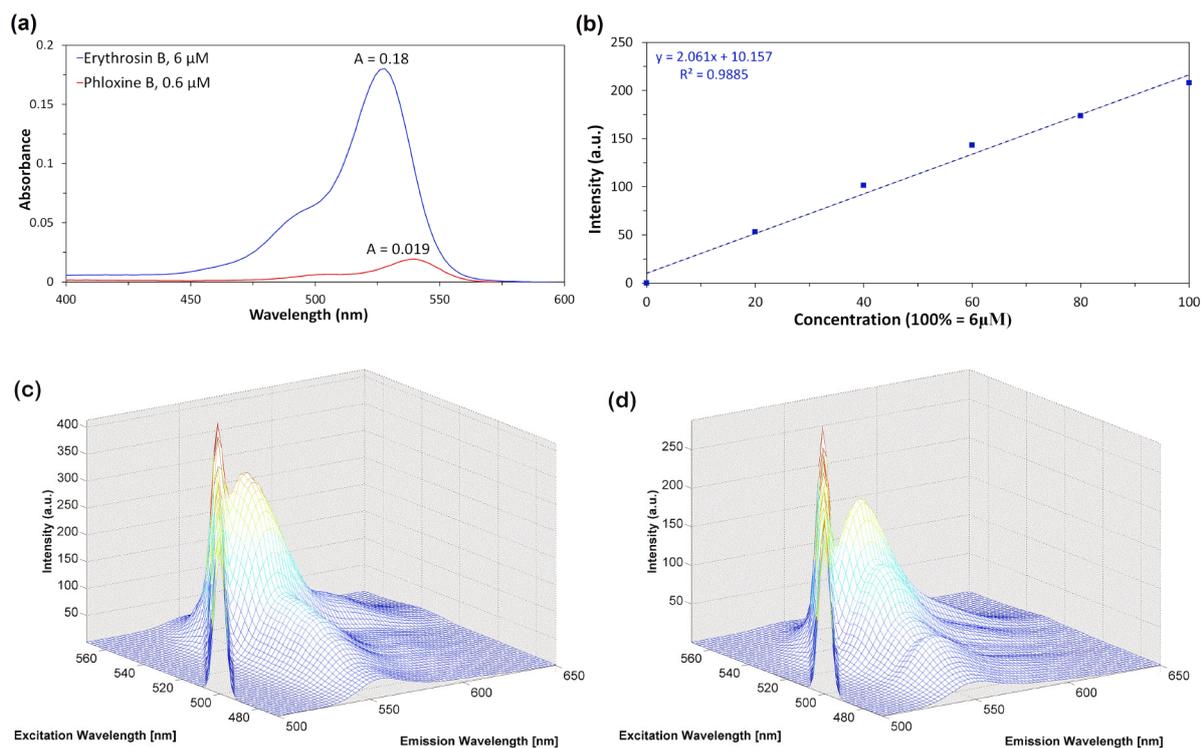
The only big difference between the real and modelled data in the excitation spectra were observed between the two electronic states ( $\lambda_{ex} = 300$ – $350$  nm), where mostly noise was present. In the emission spectra (Figure 39 b), a small discontinuity between the real and modelled spectra was observed at  $\sim 495$ – $515$  nm, which corresponded to the location of second order Rayleigh scatter. Therefore, it was possible that a small amount of residual scatter was

present in the data, which would explain the slightly higher anisotropy values obtained in this region. Both bilinear and trilinear models accurately retrieved perylene anisotropy for the  $S_1$  state, and the bilinear model was in slightly better agreement with the real anisotropy values (Table A- 5 vs Table 13). This was also observed in  $S_2$  region, where the trilinear model slightly overestimated the anisotropy. In the perylene example, where Rayleigh scatter was almost perfectly removed, the bilinear model worked better because it offered more flexibility to the spectra to best fit the data and reproduce the exact anisotropy. In contrast the trilinear model restrained the shape more strongly which led to the discrepancy observed. This was in contrast with cases where residual scatter (shot noise) in the data can affect the shape of resolved spectra, and where trilinear models can help to fix that issue (*i.e.* fluorophore mixtures). However, the differences were small and are only significant when the magnitude of the anisotropy is small, such as with the  $S_2$  state here. Trilinear models have other advantages, such as avoiding rotational ambiguities and preserving a common spectral shape for same component across samples, which is of more practical significance in cases where the components have similar emission intensities. This advantage was seen in case of the impurity, where the trilinear model better resolved the impurity profile. Modelling of this second very weak ( $\sim 0.2\%$  variance) impurity component was more challenging because the signal to noise ratio was poor. From the excitation (Figure A- 14, A) and emission (Figure A- 14, B) spectra, it seemed that within the band ( $\lambda_{\text{ex}} = 290\text{-}350\text{ nm}$ ), both bilinear and trilinear models were able to accurately reproduce the real anisotropy  $r = 0.310 \pm 0.024$ .

#### 4.4.3.5 Fluorophore mixture modelling

To further validate this chemometric methodology for accurate anisotropy recovery via MCR, we needed to examine a case where there were two fluorophores with overlapping emissions of near equal intensity but with different anisotropies. This is more representative of the situation existing in proteins. We used a mixture of Erythrosin B ( $\sim 6\ \mu\text{M}$ ) and Phloxine B ( $\sim 0.6\ \mu\text{M}$ ), analysed in HEPES buffer at pH 7.2 and  $25^\circ\text{C}$  (see Table A- 8 and section 2.3.2 for more details). These concentrations yielded equivalent fluorescence intensities and minimized IFE<sup>171</sup>. In addition, the short lifetimes (for example 75 ps for Erythrosin B in water<sup>321</sup>) of these fluorophores should further minimize depolarization effects. The absorbance and the uncorrected fluorescence EEM spectra (Figure 40) of Phloxine B and Erythrosin B show that there is a small degree of IFE for EB, expected because of the relatively high absorbance ( $\sim 0.18$ ) for  $6\ \mu\text{M}$  EB compared to ( $\sim 0.019$ ) for  $0.6\ \mu\text{M}$  PB.

#### 4. Accurate anisotropy recovery from fluorophore mixtures using MCR



**Figure 40: (a) UV-Vis spectra of 6  $\mu\text{M}$  Erythrosin B and 0.6  $\mu\text{M}$  Phloxine B; (b) Plot of maximum fluorescence intensity versus Erythrosin B concentration; EEM spectra of: (c) 6  $\mu\text{M}$  Erythrosin B and (d) 0.6  $\mu\text{M}$  Phloxine B. All data were averaged from three replicate measurements.**

MCR modelling was then implemented using a similar process as for perylene. The EEM data were Rayleigh scatter corrected (Figure A- 20) and then MCR models were built for each polarization setting using CWA data. Initial estimates were obtained using the pure fluorophore spectra. The non-negativity constraint was used on both concentration and pure spectra profiles, as well as normalization to equal length (Euclidean) in order to avoid intensity ambiguities between components. If no other constraints were applied, two components were resolved, but the components were not pure (Figure A- 21).

To solve this issue, we implemented the “*correspondence among species*” option available in MCR-ALS 2.0 toolbox for multiset data structures<sup>314</sup>, to inform the model about the presence/absence of both species in the augmented mode. This constraint, which extends the concept of selectivity and local rank constraints to a multiset structure, is particularly useful for known chemical systems, as it reduces the rotational ambiguities of the solution<sup>228</sup>.

**Table 14: Estimation of rotational ambiguities per component for the best EB-PB-EEM-MCR MCR model. Built using 1: normalization, 2: non-negativity, 5: total trilinearity, and 6: correspondence among species.**

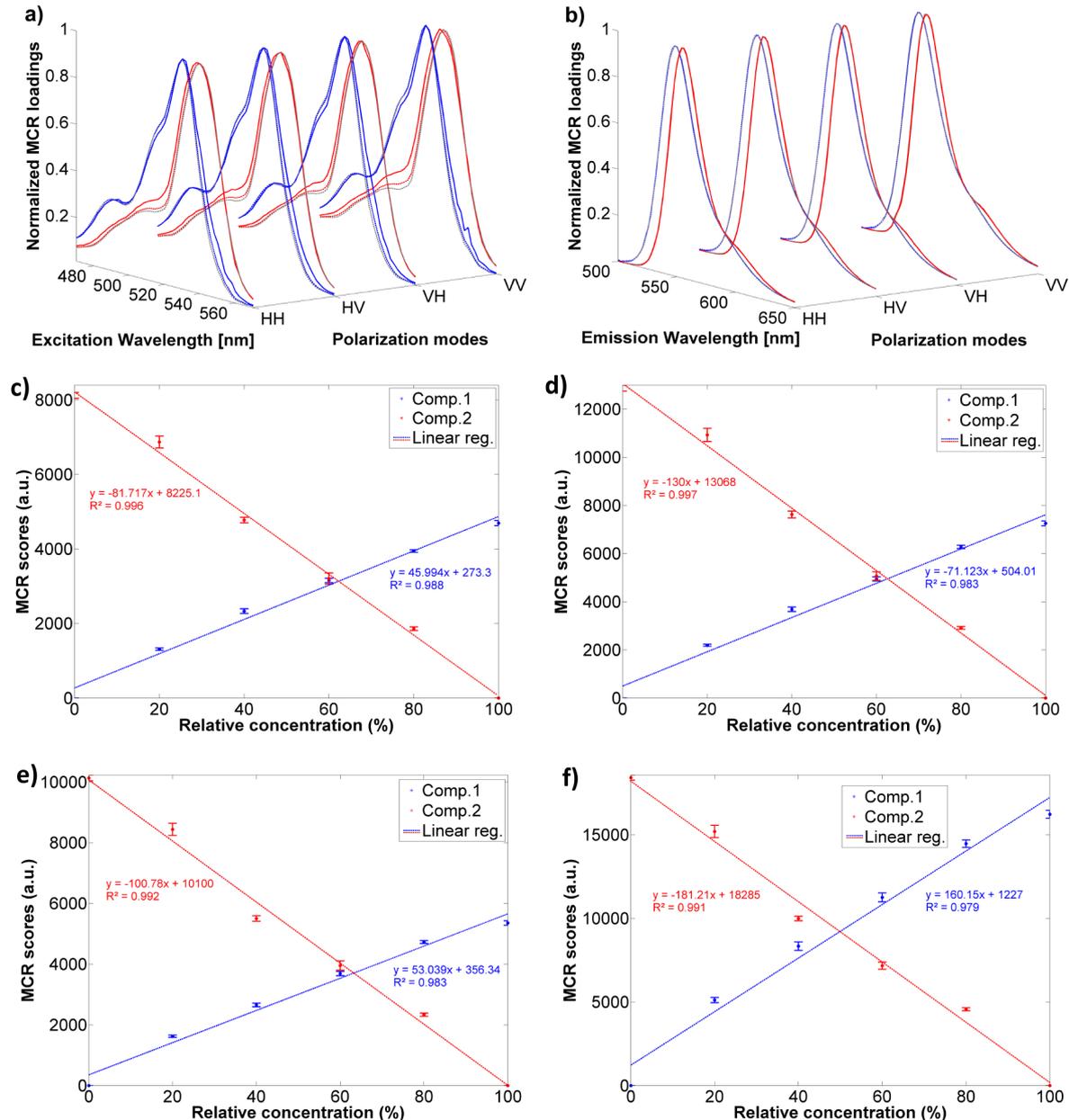
1,2,5,6	Component 1				Component 2			
	HH	HV	VH	VV	HH	HV	VH	VV
<b>f<sub>n</sub> initial</b>	0.453	0.448	0.436	0.596	0.754	0.756	0.768	0.628
<b>f<sub>n</sub> max</b>	0.453	0.448	0.436	0.596	0.754	0.756	0.768	0.628
<b>f<sub>n</sub> min</b>	0.453	0.448	0.436	0.596	0.754	0.756	0.768	0.628
<b>f<sub>n</sub> max-f<sub>n</sub> min</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Rotational ambiguities were removed by application of constraints 1,2,6 (Table A- 9, Appendices). However, we noted that the spectral profiles (especially for PB) were not well resolved, with presence of a small shift in PB, and noise in the loadings (due to residual scatter in data, particularly for the VV, Figure A- 22). This negatively impacted the accuracy of the calculated anisotropy (Table A- 10). The trilinear constraint was applied to rectify the problem, and this improved the shape of recovered components (less noise and no band shifts), which produced more accurate component anisotropy values (Table 16).

**Table 15: Evolution of R<sup>2</sup> and LOF of the MCR models for the EB-PB-EEM data using combinations of constraints: normalization (1), non-negativity (2), total trilinear (5), and correspondence among species (6). Percentage of variance captured by each MCR model component (Fit %X) for 100 iterations.**

EB-PB-EEM MCR model (constraints 1,2,6)				
Comp	HH (%X)	HV (%X)	VH (%X)	VV (%X)
<b>1</b>	26.94%	26.84%	24.36%	48.43%
<b>2</b>	73.01%	73.11%	75.60%	51.45%
<b>Sum</b>	99.95%	99.95%	99.96%	99.88%
<b>Lack of Fit</b>	2.32%	2.29%	2.10%	3.45%
EB-PB-EEM MCR model (constraints 1,2,5,6)				
Comp	HH (%X)	HV (%X)	VH (%X)	VV (%X)
<b>1</b>	26.42%	25.97%	24.35%	47.32%
<b>2</b>	73.37%	73.85%	75.45%	52.35%
<b>Sum</b>	99.79%	99.82%	99.80%	99.67%
<b>Lack of Fit</b>	4.57%	4.25%	4.45%	<b>5.73%</b>

LOF increased with application of the trilinear constraint (Table 15) and was probably due to the presence of residual scatter in the EEM. Wider excitation slits (10 versus 5 nm) were used for the fluorophore mixtures compared to perylene, which generated wider and stronger Rayleigh scatter bands. These were more difficult to correct because of the increased shot noise and was best seen in the VV data as Rayleigh scatter is strongly polarized (Figure A-23, Appendices). Here after scatter correction, significant residual scatter was observed all along the Rayleigh band which correlated with the higher LOF obtained for VV dataset.



**Figure 41: (Top row) Overlap between real excitation/emission spectra (full line-red/blue) and spectra loadings obtained from EB-PB-EEM-126 model (dash-grey line) and EB-PB-EEM-1256 model (dash-red/blue); (middle & bottom rows) MCR scores obtained for each pure component (EB-PB-EEM-1256 model) for the HH (c), HV (d), VH (e), and VV (f) dataset.**

Using this finalised set of constraints, the best MCR model used for anisotropy calculations, were built for each polarized dataset (Figure 41). The first component corresponded to Erythrosin B ( $\lambda_{\text{ex/em}} \text{ max} = 530/548 \text{ nm}$ ) and the second to Phloxine B ( $\lambda_{\text{ex/em}} \text{ max} = 540/554 \text{ nm}$ ). Some slight deviations were observed between excitation loadings and real components (deviations slightly higher for bilinear model compared to trilinear), but the emission loadings overlapped perfectly with the real spectra. Component 1 explained more variance in the VV data compared to other polarization settings (Table 15) which was because Erythrosin B had a larger anisotropy ( $r = 0.243$ ) than Phloxine B ( $r = 0.054$ ) because of its much shorter fluorescence lifetime, 75 ps vs.  $\sim 1 \text{ ns}$ .

We are assuming that because both fluorophores have similar molecular masses, and structures, that they will also have similar rotational correlation times in this solvent system<sup>78</sup>. Erythrosin B was thus less depolarized and has stronger emission intensity in the VV configuration. Component 1 deviated slightly from linearity, which was due to the small self-absorption IFE mentioned above, and the effects was largest in the VV configuration. MCR scores (Figure 41) showed that component 2 (PB) varied linearly in all configurations, indicating that there was no detectable IFE or FRET. This was expected because Phloxine B concentration was an order of magnitude lower than Erythrosin B.

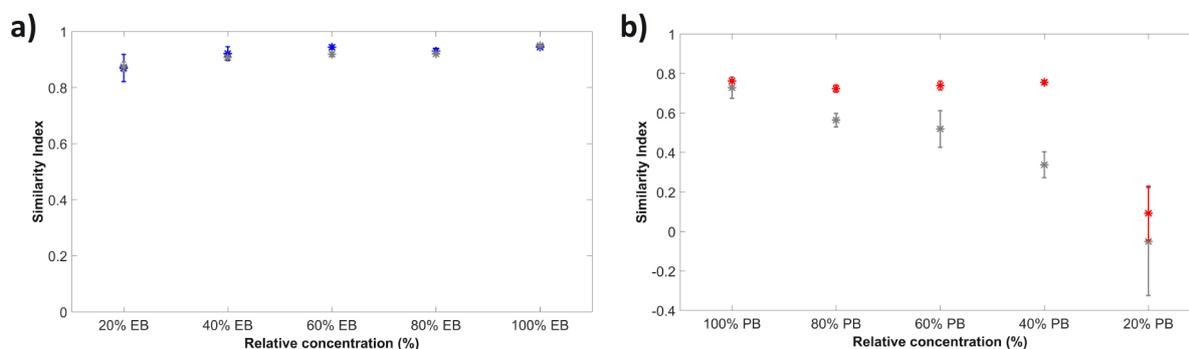
The data from the best model was used to calculate individual component anisotropy. Firstly, the *aniso*-EEM plots for each resolved component for all mixtures were extracted (see Figure A- 24, Appendices, for some examples). It appeared that the *aniso*-EEM plots in case of mixtures were smooth and uniform for both compounds over the full  $S_1$  emission band region ( $\lambda_{\text{ex/em}} = 520\text{--}620/470\text{--}570 \text{ nm}$ ), and correlated well with the data from the pure fluorophore solutions. By averaging the data across the entire  $S_1$  region, a very good agreement was obtained between recovered anisotropy and values from pure EB and PB in solution (Table 16), confirming that ARMES could accurately retrieve the anisotropy values of pure fluorophores in mixture, without any rotational ambiguities.

We quantified the degree of similarity between pure and recovered *aniso*-EEM plots using a similarity index<sup>317</sup> (Figure 42). Due to variations in the size of the *aniso*-EEM plots between samples (because of the 10% threshold setting, section 4.3.2) we calculated the similarity index using a specified area that was the same for all samples and contained no NaN valued data points (Figure A- 25).

**Table 16: Real and modelled (calculated from MCR components) averaged anisotropy values (EB-PB-EEM-1256 model, with the trilinear constraint), for each mixture concentrations, across the  $S_1$  ( $\lambda_{ex/em}$ ) = 520–620/470–570 nm bands.**

	Modelled anisotropy $\pm$ STD	
	Component 1 (EB)	Component 2 (PB)
Pure PB EB solutions, real anisotropy values.	<b><math>0.243 \pm 0.002</math></b>	<b><math>0.054 \pm 0.002</math></b>
0% / 100%	/	<b><math>0.049 \pm 0.001</math></b>
20% / 80%	$0.244 \pm 0.006$	$0.043 \pm 0.001$
40% / 60%	$0.246 \pm 0.004$	$0.045 \pm 0.002$
60% / 40%	$0.238 \pm 0.003$	$0.048 \pm 0.002$
80% / 20%	$0.234 \pm 0.002$	$0.070 \pm 0.008$
100% / 0%	<b><math>0.241 \pm 0.003</math></b>	/
Avg/std	$0.241 \pm 0.005$	$0.051 \pm 0.011$

For EB, an index of  $>0.90$  for most samples indicates that the recovered components for all polarization configurations were good and that the recovered *aniso*-EEM plots were accurate (to within  $\pm 4\%$ ). We also note that the agreement was marginally better for the trilinear model, confirming that trilinear model was slightly better at recovering the anisotropy values. However, for PB, because of its much weaker absorption, the excitation loadings differed slightly from real spectra at wavelengths below 520 nm (Figure 41). The lower magnitude anisotropy of PB ( $\sim 22\%$  of the EB value) also makes accurate anisotropy component recovery in all polarization configurations a more challenging task, because shot noise from the second fluorophore will cause significant problems as seen in Figure 42-b where the difference between real and recovered *aniso*-EEM plots gets larger as the proportion of EB fluorescence (and its constituent shot noise) increases. All of these effects resulted in less accurate recovered anisotropies ( $\sim \pm 0.01$ ) and the similarity index only gave good values for  $\lambda_{ex} > 520$  nm, where absorbance was greatest (Figure A- 25).



**Figure 42: Similarity indices calculated between anisotropy plot extracted from real pure compounds in solution, and the recomposed components. (Left) Results for EB, with results from trilinear model (blue) and bilinear model (grey), (Right) Data for PB, with results from the trilinear (red) and bilinear models (grey).**

## 4.5 Conclusions

These results show that it is possible to accurately extract the anisotropy of individual components using MCR from relatively simple small-molecule fluorophore mixtures once appropriate data organisation and constraints have been applied. However, we saw that replacing TFP with UV transmitting filters (dWGP) resulted in the elimination of the spectral reshaping previously observed in TSFS measurements<sup>14</sup>. This had consequence in terms of data structure which made CWA not suitable anymore for TSFS. Instead, we found that SWA with non-negativity, and unimodality constraints could generate sufficiently accurate anisotropy data, at least on the perylene data.

Overall, however, MCR modelling of TSFS data was less efficient than MCR modelling of EEM data which was, for example better able to cleanly resolve the impurity in the perylene solutions. A pre-requisite for the latter was efficient removal of the 1<sup>st</sup> order Rayleigh scatter, which was achieved by modelling<sup>222</sup> rather than interpolation<sup>306</sup>, increasing the working range of perylene as multidimensional anisotropy standard. For the EEM-MCR modelling the optimal settings were to use normalization, non-negativity, and total trilinearity constraints, with correspondence among species being required for mixtures where the species had similar emission intensities. Recovery of component anisotropy via chemometric modelling was also achieved and was accurate to  $\pm 4\%$  for relatively large magnitude anisotropies (*e.g.* Perylene S<sub>1</sub> and EB). For the lower magnitude anisotropy ( $<0.06$ ) fluorophores accuracy decreased ( $\pm 20\%$ ) due to the impact of shot noise from the overlapping other component and relative measurement error. This might be rectified, by changing the measurement conditions to improve the signal to noise ratio, which should result in less variation in the recovered components and thus the calculated anisotropies. Despite this, the fact that these very low anisotropy components were recovered was still very significant.

Being able to recover the spectral and anisotropy information using MCR with either TSFS or EEM measurements provides a useful tool for accurately characterizing molecular systems where the fluorophores are not interacting (FRET or IFE) with each other and where significant size, mobility, or lifetime changes are occurring (*e.g.* fluorophore diffusion in polymer matrixes). The two sample systems used here also provide very simple and robust test systems for validating the TSFS/EEM measurements and chemometric data analysis elements of ARMES. The next stage is to see if it is feasible to analyze/model the aggregation of insulin, which is a much more complex situation due to FRET interactions between fluorophores.

## **5. Elucidating the intrinsic fluorescence of human insulin and its different association states**

This chapter is divided in three sections. The first part is a general introduction about insulin, covering aspects of its discovery, to the existence of natural soluble oligomers in solution and their molecular structure. It also points out the need for a new analytical method capable of rapidly discriminating these oligomers and ends with a particular focus on tyrosine and insulin intrinsic steady-state fluorescence, explaining how this method has been used to date to characterize that molecule. The second part starts by showing some results that can be achieved from UV-Vis spectroscopy before explaining with the monomer data, that the presence of rank deficiencies makes the chemometric resolution of fluorescence a difficult task. The last section demonstrates how the ARMES measurement can be split into several tools that are all very useful to characterize qualitatively or quantitatively these soluble oligomers. We also show how the rank deficiency issue was overcome, resolving for the intrinsic fluorescence of insulin.

### **5.1 Introduction**

A protein can comprise of tens (insulin) or hundreds of amino-acid residues (IgG), of which many can be fluorescent. These fluorophores are usually located in close proximity (<15 nm) to each other and interact photophysically (FRET, electron transfer, quenching, variable solvent exposure etc.) to generate complex emission that is difficult to resolve due to the high spectral overlap<sup>28</sup>. ARMES that combines anisotropy ( $r$ ) with MDF measurements and factor-based methods appeared as a promising tool towards their resolution<sup>14</sup>.

Protein analysis by ARMES is also very attractive as it uses the molecule's intrinsic fluorescence and is therefore a fast, non-destructive, sensitive and inexpensive method. It is also non-invasive and preserves the structure and behavior of protein native states. In this chapter, we use this method to elucidate the intrinsic fluorescence of human insulin, with a particular focus on the natural soluble oligomers that insulin can form in solution. Monomeric insulin is known to exist in equilibrium with oligomers like dimers or hexamers, the equilibrium being highly dependent on the pH, temperature, and the ionic strength of the solution<sup>74</sup>. This chapter introduces general concepts about insulin and advances some

hypotheses to explain why its intrinsic fluorescence was never resolved until now. Also, due to the key location of fluorophores in regions responsible for multimerization, slight spectral shifts are produced that enable one to break the linear dependencies between fluorophores and start to open the way towards the resolution of the various families of tyrosine residues. Robust models are developed to discriminate these soluble aggregates in solution.

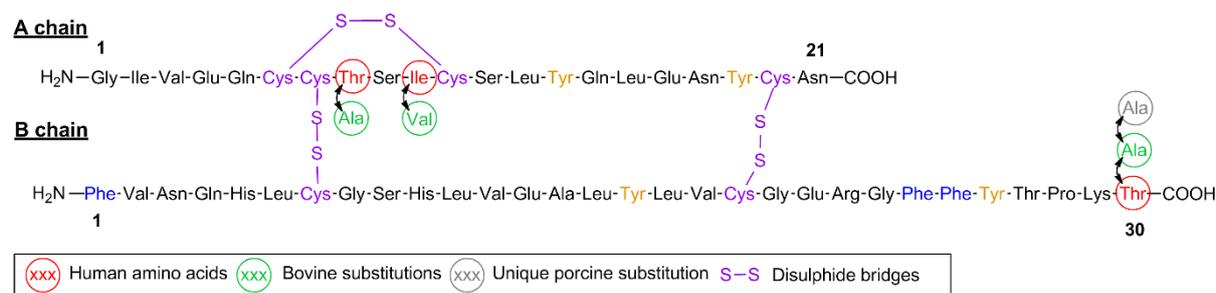
## 5.2 The molecule of insulin

The insulin molecule was isolated for the first time in 1921 by Banting and Best and has since been produced and commercialized to treat human diabetes. According to the World Health Organization (WHO) that recently predicted a diabetes epidemic for the near future<sup>322</sup>, it is clear that insulin will become even more important. For further information about the history behind insulin's discovery, the evolution of insulin formulations as well as a short description of the insulin market and what it is expected to become, the reader is invited to consult Appendix 7.10.

### 5.2.1 Insulin structure

#### 5.2.1.1 The primary structure of insulin

Insulin (in its monomeric form) is a disulphide-bridged peptide of 51 residues, characterised by a molecular weight of 5808 Da for the human form, 5734 Da for the bovine form, and 5778 Da for the porcine one. This indicates very small primary structure differences between insulin sources as shown in Figure 43:



**Figure 43. Illustration (ChemDraw) of the primary structure of monomeric human insulin, along with substitutions encountered in bovine and porcine forms of insulin. Tyrosine (Tyr) and phenylalanine (Phe) residues are highlighted in orange and blue respectively.**

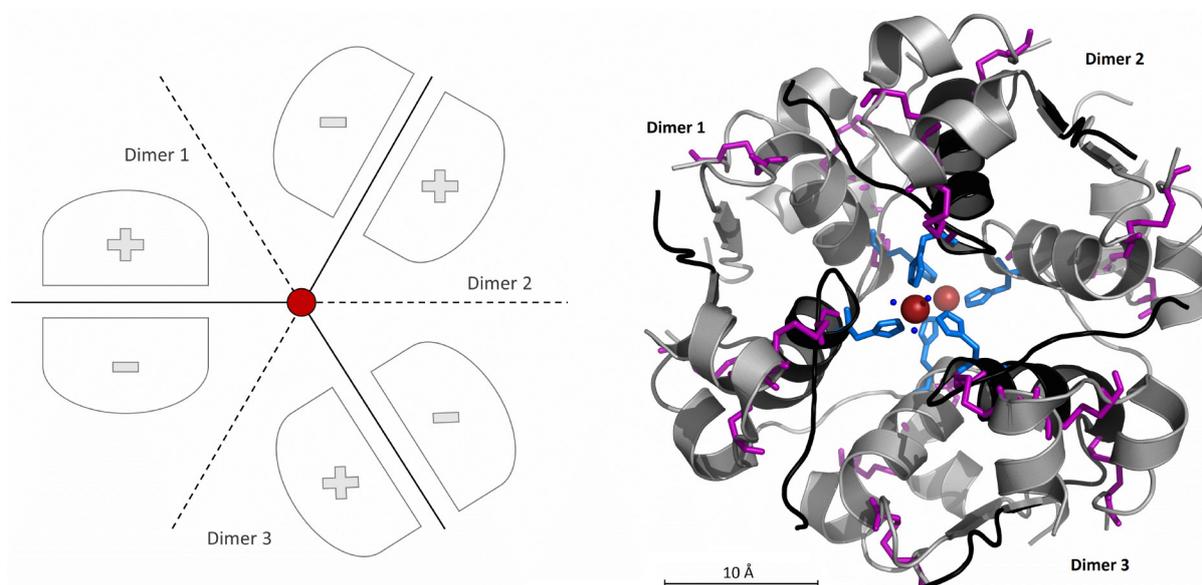
Human insulin differs from bovine and porcine insulin respectively by 3 and 1 residues, explaining why both animal forms were used in the past to treat human diabetes, but also why the protein fold is essentially maintained between sources. Insulin is indeed considered as one of the most primitive proteins and seems to have been highly conserved during evolution. The

amino-acid substitution rate for insulin was estimated to be  $10^{-9}/\text{locus}/\text{year}^y$ , resulting in insulin forms that vary by only 38% between human and hagfish<sup>323</sup>, one of the earliest vertebrate ancestors.

Paradoxically, even if insulin is believed to have a structure that originated more than one billion years ago<sup>324</sup>, its primary structure was only elucidated in the early 1950s after the crucial work of Sanger, and more details about this discovery are given in Appendix 7.11.

### 5.2.1.2 Elucidation of the crystal structure of insulin

It was only in 1969 that the first complete crystal structure of insulin was resolved by Hodgkin *et al.*<sup>325</sup>, confirming that insulin crystallises as a hexamer formed from three dimers coordinated around two Zn atoms (called 2-Zn insulin, see section 7.12 for historical details).



**Figure 44: Illustration of porcine 2-Zn insulin hexamer assembly with (left) diagram inspired from Blundell<sup>326</sup> showing three antiparallel (+/-) dimers arranged along a twofold symmetry axis (full line/dash) in the plane of the page, orthogonal to the threefold axis shown with red dot and pointing towards the reader. (right) Hexamer structure from PDB file 4INS<sup>327</sup> (plotted using PyMol), showing in purple the disulfide bridges, in light blue the B10 Histidine residues; both Zn atoms are represented as red-brick spheres, and the water molecules coordinated to first Zn atom as dark-blue spheres. For each dimer, the A chain and B chain are coloured respectively in black and grey.**

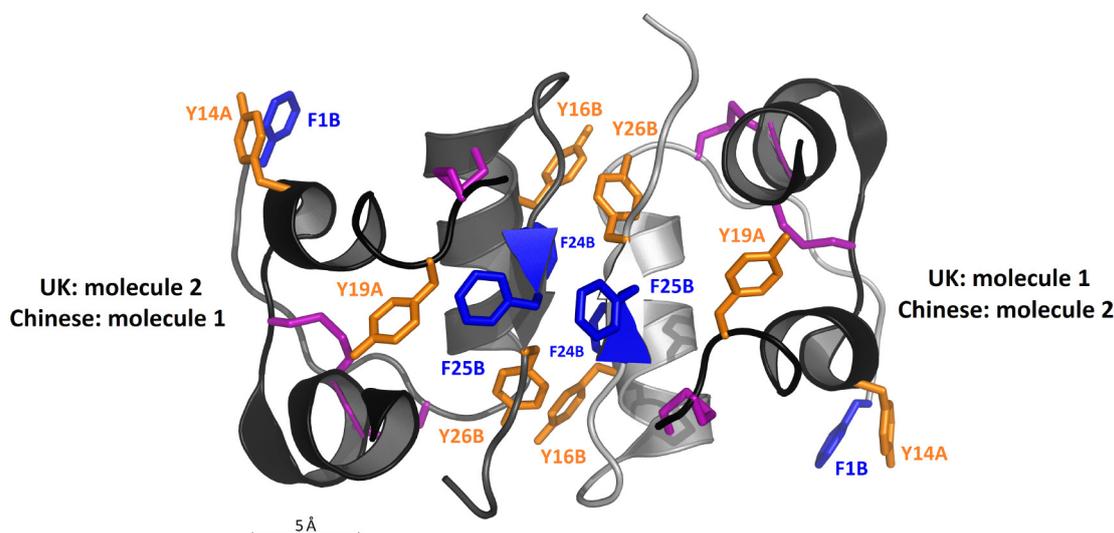
The structure chosen to represent the best 2-Zn pig insulin is the one published by Baker in 1988<sup>327</sup>, mainly because it is the most cited insulin structure and because it gathers all the main authors that contributed to insulin structure (~ final structure). The core of the information though is similar to what was shown in 1969, *i.e.* that three dimers are arranged around two Zn

<sup>y</sup> A locus is a term normally applied to indicate the precise location of a gene on a chromosome but can also be applied to protein field to indicate a precise location in amino-acid sequence.

atoms and related by a threefold symmetry axis (Figure 44), giving to the insulin hexamer a symmetrical geometry<sup>328</sup>.

Each dimer is arranged along a twofold symmetry axis, orthogonal to the threefold axis. The hexamer has a flattened spheroid shape (or torus shape) estimated to have a diameter of  $\sim 50 \text{ \AA}$  and a height of  $35 \text{ \AA}$ <sup>326</sup>. Both Zn atoms are located along the threefold axis,  $\sim 16 \text{ \AA}$ <sup>327</sup> apart from each other. Each Zn is octahedrally co-ordinated with the imidazole groups of B10 histidine from the three dimers and three water molecules (Figure 44). It is important to also mention that the 2-Zn rhombohedral crystal is the form naturally adopted by insulin when stored in the human body<sup>329</sup>. The pancreas is known to be the organ with the highest zinc concentration<sup>330</sup>, and each  $\beta$ -cell of Langerhans islets contains more than 10,000 secretory granules<sup>331</sup> where hexamers are densely stored approximately  $50 \text{ \AA}$  apart<sup>332</sup> as crystalline arrays, waiting for release into the blood.

Looking in more detail at the hexamer structure, it was realised that the perfect twofold axis that was predicted for the dimers was actually only approximate<sup>325-326,333</sup>, due to the forces engaged in crystal packing<sup>334-335</sup>. It was obvious that for each monomer, the side-chain of residue Phe25B<sup>z</sup> had different orientations (Figure 45), lying against Tyr19A for the left molecule, or lying across the twofold axis on the right one.



**Figure 45 : Typical asymmetric insulin dimer found in 2-Zn porcine hexamer insulin (PDB file 4INS). Tyrosines (Y) and Phenylalanines (F) are coloured respectively in orange and blue; disulfide bridges are in purple. A and B chains for both monomers are coloured respectively in black and dark/light grey. The UK and Chinese conventions used to name both molecules are shown on each side.**

<sup>z</sup> Phe25B can be read as Phe: Phenylalanine, 25: residue number, B: chain B. This convention, often used in protein chemistry, is adopted in the thesis text whereas F25B (shorter) is used in figures. Another important abbreviation is Y or Tyr for Tyrosine.

Dimers appeared thus to be asymmetric (or non-crystallographic) and a convention<sup>336</sup>, reversed between UK and China and based on direction of Phe25B, was adopted to name these two molecules (Figure 45).

There are a few cases however where the dimers can have a real twofold symmetry axis (crystallographic). This is encountered particularly when divalent cations are absent from solution, enabling the dimers to form cubic crystals. For example in the absence of zinc, porcine insulin adopts this type of structure at alkaline pH ( $\sim 9$ )<sup>337-338</sup>. It is interesting to note that the residues involved in the formation of the symmetrical dimers (thus present in the region of interaction) were found to have the same orientation as in the non-symmetrical dimer. The monomer forming the symmetrical dimer is also very similar to the molecule 2 (UK convention). For a very primitive species like hagfish, Cutfield *et al.*<sup>339</sup> found out that insulin does not crystallise as a hexamer at pH 6 but rather as the symmetrical dimer, with a monomer conformation also close to molecule 2 from the 2-Zn porcine insulin.

This latter conformation is considered the most stable<sup>335,338</sup> for monomeric insulin, explaining why it is adopted by most of these subunits in crystals. But it also seems to be the conformation adopted by insulin in blood circulation<sup>340-342</sup> (*i.e.* in solution) as suggested by crystal packing structure analysis<sup>335</sup>, molecular dynamics simulations<sup>343</sup>, or energy minimization studies<sup>344</sup>.

### 5.2.1.3 Is the insulin solution structure much different from crystal structure?

Insulin is stored as hexamers in islets  $\beta$ -cells.  $\beta$ -cells are highly connected to blood vessels through permeable tissues made up of small pores called fenestrae<sup>340</sup> and this enables them to “sense” the plasma nutritional state. When glucose levels are too high, a complex coordination between the million islets allow  $\beta$ -cells to release in a pulsatile manner the appropriate amount of insulin hexamers into the blood<sup>345</sup>. The huge dilution following exocytosis and blood release<sup>346</sup>, as well as a change in pH and  $Zn^{2+}$  concentration<sup>347</sup>, causes dissociation of hexamers into monomers which are considered to be the active form and the only form present in the blood<sup>328,348-350</sup>. To be more exact, dimers are probably also present in the portal or peripheral blood (initial release phase), but progressive dilution leads to predominance of monomers in circulating blood: it is generally considered that the intragranular insulin concentration is  $\sim 120$  mM<sup>351</sup>, whereas circulating blood concentration is only  $\sim 0.1$  to  $1$  nM<sup>352-353</sup>. Monomeric insulin also seems to be the only form<sup>aa</sup> recognised by

---

<sup>aa</sup> Insulin hexamers have for example been proven to be ineffective to form the insulin-receptor complex<sup>354</sup>.

insulin receptors, as the formation of the insulin-receptor complex involves many of the hydrophobic residues engaged in dimer interface<sup>355-356</sup>.

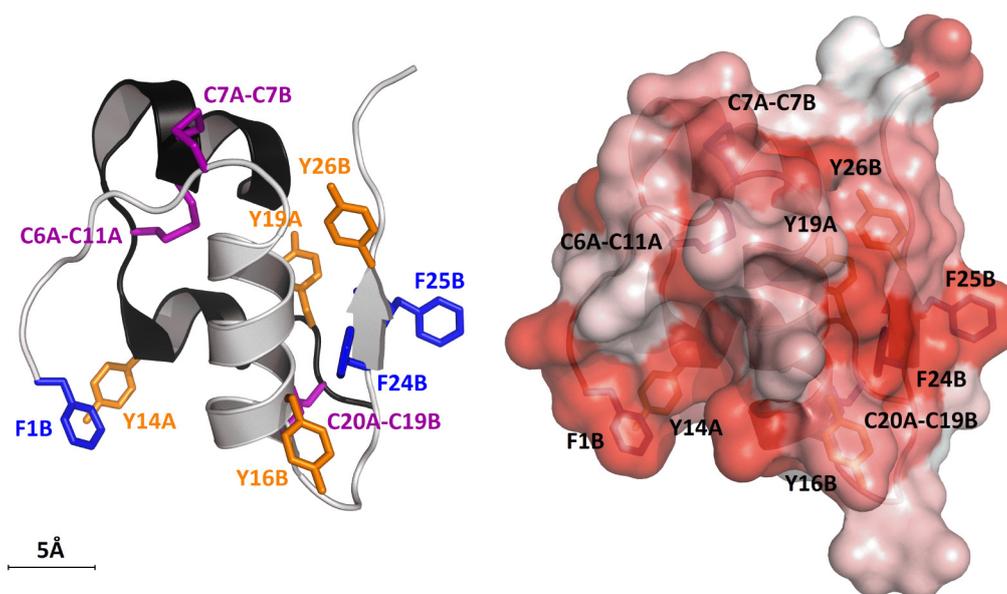
Because insulin is transported and physiologically active as the monomer, it is critical to know if the conformation derived from the X-ray analysis of hexamer crystals exists also in solution and in the monomeric state or if crystal packing forces modified it<sup>357</sup>. This is also important for the work developed in this chapter, as we need structures to explain changes in intrinsic protein fluorescence, and we want them to be reliable when cross referencing with our experimental work. To find more details about the crystal and solution structures of insulin in the various oligomeric states relevant to this chapter, the reader is invited to read section 7.13. The main points to remember are:

(i) insulin's native fold is maintained at low pH but also across the formation of higher order aggregates; (ii) monomers are mainly present at pH 2 in acetic acid and can form at the same pH in mineral acids dimers, that are highly similar to those found in hexamer structures assembled at neutral pH<sup>358</sup>; (iii) there are differences between solution and crystal structures, especially more flexibility is seen in free monomer in solution. Disorder is often observed at the N- and C-terminus of B chain and N-terminus of A chain, synonymous of high flexibility crucial in receptor recognition<sup>359</sup>; (iv) Conformations 1 or 2 have overall very little or no consequences on the distance between Tyr residues; the position of side chain residues like PheB25 can vary but the quantum yield associated with Phe residues are too low to have consequences on emission; (v) whether crystal or solution structure is considered, insulin tertiary structure is overall stable<sup>349,360</sup> and maintained in highly ordered aggregates, with crystal packing forces having little impact on insulin folding.

Overall and from reviewing all the sources (section 7.13), it was thus acceptable to use the crystal structure (PDB file 4INS) to explain our fluorescence observations from different solutions.

#### **5.2.1.4 The structure of monomeric insulin and the existence of T and R states**

The monomer structure presented below was extracted from 2-Zn porcine hexameric structure. The molecule 1 conformation was chosen for representation because 20% acetic acid was used in the thesis to prepare monomer solutions (Figure 46). Looking at the arrangement of chains, it seems that the B chain tends to wrap the compact A chain as result of folding by disulfide bridges, which is often seen as a complex arrangement for such a small protein<sup>340</sup>.



**Figure 46 :** (Left) Representation of monomeric insulin in molecule 1 conformation, extracted from 2-Zn porcine hexamer crystal (PDB file 4INS) with tyrosines (Y), phenylalanines (F) and disulfide bridges (C) shown. A and B chains are coloured in black and light grey respectively. (Right) Same molecule for which the surface was coloured according to Eisenberg hydrophobicity scale<sup>361</sup>; the red corresponds to highest hydrophobicity, highlighting the right face and bottom left enclave, responsible respectively for dimer and hexamer formation.

The A chain of 21 residues is organised around two  $\alpha$ -helices<sup>bb</sup> arranged as two antiparallel stretches<sup>340,362</sup> (GlyA1-ThrA8 and SerA12-AsnA18 represented as black chains, Figure 46). These stretches are brought to the same plane by the presence of a turn (SerA9-CysA11) and a covalent internal bridge (CysA6-CysA11), which have for other consequence to set the side chains IleA2 and TyrA19 into van der Waals contact, a contact essential for insulin structure as it maintains the GlyA1-ThrA8 close to the hydrophobic core<sup>363</sup>.

The structure of the B-chain is a bit more complex and composed of a central  $\alpha$ -helix running from residues SerB9 to CysB19, flanked by two  $\beta$ -turns<sup>cc</sup> (CysB7-HisB10 and GlyB20-GlyB23). The first turn brings the extended B1-B6  $\beta$ -strand in contact with the central region of the A chain, forming side-chain hydrogen bonds that are important for fold stability. Examples of these are LeuB6-CysA6, GlnB4-CysA11<sup>327,364</sup> or HisB5 that is packing between CysA7 and IleA10<sup>365-367</sup>. Other inter-chain hydrogen bonds, such as TyrA19-PheB25<sup>368</sup> and AsnA21-GlyB23<sup>369-370</sup>, as well as ionic interactions<sup>370</sup> also contribute to fold stability. The

<sup>bb</sup> The  $\alpha$ -helix is generated by hydrogen bonds between the N-H and C=O backbones of amino-acids located a couple of positions away to each other. It results in right-hand helical structures that are represented as distinct spirals in Figure 46.

<sup>cc</sup> A  $\beta$ -turn is a very common motif, made of four residues and responsible for change in backbone direction. It is defined by presence of H-bond between N-H and C=O group of first and last of four amino-acids.

second turn forms the V-shape observed with the extended  $\beta$ -strand PheB24-ThrB30 and enables a close anti-parallel packing against the B-chain  $\alpha$ -helix. A  $\beta$ -sheet<sup>dd</sup> structure is formed by residues PheB24-PheB25-TyrB26<sup>372</sup> (represented as an arrow Figure 46), where residues PheB24 and TyrB26 interact with LeuB15/LeuB11/ValB12<sup>340-341,366-367</sup> through hydrogen bonds and van der Waals interactions; LysB29 and ThrB30 are on the other hand less well ordered<sup>362</sup>. Native insulin is thus a compact globular protein with 58%  $\alpha$ -helical and 6%  $\beta$ -sheet structure<sup>373</sup>.

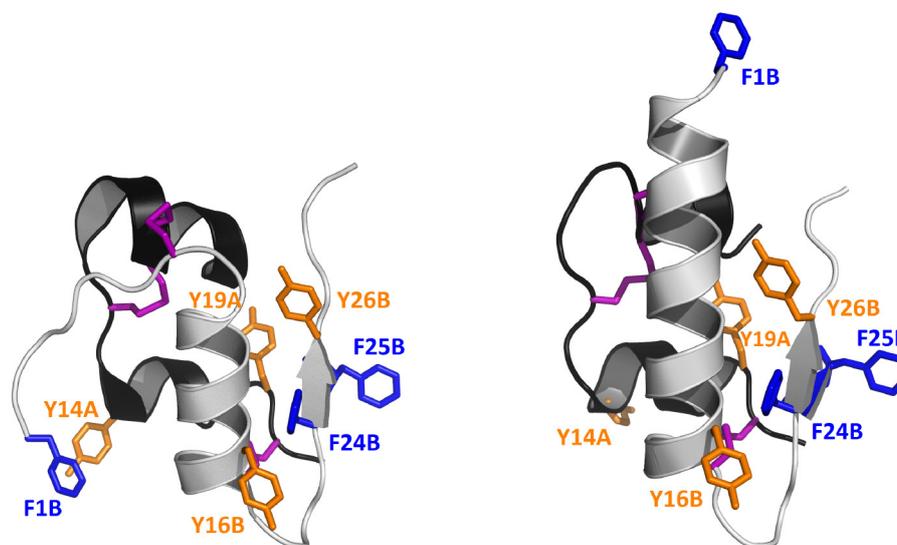
Insulin also has in its centre a strongly hydrophobic core which is crucial for bringing the A and B chains together. The residues surrounding this core form a dense cluster of hydrophobic interactions that result in a tightly packed structure that are very important for protein folding and stability, and for protein-protein interactions<sup>374</sup>. In the case of insulin, this hydrophobic core was highly preserved during evolution and is composed<sup>327,341</sup> of residues that are completely buried (IleA2, LeuA16/B6/B11/B15, TyrA19 and AlaB14). These are in contact with other hydrophobic residues located at the edge of the core, regarded as being peripheral (IleA13, ValA3/B12/B18, PheB24 and TyrB26), which themselves are in contact with the rest of the hydrophobic residues<sup>327</sup> (including PheB1, PheB25 or LeuB17<sup>370</sup>). All these peripheral residues are partially exposed and contribute to the hydrophobic faces involved in the formation of dimers and hexamers, and possibly also in receptor binding (Figure 46, middle). A and B chains are also brought together by the buried CysA6-CysA11 and CysA20-CysB19 disulfide bridges (CysA7-CysB7 is exposed)<sup>327</sup>, that are strongly hydrophobic and known to participate strongly in the stability of globular proteins in general<sup>375</sup>.

Even if the structure of monomeric insulin seems well defined and the fold always maintained, the reality is a bit more complex. Schlichtkrull<sup>376</sup> was the first scientist in the 1950s to find that insulin hexamers needed at least two Zn atoms to crystallise. However, he also discovered that a second form of rhombohedral crystals appeared when the concentration of sodium chloride was greater than or equal to 6% (w/w)<sup>ee</sup>. This second hexameric form crystallized around four Zn atoms and was believed for a long time to be very similar to the 2-Zn structure as it also comprises asymmetric dimers<sup>377</sup>. It was only in 1976 that Bentley *et al.*<sup>378</sup> elucidated fully the structure of this hexamer, and discovered a surprising large change in the conformation of only one of the monomer per dimer. The N-terminal end of the B-chain

<sup>dd</sup>  $\beta$ -strands are encountered whenever several amino-acids adopt an extended conformation. A  $\beta$ -sheet is formed when  $\beta$ -strands overlap, forming H-bonds between strands which can run parallel or anti-parallel to each other. After  $\alpha$ -helix, this is the second most encountered motif in proteins<sup>371</sup>.

<sup>ee</sup> In our experiment, we used 0.1 M NaCl that corresponds to 0.58% (w/w) which is well below 6%.

(PheB1-GlyB8) that normally exists as an extended strand, seemed to become a simple extension of the central  $\alpha$ -helix, generating a long PheB1-CysB19  $\alpha$ -helix (Figure 47).



**Figure 47: Representation of T (left) and R (right) state of monomeric insulin, built respectively from PDB file 4INS<sup>ff</sup> and 1ZNJ.**

Residues in the N-terminus can move up to 30 Å<sup>383-384</sup>, resulting in the loss of important H bonds<sup>364</sup> such as PheB25-TyrA19, LeuB6-CysA6, or GlnB4-CysA11, and generating elliptical cavities around HisB5 as the CysA7-CysB7 bridge also rearranges. The monomeric sub-units forming the 4-Zn hexamer are thus too different to preserve the approximate two-fold symmetry observed in the 2-Zn molecule. To avoid confusion, this new PheB1-CysB19  $\alpha$ -helix conformation was called the monomeric R (for Relaxed) state, and the one previously encountered the T (for Tensed) state<sup>385-386</sup>. The 2-Zn and 4-Zn structure were then named accordingly as T<sub>6</sub> and T<sub>3</sub>R<sub>3</sub>, both structures existing in reversible equilibrium<sup>387</sup>.

To find more details about the consequences of the T/R transition on insulin hexamer, as well as the potential importance of this transition in biology (insulin-receptor binding), the reader is invited to read section 7.14. Related to this chapter work, what is important is that due to experimental conditions used<sup>gg</sup>, we can say that we are working only with the T<sub>6</sub> 2-Zn

<sup>ff</sup> We could not find a satisfying PDB file for the wild-type 2-Zn T<sub>6</sub> human insulin at room temperature, for the simple reason that 4INS structure derived from porcine insulin is already considered to be the best approximation for it. The only T<sub>6</sub> structures available are else from dehydrated crystal<sup>379</sup> or obtained at 120K<sup>380</sup>. In both cases 4INS structure was used as reference, to observe the modifications brought by these experimental conditions. Other PDB files are available but use different divalent cations<sup>381</sup>, or show in the structure additional  $\beta$ -sheet content (*i.e.* 4EX0) never mentioned in any other papers<sup>382</sup>. For that reason and considering the extreme similarity between human/porcine sequence, we used 4INS in the thesis as the most accurate representation of the human structure.

<sup>gg</sup> In our experiment, we used 0.1 M NaCl that corresponds to 0.58% (w/w) which is well below 6%. Phenol was not added to the solution either. In these conditions, the T<sub>6</sub> configuration is more stable than R<sub>6</sub> and is the form

hexamer. The T-state, which was found to be the form stabilised in acidic pH<sup>389</sup>, is also adopted by monomers or dimers when they are the predominant species in solution. In fact, unless they form hexamers, the R-state has never been observed for monomers or dimers<sup>328,350,386,388,390</sup>.

### 5.2.2 The natural aggregation surfaces of insulin

The insulin monomer possesses two hydrophobic faces (Figure 46) that contribute to the formation of multimers like dimers and hexamers. These soluble and non-covalent oligomers (also called the association states of insulin) are brought and held together by many different types of weak interactions such as hydrogen bonding, hydrophobic effect, ionic and van der Waals forces. It is important at this stage to differentiate them from the insoluble, linear, sometimes “thick” fibrils (tens of nm in diameter) that insulin can produce when exposed to destabilising conditions<sup>71,367,391</sup>. The appearance of fibrils was observed originally for insulin at low pH and elevated temperatures ( $> 80^{\circ}\text{C}$ )<sup>392-394</sup>, but a variety of other parameters such as the presence of hydrophobic air-water interfaces during agitation<sup>395-396</sup>, organic co-solvents<sup>397-398</sup>, high concentration of salt<sup>399</sup>, or denaturant<sup>400</sup> have also been reported to facilitate that conversion. Whereas the various association states of insulin maintain the protein fold, fibrils that are typical of amyloidogenic peptides<sup>401</sup> are believed to come from the partially unfolded monomer<sup>71,367,391</sup>.

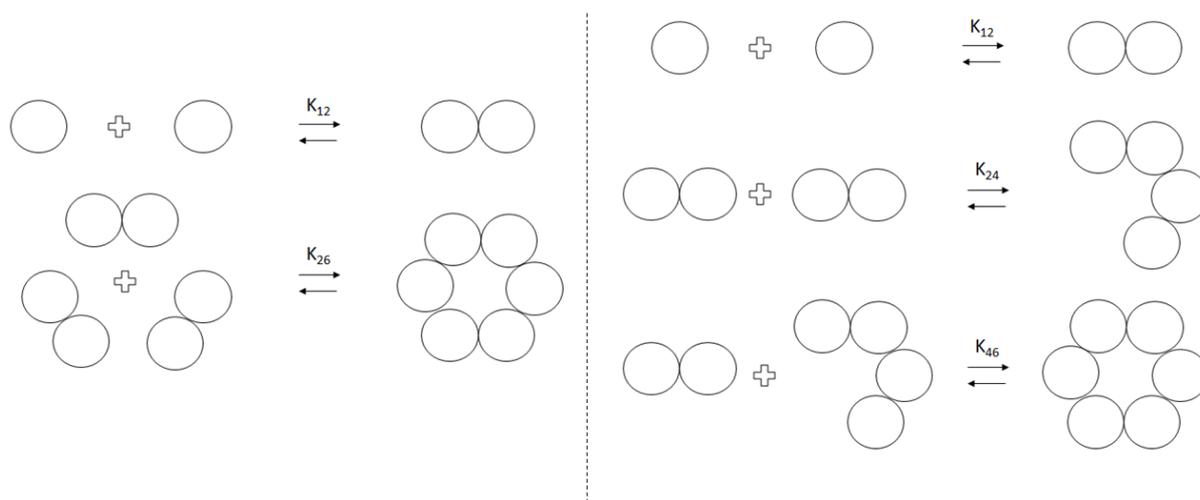
Insulin was found to be fully monomeric at neutral pH if the concentration was below  $1 \text{ ng.mL}^{-1}$  ( $\sim 10^{-10} \text{ M}$ )<sup>349</sup>. At higher concentrations in acid pH, or neutral pH, but in the absence of zinc, dimers become the predominant species and form mainly hexamers if zinc is present<sup>329,341,349,360,402</sup>. Ultimately the equilibrium that exists between species in solution depends on concentration and on pH, temperature, ionic strength, presence of divalent cations etc. explaining why different association constants are found in the literature. For example, the dimer-hexamer association constant  $K_{2,6}$  was found to be 400 times higher<sup>403-404</sup> in the presence of zinc ( $15 \times 10^{10} \text{ M}^{-2}$ ) than without ( $4 \times 10^8 \text{ M}^{-2}$ ). Sedimentation equilibrium studies, particularly suited to measure association equilibria<sup>69</sup>, also gave different  $K_{1,2}$  values for monomer-dimer equilibrium at pH 2 ( $8 \times 10^3 \text{ M}^{-1}$ ) vs pH 8 ( $2 \times 10^5 \text{ M}^{-1}$ )<sup>349,360,402</sup>.

Several association models were proposed to best fit the sedimentation data and try to deal with the concentration dependency of insulin in solution. From all these models, only one involving monomers, dimers, and hexamers as the main species<sup>403-405</sup>, worked best

---

present in solution<sup>388</sup>. Also the insulin batch used in this thesis contained 0.4% (w/w) of zinc (see section 7.16 in Appendices), accounting according to other studies<sup>17,72,75</sup> for two molecules of zinc for six insulin molecules (= 2 Zn T<sub>6</sub> insulin).

independently of pH or presence of zinc. It is still the most widely accepted<sup>69</sup> and a variant, including the tetrameric state<sup>360,402,406-407</sup> seen by Uversky *et al.* as transiently populated<sup>17</sup>, was similarly proposed (Figure 48).



**Figure 48: Illustration of the models believed to best represent the association of wild insulin in solution. Monomers, dimers and hexamers are the key species. The model on the right is a variant including a partially populated tetrameric state.**

Not surprisingly, what matters is that the dimer and not the monomer is seen as the additive species. Models involving for example the addition of one monomer at a time to yield bigger assemblies, were proven to give negative values for some constants, a poorer fit and a bigger error than the models proposed (Figure 48)<sup>329,360</sup>, meaning that they do not represent well insulin polymerization<sup>17</sup>. Similarly, models going through trimer or pentamer states gave negative values for polynomial coefficients and a larger error of fit<sup>402</sup>, making their existence unlikely.

The reason why tetramers are suspected to be transiently populated is because spectroscopic features specific to that state could not be found<sup>17</sup>. The near-UV CD difference spectra of the tetramer for example, could be calculated as a simple weighed sum of dimer and hexamer spectra. The authors had similar intentions to ours in the way that they wanted to use spectral parameters to predict insulin association states. However, clusters of points emerged only if a difference spectrum was calculated and plotted against another physicochemical constant, in that case near-UV CD difference spectrum versus  $K_{SV}$  (Stern-Volmer constant calculated from fluorescence quenching by acrylamide). This is of course not ideal and even with that method, dimers/tetramers formed a single cluster. Furthermore, to date no one has proposed yet a crystal or solution structure by NMR of the tetrameric state. There is therefore some ambiguity around that state. On the other hand, it is natural to think that the chances to

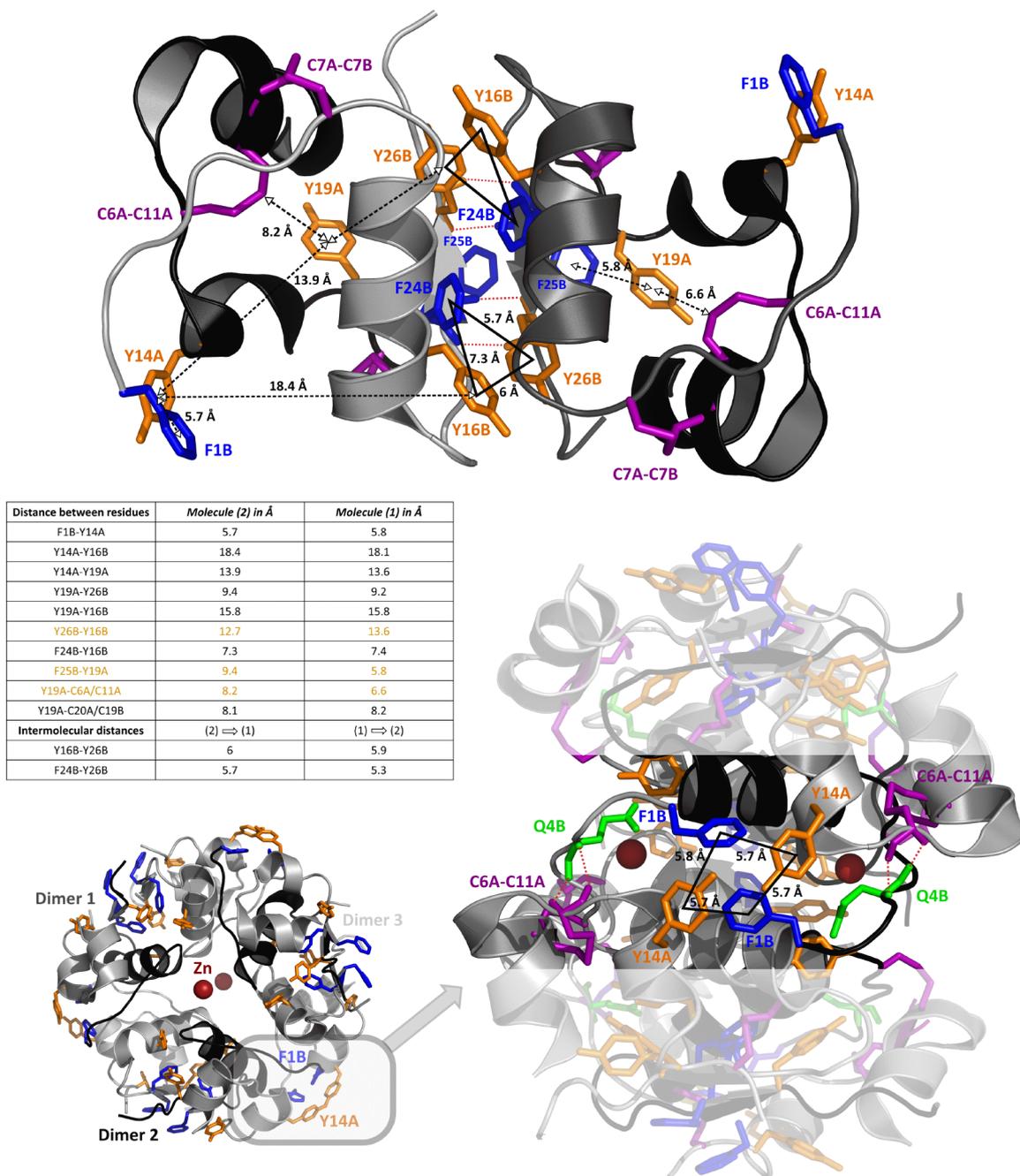
see three dimers interacting exactly in the same time to form one hexamer are very low. Therefore, two dimers first have to meet before forming a bigger structure, which make the tetrameric state more plausible.

Overall, multiple studies<sup>17,69-73</sup> have shown that each association state of insulin could be “isolated” using the conditions listed in (see section 2.4.1). The use of such buffers enabled the avoidance of the concentration dependency otherwise necessary to generate different associations, which would not suit fluorescence analysis (very large intensity variation, increasing inner filter effects, etc.).

The surface responsible for the dimerization of insulin is made mostly of hydrophobic or partially hydrophobic B-chain residues<sup>327,329,408</sup>: GlyB8, ValB12, TyrB16, GlyB23, PheB24, PheB25, TyrB26, ProB28. The goal here is not to describe the tens of interactions occurring at the dimer interface (more details can be found in work of Baker *et al.*<sup>327</sup> or Blundell *et al.*<sup>329</sup>), but to focus only on the interactions involving chromophores. These are probably part of the most important interactions anyway, considering that the dimer structure is mainly determined by the formation of an antiparallel  $\beta$ -sheet between residues B24-B26 of each monomer, all of which are fluorescent<sup>341</sup>. Four important intermolecular hydrogen bonds are formed and contribute to the dimer stability with support of multiple hydrophobic interactions between side chains of ValB12, TyrB16, PheB24, PheB25, TyrB26, and ProB28<sup>327,329</sup>. All these residues (except PheB25) that were partially exposed to the solvent, end up completely buried in the dimer interface (Figure 49) approximated as a square surface with sides  $\sim 23 \text{ \AA}$  long<sup>327</sup>. The packing arrangement between the aromatic residues is also particularly interesting. The residues TyrB16, PheB24, and TyrB26 form a complex network of loose contacts ( $\sim 4 \text{ \AA}$  long) with their antiparallel-facing homologues, generating shielded aromatic cages. Another aromatic contact to consider involves PheB25 in molecule 2 conformation and the adjacent TyrA19<sup>327</sup>.

Compared to the dimer interface, the residues that form the surface responsible for insulin hexamerization are both polar and non-polar, and from A and B chains: LeuA13, TyrA14, GluA17, PheB1, ValB2, GlnB4, GlnB13, AlaB14, LeuB17, ValB18, CysB19, and GlyB20<sup>327</sup>. The packing between dimers is much looser than packing between monomers<sup>327,340</sup>. As shown in Figure 49, the residues CysA11 (C11A) and GlnB4 (Q4B) of interacting dimers are hydrogen bonded at two positions. This connection between dimers is strengthened by a weak ionic interaction between PheB1  $\text{NH}_3^+$  and GluA17  $\text{COO}^-$  (not shown on figure)<sup>409</sup>. Both TyrA14

residues are also in close proximity, forming through hydrophobic interactions and with the adjacent LeuA13 a cavity under which both PheB1 sidechains can fit<sup>410</sup>. This cluster of four fluorophores (Figure 49) is another example of an aromatic cage, situated this time on the surface of the hexamer. This leaves both TyrA14 residues exposed to the solvent.



**Figure 49:** Representation in the dimeric (top) and hexameric (bottom right) state of some important interactions and distances between aromatic residues. The intermolecular H-bonds are represented by thin red-dashes and the aromatic cages by thick black lines. For clarity, not all distances in dimeric state were mentioned as they might vary due to asymmetry; they were instead all reported in the table attached (see yellow for changing ones). These distances were measured between phenyl ring centres, using PDB file 4INS.

To date, many methods have been used to study the association states of insulin, such as sedimentation equilibrium mentioned already<sup>349,402-403,405-406</sup>, circular dichroism<sup>360,405,407,411-414</sup>, concentration<sup>hh</sup> difference spectroscopy<sup>415-417</sup>, small angle X-ray scattering (SAXS)<sup>17,71-72,400,418-419</sup>, static and dynamic light scattering<sup>69,404,419-421</sup>, <sup>1</sup>H NMR<sup>422-425</sup>, mass spectrometry<sup>426-428</sup> with development of electrospray technique that enabled one to maintain non-covalent interactions in the gas phase, and other less “obvious” methods such as osmotic pressure<sup>429</sup> or pulsed-field gradient NMR enabling one to get indirectly the hydrodynamic radius of insulin through measurement of its self-diffusion coefficient<sup>430</sup>.

### 5.3 Measuring protein aggregation: from traditional methods to motivations for new alternatives

The self-association of insulin is only one example of aggregation and it is important to remember that protein aggregation, in general, remains one of the biggest challenges affecting the manufacture and safe use of biopharmaceuticals<sup>1</sup>. The presence of aggregates in therapeutic products can decrease bioavailability<sup>3</sup> and increase the risk of adverse immune response<sup>4</sup>. The problem is that at every step of the bioprocess chain *i.e.* production (cell culture, purification, formulation, etc.), storage, transport or delivery to the patient, protein aggregation can occur as a result of changes in a multitude of factors such as temperature, pH, protein and/or salt concentration, light, addition of co-solute, additive etc<sup>5,9,431-434</sup>. Even standard procedures such as mixing, filling, filtering, pumping can favour protein aggregation by contact with air/water or water/container interfaces<sup>9</sup>. Freezing the protein for storage or transport can also change its structure<sup>5,435</sup>.

Aggregates is a term that gathers every species that are multimers of the molecule of interest<sup>5,9</sup>. What makes the analysis of protein aggregates so difficult in analytical chemistry, is that these species can be very diverse in terms of their form (amorphous, spherical, fibrillar), the type of linkage (covalent versus non-covalent), the potential for reversibility, the folding state of the protein used to assemble them (native versus unfolded) and their size (nm up to mm)<sup>5,9,431-434,436-437</sup>. An informative review by den Engelsman *et al.*<sup>9</sup> details the strategies that are available to measure protein aggregates during biopharmaceutical product development. It is now well known that not a single analytical method can cover the entire size range of

---

<sup>hh</sup> Concentration difference spectroscopy uses UV-Vis absorbance spectra. The principle is to keep constant the total quantity of light-absorbing material, by varying simultaneously the concentration of solution and pathlength of light. The difference spectra will then be other than 0 only if association occurs, as spectral properties will change. This proved to be useful for insulin only for monomer-dimer equilibrium<sup>415</sup>.

aggregates. Large visible particles ( $> 100 \mu\text{m}$ ) can be detected by human eye in quality control (QC) labs, whereas methods such as optical microscopy, turbidimetry or light obscuration are well-established for detection of the largest subvisible particles (1-100  $\mu\text{m}$ )<sup>5,9</sup>. The latter is often seen as the gold standard in biopharma industry for particles in this size range. In US<sup>438</sup> and European pharmacopoeia<sup>439</sup> it is suggested as a reference method for analysis of particulate matter in injections. We are interested here in measuring small soluble aggregates (1-50 nm) which are, by being the smallest subvisible particles, more difficult to measure. A selection of the most common/applied method for small aggregates measurement with some advantages/drawbacks, are summarised in Table 17.

Among all these methods, size exclusion chromatography (SEC) is the gold standard in the biopharma industry<sup>9,437,440</sup>. SEC is able to analyse both covalent and non-covalent aggregates with good robustness and precision<sup>9</sup>. However, several concerns need also to be discussed. In particular, a significant loss of multimers can be observed during chromatography due *e.g.* to the physical filtration of bigger particles on column frits, but also and mainly due to the adsorption of aggregates onto column stationary phase<sup>441-444</sup>. This adsorption is non-specific, meaning that the protein of interest and/or aggregates can interact with column-bed. For proteins, this leads to abnormal peak shape/elution behaviour and low protein recovery, whereas for aggregates that may have a much higher affinity with the column-bed due to their denatured “sticky” state, this means no or under-detection which can have critical consequences in product safety<sup>441-444</sup>. Other sources of multimer loss (affecting more non-covalent species) result from the dissociation of aggregates in the presence of hydrodynamic shear forces, or dissociation due to unavoidable dilutions during sample preparation & chromatographic run, which affect the reversible equilibrium of small oligomers<sup>5,9,437,441,443,445</sup>.

Conversely, new aggregates can also be formed during the course of an SEC experiment as aggregates that are adsorbed onto the column, can further denature, and bind to other particles to form new aggregates<sup>437,441,445</sup>. Ironically, this was also by trying to counteract the electrostatic interactions responsible for column adsorption (by using *e.g.* high ionic strength mobile phase) that new aggregates were found in solution<sup>437,441-443</sup>. Therefore, the entire distribution of oligomers present in the original sample can change (formation & dissociation) during SEC, if the sample needs a mobile phase of different composition to be eluted. This is unfortunately the case in many SEC studies as additives such as organic co-solutes, denaturant salts, or different eluting conditions (pH, temperature, ionic strength) are very often necessary to increase sample recovery<sup>437,441-443</sup>. Weakly-bonded aggregates are of course the type of

aggregates that are the most sensitive to these changes<sup>9,443</sup>, which is of high concern for the analysis of many native oligomers including insulin.

The best example to illustrate these points is probably the study by Tantipolphan *et al.*<sup>18</sup>, where the authors specifically studied insulin elution behaviour during SEC analysis. The authors used high pressure, HP-SEC, which is the reference method given by both US<sup>446</sup> and EU<sup>447</sup> pharmacopoeia for the analysis of aggregates in insulin formulations. This is surprising considering that HP-SEC should in theory be less suited for the analysis of weakly-bonded aggregates than FPLC-SEC, where the pressure and thus the risk of dissociation is much lower. Another surprising point is that both pharmacopoeias recommended the use of an acidic mobile phase with 1 g/L L-arginine/acetonitrile/glacial acetic acid in 65:20:15 (v/v/v) proportions, which is different from most insulin formulations that are developed mostly at neutral pH<sup>448</sup>. Considering the high dependency of insulin association states to pH, it is highly probable that the SEC results will be different from the original formulation<sup>18</sup>. The authors were thus interested in measuring the HP-SEC behaviour of insulin at neutral pH. They proved that during SEC separation, insulin spontaneously re-equilibrates to adapt to the new mobile phase composition. SEC analysis is thus totally insensitive to the original association states present in the formulation, perfectly illustrating the issues encountered with SEC<sup>18</sup>. Furthermore, peak tailing was observed due to re-equilibration and interactions between insulin and the column. The addition of L-arginine, known to suppress column adsorption<sup>442,449</sup>, worked well in removing the tailing. But the authors proved that it also interfered with the association states of insulin which shows the great difficulty in obtaining the right mobile phase for reliable SEC analysis of soluble non-covalent oligomers.

SDS-PAGE and CE-SDS are the only two other methods to be QC-compatible<sup>9</sup> for the analysis of protein aggregates. The issue with these is that neither can detect non-covalent aggregates. Analytical ultracentrifugation by sedimentation-velocity (AUC-SV) would be an interesting alternative (Table 17) except that as highlighted by Philo<sup>437</sup>, the throughput is very low with measurements generally taking several hours. Added to the necessary dilutions and high cost for the equipment, this method does not appear as a viable alternative for *e.g.* monitoring in near real-time the protein aggregation state during biopharma process. Dynamic light scattering (DLS) is another good alternative that has increased in popularity as a screening tool in the biopharma industry<sup>450</sup> as it is non-destructive, easy to use and requires no sample manipulation other than filtration, therefore avoiding issues associated with dilution and mobile phase composition. However, even if its sensitivity is very high for large particles<sup>9</sup>, it was

reported to not be suitable for small soluble oligomers because particles need to differ by a factor  $\sim 8$  in molecular size to generate different peaks<sup>437</sup>. The resolution of species with similar molecular weight is thus very low. The data analysis is also quite complex and highly dependent on the algorithm used to convert the measured intensity distributions to fraction weights<sup>5</sup>. This mathematical conversion is based on various assumptions that are all potential sources of error, and are responsible for the low robustness of the method<sup>437,441</sup>.

**Table 17: Advantages and drawbacks of some analytical methods (modified from<sup>9</sup>) relevant to the analysis of small protein aggregates during biopharmaceutical development.**

Method	Advantages	Drawbacks
SEC	<ul style="list-style-type: none"> <li>• Robust</li> <li>• Low cost</li> <li>• Sensitive &amp; precise</li> <li>• Possibility of being coupled with various detectors</li> </ul>	<ul style="list-style-type: none"> <li>• Time consuming</li> <li>• Destructive method</li> <li>• Low resolution (small oligomers)</li> <li>• Sample dilution</li> <li>• Matrix/column interactions</li> </ul>
DLS	<ul style="list-style-type: none"> <li>• Non-destructive</li> <li>• Highly sensitive (large aggreg.)</li> <li>• Easy to perform</li> <li>• Wide concentration/size range</li> </ul>	<ul style="list-style-type: none"> <li>• Not very robust</li> <li>• Data analysis complicated</li> <li>• Low resolution (small oligomers)</li> <li>• Very sensitive to contamination (filtration)</li> </ul>
AFFF*	<ul style="list-style-type: none"> <li>• Wide size range</li> <li>• Possibility of being coupled with various detectors</li> </ul>	<ul style="list-style-type: none"> <li>• Low precision</li> <li>• Sample dilution</li> <li>• Membrane interactions</li> <li>• Need expert technical staff</li> </ul>
AUC-SV*	<ul style="list-style-type: none"> <li>• High resolution</li> <li>• Wide concentration/size range</li> <li>• Measure size and shape</li> <li>• Non-destructive</li> <li>• Sensitive &amp; precise</li> </ul>	<ul style="list-style-type: none"> <li>• Very time consuming</li> <li>• Complex data</li> <li>• Expensive instrumentation</li> <li>• Need expert technical staff</li> <li>• Sample dilution</li> <li>• Low throughput</li> </ul>
SDS-PAGE Or CE-SDS*	<ul style="list-style-type: none"> <li>• Easy to perform</li> <li>• Low cost</li> <li>• Fast separation (CE-SDS)</li> </ul>	<ul style="list-style-type: none"> <li>• Protein staining (SDS-PAGE)</li> <li>• Low precision</li> <li>• Sample dilution</li> <li>• Does not detect non-covalent aggregates</li> </ul>
SAXS	<ul style="list-style-type: none"> <li>• High resolution</li> <li>• No sample prep needed</li> </ul>	<ul style="list-style-type: none"> <li>• Expensive</li> <li>• Need expert technical staff</li> <li>• Time consuming</li> </ul>

\*AFFF stands for asymmetrical field-flow fractionation, AUC-SV for analytical ultracentrifugation by sedimentation-velocity, CE-SDS for capillary electrophoresis-SDS.

There is therefore an opportunity for a new analytical method. As pointed out by Slew<sup>450</sup> and suggested by the Food and Drug Administration (FDA)<sup>451</sup> in a report addressed to therapeutic protein manufacturers, industries that were asked to track the presence of  $>1$  micron subvisible aggregates in therapeutic product, are now encouraged with the development of new analytical methods to test for smaller particle sizes. In particular the FDA as well as several authors<sup>9,445,437,441</sup> promote more and more the use of orthogonal methods to complement SEC.

Any new methods should ideally be fast, non-destructive, have high throughput and require no sample manipulation. These are all characteristics inherent to intrinsic protein fluorescence, as well as being inexpensive and highly sensitive.

Fluorescence spectroscopy (or microscopy) is in fact often used as a method to detect protein aggregation or fibrillation<sup>5,9</sup>. However it almost always refers to the use of extrinsic dyes, such as Congo red, Nile red, Thioflavin T (ThT), or 1-anilinonaphthalene-8-sulphonic acid (ANS)<sup>27,452-453</sup>. One issue associated with the use of probes is that the method is intrusive as it requires sample manipulation and staining protocols that can be time consuming. Also, probes can interfere in the misfolding process and aggregation behaviour of proteins<sup>454-456</sup>, modifying the kinetics and pathways of fibrillation. Intrinsic fluorescence is thus in theory much more attractive but suffer from a complex emission space that cannot be apprehended with a standard 2-D measurement. The only way to decipher all that information and achieve the full potential of intrinsic fluorescence, is to solve the complex spectral overlap by associating multidimensional measurements with factor based chemometric methods.

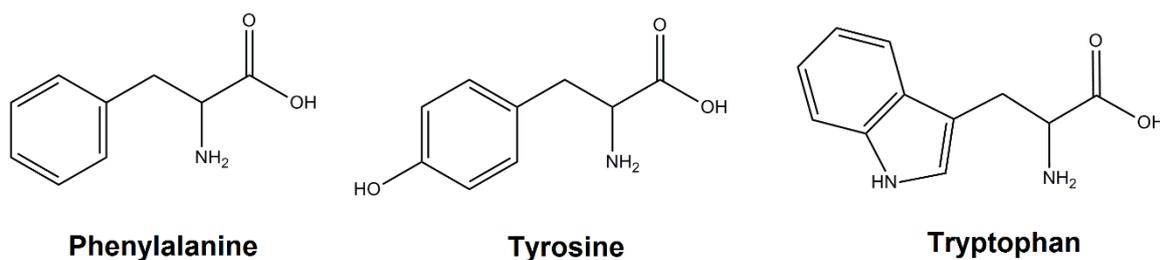
## 5.4 The steady-state fluorescence of Tyrosine and Insulin

### 5.4.1 Tyrosine fluorescence

Tryptophan (Trp) tyrosine (Tyr) and phenylalanine (Phe) are responsible for the intrinsic fluorescence of proteins<sup>52</sup>. Since the early 1960s, proteins were classified in two classes based on the presence (“class B”) or the absence (“class A”) of Trp residue in their sequence<sup>53</sup>. The fluorescence of class B proteins is usually dominated by Trp emission, not because Trp has a higher quantum yield than *e.g.* Tyr (Table 18) or is present in greater quantities, but because both Tyr and Phe efficiently transfer their energy to Trp<sup>28,52-53,457</sup>, the latter absorbing at longer wavelengths than them (and with higher efficiency).

**Table 18: Fluorescence characteristics of aromatic amino acids responsible for protein fluorescence, determined at neutral pH and room temperature in 0.1 M phosphate buffer (data from PhotoChemCad 2.1 software<sup>31-32</sup>).**

Fluorophore	Absorptivity ( $M^{-1} cm^{-1}$ )	Quantum yield	$\lambda_{ex} / \lambda_{em}$ (nm)
Tryptophan	5579	0.12	280 / 354
Tyrosine	1405	0.13	274 / 303
Phenylalanine	195	0.022	257 / 280

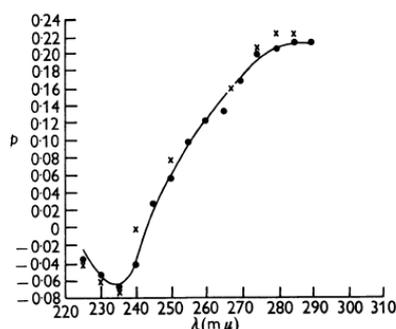


**Figure 50: Chemical structures of the aromatic amino acids responsible for the intrinsic fluorescence of proteins, drawn with ChemDraw.**

Insulin only has Tyr and Phe as fluorophores which classifies it as “class A” protein. Due to the low quantum yield/ absorptivity of Phe residues and the fact that Tyr are acceptors of Phe emission<sup>28,458</sup>, the fluorescence of class A proteins is generally dominated by Tyr emission.

Whereas Trp has complex emission properties due to the presence of nearly isoenergetic excited states  $^1L_a$  and  $^1L_b$ <sup>ii</sup> (see section 7.15 in Appendices for more details), Tyr owes its fluorescence to the presence of a phenol group (Figure 50) where both excited states have a too large energy gap to overlap<sup>460</sup>.  $^1L_a$  absorbs at wavelengths below 250 nm (maximum  $\sim 223$  nm) and  $^1L_b$  absorbs at longer wavelengths 260-290 nm (maximum  $\sim 274$  nm), with emission that occurs in both cases at 303 nm<sup>28,461-462</sup>.

As opposed to Trp which is characterised by a complex anisotropy in the 250-300 nm band, one consequence of this non-overlap is that Tyr anisotropy is simpler. Because the absorption dipoles associated with these transitions are orthogonal<sup>461,463</sup> (Figure 52), negative values are obtained for the high energy band ( $^1L_a$  with orthogonal absorption/emission dipoles) that progressively increase to positive values when reaching the low energy band ( $^1L_b$  with collinear absorption/emission dipoles) up to 290 nm<sup>53,457</sup> (Figure 51).

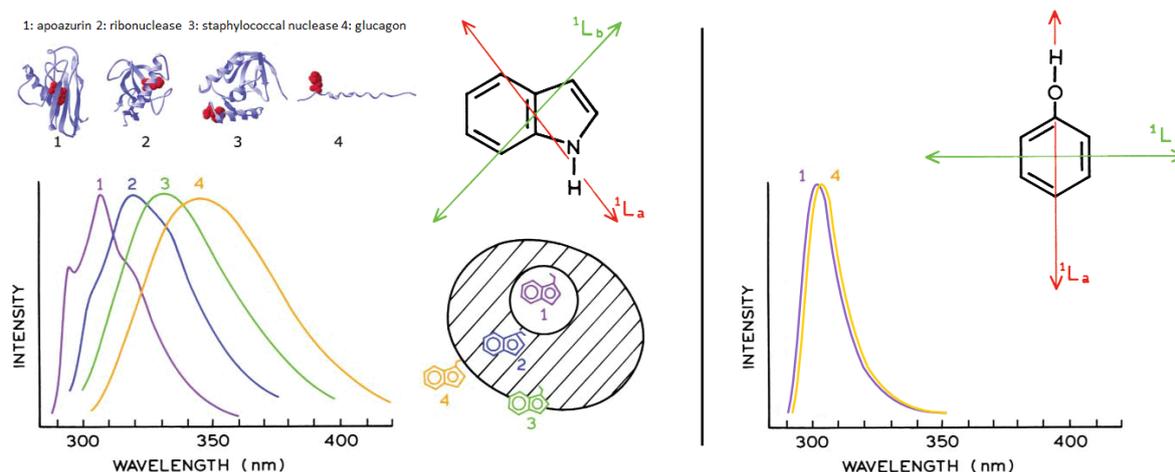


**Figure 51: Polarization spectra of cresol (×) and tyrosine (●), 0.5 mM in propylene glycol, at  $-70^\circ\text{C}$ . Reproduced from<sup>457</sup>, with permission from Portland Press.**

<sup>ii</sup> Platt's notation<sup>459</sup> are often used to describe the photophysics of aromatic amino acids.  $^1L_b$  and  $^1L_a$  are the names given to the two lowest excited singlet states, respectively  $S_1$  and  $S_2$ .

A second consequence is that as opposed to Trp for which the emission (like in proteins) can arise from  $^1L_a$  and shift up to  $\sim 50$  nm depending on the polarity of its microenvironment, Tyr solely emits from  $^1L_b$  and shows a fluorescence that is largely insensitive to the solvent polarity<sup>28,461</sup> (Figure 52). What makes the  $^1L_a$  transition in Trp particularly sensitive to solvent polarity is the fact that it involves directly the indole nitrogen that can undergo hydrogen bonding with its microenvironment. The  $^1L_b$  in Tyr on the other hand does not involve any potential atoms that can increase the sensitivity of that dipole to hydrogen bonding (which is in fact comparable to the  $^1L_b$  in Trp).

The insensitivity of Tyr fluorescence to a change in its microenvironment was for example observed during the fibrillation of bovine insulin where the maximum emission was obtained at 305 nm during a 22 hour experiment<sup>454</sup> (see Figure 53). But it was also proven using *N*-acetyl-L-Tyrosinamide (NAYA) in various solvents and in the presence of various ligands<sup>33,464</sup>, which is often used to model the Tyr residue in proteins. In fact, it seems that the absorbance or excitation spectra of Tyr are more prone to show some slight shifts with change in the environment than does the emission. This particular point, as well as the meaning of very small spectral shifts that can occur in the emission, are further developed in sections 5.7 and 5.10.1. Nevertheless, this lack of environmentally induced spectral shifts in the Tyr emission reduces drastically the attractiveness of using Tyr intrinsic fluorescence for diagnostic purposes compared to Trp.



**Figure 52: Illustration of the difference between class B (left) and class A (right) protein in terms of Stokes shift upon exposure of the aromatic residue to its environment. A schematic representation of Indole/Phenol excited states, responsible for the fluorescence of Trp/Tyr respectively, is also included. Modified and taken from<sup>28</sup>, with permission of Springer Nature.**

One mechanism which can shift Tyr fluorescence to the red is the ionization of the phenolic group in the ground or excited state. At high pH (~11), the phenolic hydroxyl group which has a  $pK_a \sim 10.3$ <sup>28,465-466</sup> in ground state can ionize, forming tyrosinate<sup>461-462,465-466</sup> and resulting in strong static quenching of Tyr fluorescence. Tyrosinate has an absorbance red shifted to 295 nm and a fluorescence emission maximum around 345 nm, explaining how it can easily be mistaken for Trp emission. The issue with tyrosinate (which is also relevant to our study) is its potential fluorescence even at neutral pH. This is made possible due to the fact that tyrosine is much more acidic in the excited state, with a  $pK_a$  between 4 and 5<sup>28,461,465-466</sup>. Ionization could occur during the lifetime of the excited state (dynamic quenching), if a suitable base is present. Excited state proton transfer has for example been observed at neutral pH for Tyr mixed with increasing concentration of acetate<sup>465</sup>, but also for several proteins such as Bovine testes Calmodulin<sup>467</sup>, peptides isolated from an Indian cobra<sup>468</sup> etc. The presence of tyrosinate emission in proteins at neutral pH being dependant on both base concentration and the degree of exposure of Tyr to the solvent<sup>28</sup>.

The fluorescent properties of Tyr can also change upon radicalization. This process that is either photoactivated or enzyme-catalysed (*e.g.* peroxidase), forms two tyrosyl radicals that cross-link to form a new molecule called dityrosine<sup>469-470</sup>. Besides being potentially present in proteins either naturally or following oxidative stress<sup>471-472</sup>, dityrosine is interesting as a fluorescent probe because it absorbs and emits at longer wavelengths ~ 315/420 nm<sup>454,469-470</sup> respectively. Dityrosine is thus an efficient protein marker, which has been very useful to monitor the presence of glycoprotein product in a bioprocess<sup>133</sup>, or to follow oxidative stress in proteins<sup>471-472</sup>. The other interesting aspect about dityrosine is that it can form covalent intra or intermolecular bridges. This was for example observed in the case of insulin under enzymatic<sup>469-470</sup> or UV-exposure, where a covalent cross-linked dimer was formed<sup>473</sup>. The formation of dityrosine occurs with simultaneous quenching of Tyr emission.

#### **5.4.2 Fluorescence methods used to measure insulin aggregation**

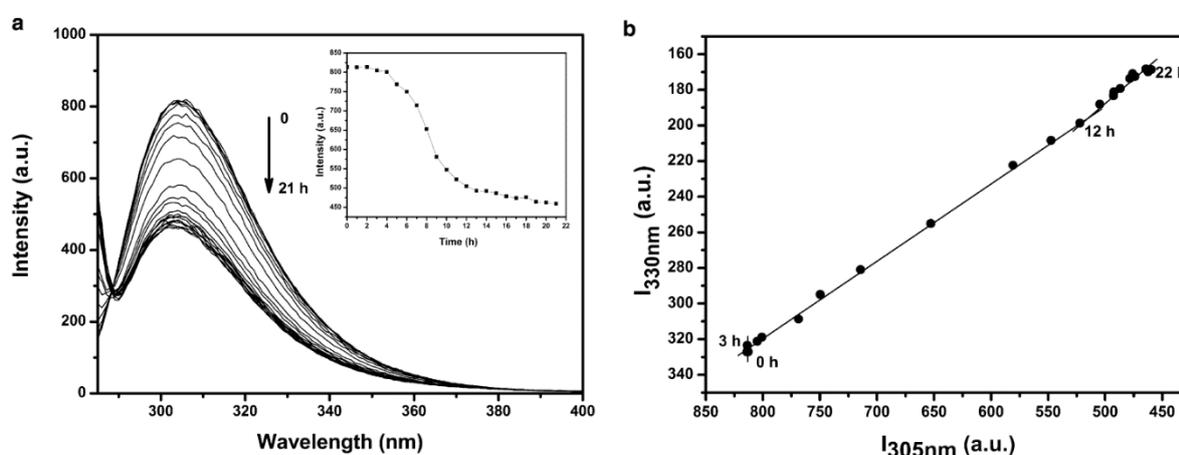
The traditional approach of using fluorescence measurements for insulin analysis involved mainly the use of extrinsic fluorophores. ThT for example displays an enhanced fluorescence upon binding to  $\beta$ -rich structures and is the probe of choice to detect the presence of insulin amyloid fibrils. ThT shows an increase in the emission band at 482 nm upon excitation at 450 nm which was used in many studies to monitor the kinetics of insulin fibrillation<sup>71-73,399-400,454,474</sup>. Recent studies have however shown that ThT also binds to hydrophobic cavities of non-fibrillar structures<sup>475</sup>, meaning that the specificity of that dye to

amyloid fibrils is questionable<sup>476</sup>. ANS on the other hand is able to bind to solvent exposed hydrophobic regions of proteins. The fluorescence of ANS observed at 460 nm upon excitation at 350 nm<sup>454</sup> dramatically increases upon binding with hydrophobic patches. ANS has been used to detect the partially folded state that precedes insulin fibrillation<sup>72-73,454</sup>. Drawbacks associated with use of extrinsic dyes were explained previously, but specific examples related to insulin can be found. In particular, recent studies suggests that ThT, by binding to cavities located between protofilaments, generates unavoidable structural changes in the fibrils<sup>477</sup>, thus changing the fibrillation process. Also ThT at certain concentrations and temperatures has been shown to cause some retardation in fibrillation kinetics<sup>478-479</sup>, proof of its interference with the aggregation mechanism.

An alternative approach to studying insulin aggregation by fluorescence is to use energy-transfer based methods rather than simple intensity increases (*e.g.* ANS/ThT). Hassiepen *et al.*<sup>412,480</sup> analysed insulin dimerization by using equimolar solutions of acceptor-labelled and donor-labelled monomers. These solutions were mixed in varying volume ratios respecting a fixed total concentration, and the FRET efficiency was measured as a function of the donor molar fraction ( $x_d$ ). This experimental design was interesting as it enabled the authors to avoid the huge concentration variation normally required to shift the association equilibria. By analysing the fluorescence data, they observed that the FRET efficiency was maximal for  $x_d = 0.5$  and using this data they calculated the dimerization constant  $K_{1,2} = 0.72 \times 10^5 \text{ M}^{-1}$ , which matched other literature values. The donor/acceptor pair used were 2-aminobenzoyl and 3-nitrotyrosyl, which were assumed to be small enough not to alter the chemical nature of the protein. One critical aspect of this study was to find a donor/acceptor pair with a Förster distance (here 29 Å) that matched their 32 Å separation once the dimer was formed. They also had to find a suitable site to label insulin (LysB29) and hope that it would not affect the dimer association process. The preparation of labelled insulin was not simple either, requiring multiple steps (*e.g.* dissolution, stirring, filtration, redissolution, purification by reversed phase chromatography...) that are all susceptible to change the original distribution of oligomers. This study was thus interesting but it also had several drawbacks, and the formation of hexamers was avoided.

Concerning the use of steady-state intrinsic fluorescence to study insulin aggregation, two main techniques can be found in the literature. First was the use of phase diagrams which is a method applied to fluorescence but that can be applied to any type of spectroscopy (*e.g.* on near-UV CD data for insulin<sup>17,400,474</sup>). The idea behind this method is to look at a process (*e.g.*

aggregation or denaturation) that affects the protein conformation and look at the relationship between the fluorescence intensity at two wavelengths, namely  $I_{\lambda_1} = f(I_{\lambda_2})$ . Any non-linearity in this relation reflects the sequential aspect of a protein conformational change, which enables to detect potential intermediate states. This was for example used on insulin (Figure 53, b) to detect structural changes that occur upon aggregation<sup>454</sup>. In this study, the insulin sample was continuously heated at 60°C during 22h and the aggregation process was monitored by fluorescence spectroscopy. By plotting the emission intensity at 330 nm ( $I_{330\text{nm}}$ ) against  $I_{305\text{nm}}$ , the authors observed three linear segments (0-3h, 3h-12h, 12h-22h) that suggested the existence of distinctive structural transitions during insulin fibrillation.



**Figure 53: Insulin denaturation/fibrillation detected by Tyr fluorescence (a) Fluorescence emission spectra of tyrosine during insulin aggregation. The insulin sample (0.2 mg/mL) was excited at 276 nm, with emission scans from 280 to 500 nm. The figure shows a consistent decrease in intensity of the emission  $\lambda_{\text{max}}$  (305 nm) after ~2 h of incubation, reaching equilibrium after 12 h. (Inset) A sigmoid is observed on plotting the emission intensity at 305 nm against time. (b) Fluorescence phase diagram obtained by plotting  $I_{330\text{nm}}$  against  $I_{305\text{nm}}$  shows the multiple structural transitions accompanying insulin aggregation. The experiment was done in 0.1% HCl (pH 1.9) at 60°C, with continuous stirring at 120 rpm. Reproduced from<sup>454</sup>, with permissions from Elsevier.**

The information that can be extracted from such diagrams is thus quite limited. One might also expect this method to not be reproducible, as the intensity values will surely vary from sample to sample and aggregation experiments. In fact, as shown in Figure 53, the authors did not include error bars or replicate measurements to their analysis.

Acrylamide quenching is the other main method based on Tyr intrinsic fluorescence, that was used for insulin analysis. The principle of this method was explained in section 1.3.1 and consists in measuring the decrease of insulin fluorescence upon increasing concentration of acrylamide. The data is then fitted with the Stern-Volmer equation in order to retrieve the constant  $K_{\text{sv}}$ , that gives information about the accessibility of Tyr to its environment<sup>46</sup>. In the

case of insulin, this method proved useful to look at structural changes taking place upon denaturation of the monomer<sup>400,474</sup>, but was also used to look at the various soluble oligomers<sup>17,72</sup>. However, the only conclusion that could be drawn from these measurements was that Tyr residues were more accessible in the monomer than in the hexamer because of a higher  $K_{sv}$ . Similarly to phase diagrams, the quantity of information that can be extracted from acrylamide quenching is thus pretty limited and one might question the reproducibility of such measurements.

In Figure 53, we mentioned that Bekard *et al.*<sup>454</sup> in their 2009 study used phase diagrams to look at bovine insulin fibrillation, but they also used the intrinsic Tyr signal to monitor its kinetics (Figure 53, a). After 22h of heating at 60°C, the authors noted the absence of any spectral shift (emission centered at 305 nm) as well as a strong quenching that happened 2h after the beginning of the experiment. By plotting the fluorescence intensity versus time, they observed a typical sigmoidal curve (Figure 53, a) indicative of a three steps fibrillation process, that matched with the phase diagrams results. No other information was however extracted from these spectra, which once again show how limited was the use of Tyr intrinsic signal. The authors explained the decrease in intensity observed during fibrillation by the presence of intra/intermolecular FRET between adjacent Tyr residues, and the fact that structural reorganizations probably bring some of the excited Tyr residues in effective contact with cystinyl side chains (strong quenchers of protein intrinsic fluorescence).

The presence of extensive inter-Tyr FRET is indeed one of the only/main features that has been proven to exist in native insulin. Teale anticipated the presence of efficient FRET in insulin (and other class A proteins) by calculating the average inter-Tyr distance (16.57 Å), which was within the range at which efficient energy transfer was observed in other known systems<sup>52</sup>. This was also supported by the calculations of Karreman *et al.*<sup>481</sup>, who obtained for the various donor/acceptor Phe/Tyr/Trp pairs, a critical distance,  $R_0$  well below the average radii of globular proteins (15 to 30 Å).

Table 19<sup>28</sup> summarizes from multiple studies of different proteins, the range of  $R_0$  values that can be obtained for various aromatic amino acid FRET pairs. It is important to speak of range as the microenvironment that affects the fluorescence properties of each residue involved in the FRET pair and used to calculate  $R_0$  (see section 1.3.2) is unique to each residue and vary from case to case. Eisinger *et al.*<sup>482</sup>, beside proposing  $R_0$  ranges that agreed with Table 19 values, highlighted for example the huge dependency of  $R_0$  on the quantum yield of the donor

( $\Phi_D$ ), a value often not known in proteins but dependant on the donor microenvironment<sup>28,482-483</sup>.

**Table 19: Summary of the range of critical distances ( $R_0$ ) encountered for the various aromatic amino acid FRET pairs in proteins. Phe-to-Tyr and Tyr-to-Tyr (in italic) are the two FRET pairs involved in insulin. Table taken and modified from Lakowicz<sup>28</sup>.**

Donor	Acceptor	$R_0$ (Å)
<i>Phe</i>	<i>Tyr</i>	11.5-13.5
<i>Tyr</i>	<i>Tyr</i>	9-16
Tyr	Trp	9-18
Trp	Trp	4-16

If we consider the inter-Tyr and Phe-Tyr distances in insulin (Figure 49) versus the  $R_0$  generally encountered for these FRET pairs in proteins<sup>28</sup> (Table 19), it is clear that both Phe-to-Tyr and Tyr-to-Tyr energy transfer can occur in insulin native state emission.

In fact, Weber in 1954 proved the existence of internal FRET in native insulin by fluorescence polarization measurements, under cryogenic conditions (90-95% propylene glycol, -70°C) where Brownian motions are absent. He obtained a lower anisotropy for insulin than for free Tyr<sup>53</sup> over the entire emission wavelength range (*e.g.*  $r \approx 0.08$  vs 0.16 for  $\lambda_{ex} = 285$  nm). Such differences could only be explained by the presence of extensive internal FRET between residues, as he previously proved that it is a phenomenon that strongly depolarizes the emission<sup>484</sup>. This also explained why the general quantum yield of class A proteins (0.02 to 0.08) is much lower than that of free Tyr<sup>52-53</sup>. The probability of FRET in the excited state is directly proportional to the excited state lifetime<sup>53</sup>, and thus there is a greater probability that Tyr with a high quantum yield (*i.e.* longer lifetime) transfer their energy to Tyr with lower quantum yield than for the opposite direction. FRET can thus lower the overall quantum yield of a protein, but many other quenching processes also do so. Examples of this include quenching by disulfide bridges or cysteine sulfhydryl groups for which the mechanism has not been understood yet<sup>36,461</sup>, electron charge transfer to local amide carbonyl groups from peptide backbone<sup>33,485</sup>, protolysis of the phenol group due to interaction with anions from the solvent (*e.g.* phosphate<sup>486</sup>) and/or nearby amino acid side chains (*e.g.* carboxylate groups of ionized aspartic or glutamic acid) which act as proton acceptors or partners in hydrogen bond formation<sup>461,487</sup> etc...

The presence of extensive internal FRET among Tyr residues in insulin seems therefore to be the key towards the resolution of insulin intrinsic fluorescence. However though Weber

proved its existence in the cryogenic state<sup>53</sup>, we could not find a study that was able to resolve this FRET and extract the individual Tyr emitters that contribute to the intrinsic insulin emission. Ultimately, this is what one would like to achieve in order to better monitor the changes associated with aggregation/fibrillation or denaturation processes. All the literature that we surveyed on the use of intrinsic fluorescence to study insulin aggregation seemed to suffer from a lack of information (*e.g.* phase diagrams or acrylamide quenching). In fact, intrinsic fluorescence generates a lot of information in the complex emission space. This emission cannot be fully understood with a standard 2-D measurement, as seen with the work of Bekard<sup>454</sup> (Figure 53). Instead the only way to achieve the full potential of intrinsic fluorescence and decipher that information, is to solve the complex spectral overlap by associating multidimensional measurements like EEM with factor based chemometric methods like MCR/PARAFAC. Sometimes, however, the fluorophores have too similar emission properties to be differentiated by standard EEM and more/better information is required. ARMES<sup>14</sup> by combining anisotropy ( $r$ ), related to the mobility/flexibility of constituent fluorophores, with MDF data provides this additional layer of information that may enable the differentiation of fluorophores with very similar emission properties.

Very few papers describe the collection of insulin EEM spectra<sup>jj</sup>. Therefore, all the data and modelling presented below are novel. Protein intrinsic fluorescence is often seen as a real puzzle because of the high variability of aromatic residue quantum yields/lifetimes, tightly related to the multitude of quenching processes that can happen in such an enclosed space. Associated to the strong spectral overlap between fluorophores, it becomes a very complex spectroscopic challenge<sup>28</sup>. We hope this study is a step towards the resolution of that puzzle.

## 5.5 Materials and data analysis

The reader is invited to refer to sections 2.4 for details about the sample preparation and the data collection.

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). Compared to the previous chapter, the method was changed slightly. The chemometric analysis

---

<sup>jj</sup> One paper from the lab focused on the separation of chemical defined medias and media materials that included recombinant human insulin at 2 mg/mL<sup>488</sup>. The separation was achieved by removing all EEM outliers with robust PCA (ROBPCA) and by building classification models (NPLS-DA) with the remaining samples. Another study collected insulin EEM spectra but used them simply to read the fluorescence intensity, without any multivariate treatment<sup>489</sup>.

was previously undertaken independently on all four polarization measurements, which had the advantage to confirm the robustness of modelling. However in fluorescence anisotropy theory<sup>78,86,160</sup>, the HH and HV polarizations are only used to calculate the G factor, a correction factor that cancel the wavelength dependent sensitivity difference shown by the detection channel to vertically and horizontally polarized light. Alternatively, the theory shows that the anisotropy  $r$  can be calculated from only two directions  $I_{\parallel}$  and  $I_{\perp}$  defined as  $I_{\parallel} = I_{VV}$  and  $I_{\perp} = G \times I_{VH}$ , so that  $r = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2 \times I_{\perp})$ . We therefore decided to simplify the method by adopting these notations and analysing only  $I_{\parallel}$  and  $I_{\perp}$ , as ultimately only these two measurements contain a different information<sup>kk</sup>.

$I_{\parallel}$  and  $I_{\perp}$  can also be used to calculate the total intensity  $I_T$ . The radiation emitted by a fluorophore can be defined in theory by the contribution of three orthogonal components  $I_x$ ,  $I_y$  and  $I_z$ , oriented along the space axis  $x$   $y$  and  $z$  such as  $I_T = I_x + I_y + I_z$ <sup>78,86,99,490</sup>. In the case where the polarized excitation is oriented along the  $z$  axis so that it is parallel to the fluorophore dipole ( $I_z = I_{\parallel}$ ), the radial symmetry of the emitted light (Curie's symmetry principle) imposes that  $I_y = I_x$ . Because  $I_y$  is also equal to  $I_{\perp}$ , the total intensity ( $I_T$ ) under a parallel polarized excitation can be written as  $I_T = I_{\parallel} + 2 \times I_{\perp}$ . Unpolarized light is then characterized by  $I_{\parallel} = I_{\perp}$  whereas  $I_{\parallel}$  and  $I_{\perp}$  have different contributions to  $I_T$  in the case of partially polarized light<sup>ll</sup>. The total intensity ( $I_T$ ) is thus polarization independent and is useful to calculate the fractional contributions of each fluorophore to the emission<sup>78,86,99</sup>.

$EEM_{\perp}$  is created by multiplying  $EEM_{VH}$  with the 3-D G factor. However, when the G factor is calculated from the measurements (Figure A- 29, a) the result is very noisy and not suitable for the calculation of  $EEM_{\perp}$ . In order to recover accurate  $EEM_{\perp}$  spectra and subsequent *aniso*-MDF plots, the 3D G-factor requires correction. Two approaches can be implemented: 1) pre-calculation by denoising the  $EEM_{HH}$  and  $EEM_{HV}$  spectra; 2) post-calculation by smoothing the G factor.

We tried first the post-calculation option, and denoised the G factor through PARAFAC using a one component model with non-negativity in all modes<sup>mmm</sup>. The improvement in

<sup>kk</sup> For information, in the previous chapter (Table 15) the % of variance explained by both xanthene dyes components was identical for ~ HH HV VH but different for VV polarization. HH HV and VH are carrying indeed the same information, except for the measurement intensity that varies according to the polarizers transmittance.

<sup>ll</sup> In fact, the ratio between  $I_{\parallel} - I_{\perp}$  and  $I_T$  is the definition of anisotropy (see Equation 22).

<sup>mmm</sup> Classical smoothing methods like Savitzky-Golay are not enough sufficient to denoise such data. The best way to avoid these noise issues would be to average the spectra from multiple acquisitions, or to increase the integration time. However it was not feasible here due to the long measurement times.

recovery of the G factor was certain as shown in appendix (Figure A- 29, b vs a). The obtention of this result required however some precautions. It appeared indeed that the model was highly influenced by the low G factor values present along the Rayleigh band<sup>nn</sup> (Figure A- 29, c), values that (if kept) made the modelled G factor decreasing along  $\lambda_{ex}$  (Figure A- 30, pink). To minimize this effect and the alteration of *aniso*-MDF spectra, the low G values had to be replaced. This improved the result (Figure A- 30, black) but it was not perfect. Instead we found that the best way to denoise the G factor was to use the pre-calculation option and denoise directly the Rayleigh corrected EEM<sub>HH</sub> and EEM<sub>HV</sub> datasets with a one component PARAFAC model constrained with non-negativity (Figure A- 31, pink). Like this, we removed the decreasing excitation slope that was previously observed with the post-calculation method (Figure A- 31, black). The difference anisotropy spectra with the raw/noisy data (Figure A- 31, right) also confirms the better results obtained with the pre-calculation method. Scatter region aside, this plot is indeed flatter and closer to 0 when G is calculated from the pre-calculation rather than the post-calculation method. Figure A- 32 also confirms this point and shows on top the very high stability of the G factor along the entire insulin data collection.

The other change implemented in this chapter is in the chemometric method used to resolve MDF data. MCR<sup>141,143</sup> and PARAFAC<sup>146,150,289,309</sup> are the most commonly used methods to resolve MDF measurements. In continuity with the work developed by Groza *et al.*<sup>14</sup>, it was natural in this thesis to use MCR. We saw in the previous chapter that MCR was able to recover accurately the spectral profiles and anisotropy of the individual components from non-interacting fluorophore mixtures. The use of a trilinear constraint in MCR models improved however their recovery, especially when fluorophores were strongly overlapping (*e.g.* xanthene dyes). Considering that fluorophores overlap even more in protein fluorescence and that PARAFAC is in essence a trilinear method capable of handling directly EEM data, it became logical to use PARAFAC for the protein data analysis. The absence of rotational ambiguities in the PARAFAC solution versus MCR is also another major advantage that comes with this method.

The general scheme describing the main pre-treatments / data analysis steps for applying ARMES to protein intrinsic fluorescence is given in Figure 54.

---

<sup>nn</sup> In the post-calculation method, we found that it was preferable to work with a G factor that was obtained from data where the Rayleigh band was not corrected. The correction of the Rayleigh band leaves otherwise residuals in the data that generates very erratic G factor values along the Rayleigh band, that are even harder to correct.

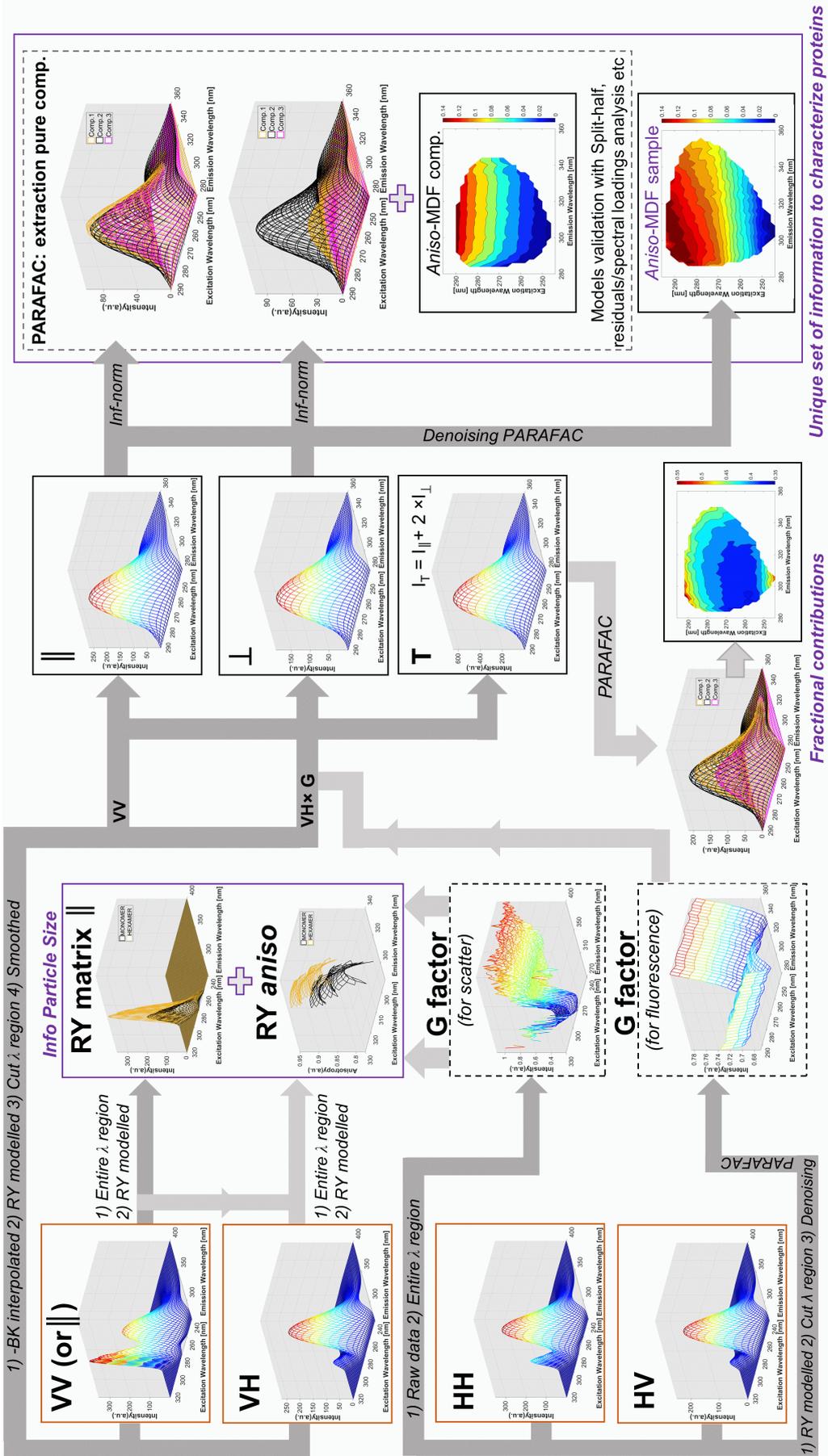
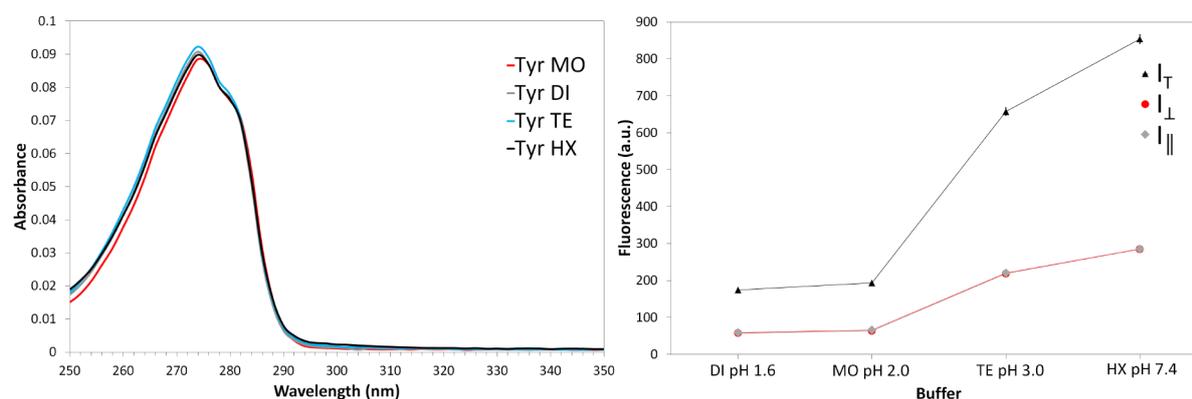


Figure 54: General scheme describing the main steps for applying ARMES to protein intrinsic fluorescence, starting from data collection (orange boxes) to main outcomes of the method (purple boxes). List of abbreviations employed: *RY* 1<sup>st</sup> order Rayleigh scatter, *BK* blank, *RY* modelled RY removed by PARAFAC, - *BK* interpolated: blank's RY interpolated prior subtraction.

## 5.6 Behaviour of Tyr in various buffers

As a control experiment we needed to verify the emission behaviour of Tyr in the different buffers used for the insulin studies, which are detailed in Table 4 (section 2.4.1). Triplicate samples of L-Tyr were thus made in each of these buffers at a concentration of  $\sim 1.8 \times 10^{-4}$  M and measured in the same conditions as for insulin samples (details in section 2.4.2). At this concentration, the absorbance is  $< 0.1$ , thus avoiding the need for IFE correction. The UV-Vis spectra as well as the maximum intensity obtained for the ARMES measurements are shown in Figure 55.



**Figure 55 : (left) UV-Vis spectra and (right) polarized fluorescence values obtained for L-Tyr in the different insulin buffers used. All data are averaged from three replicate measurements.**

Among all the buffers used, the monomeric buffer is the only one to contain a large proportion of organic compound (20% acetic acid). Acetic acid is however not a quencher of Tyr emission<sup>465</sup>. Instead, Tyr's quantum yield is known to be pH dependant<sup>462,491</sup>, which is what we also observed in Figure 55. It was shown in fact by Feitelson<sup>491</sup> that the quantum yield of Tyr was stable only between a pH of 4 and 8. In alkaline conditions, it drops due to the ground state ionization of the Tyr phenolic groups<sup>492</sup>. Below pH 4, the quantum yield drops as well but due this time to the protonation of the Tyr carboxyl group, which has a  $pK_a = 2.3$ <sup>491</sup>.

There is thus an influence of the buffer on Tyr emission intensity. Unfortunately, this data cannot be used to correct the pH dependency of the four Tyr residues present in insulin, as each residue in the protein has its own unique micro-environment (*i.e.*  $pK_a$ , refractive index, degree of solvent exposure, quantum yield) that greatly differs from the situation of Tyr in solution. The increase in intensity observed in Figure 55 is however interesting when compared to the decrease seen for insulin (Figure 63) going from dimeric to monomeric and eventually to hexameric forms in buffer. This opposite behaviour is a clear indication that the Tyr residues

must become less accessible to the solvent with increasing pH, which makes sense with the progressive burial of Tyr residues in the dimer or hexamer interfaces<sup>74,327</sup>.

Due to the potential influence of pH on insulin's emission intensity, we decided thus to build most of the insulin chemometric models (section 5.10) on data that were normalised to the maximum using the infinity-norm<sup>00</sup>. The rationale behind the normalization is to focus the separation solely on variations in the spectral shape, which is in practice more robust than intensity changes that are more sensitive to measurement and sample errors. This also enables one to build models for very high concentration insulin solutions, where IFE correction is not an option, but that are just as interesting for *i.e.* industry applications where a change in shape is enough to separate the aggregation states of insulin (5.11). The normalization also tends to stabilize chemometric decomposition and make it converge faster in the case of a given absolute (and not relative) threshold value.

Before moving on to the analysis of insulin data, it was also crucial at this point to investigate if the spectra obtained from measuring free L-Tyr in various buffers showed any spectral shifts, as that could have a strong influence on PARAFAC modelling. It was already clear from Figure 55 compared to Figure 56 that L-Tyr does not show the same red shift in the UV-Vis data going from monomeric to hexameric buffers. In order to confirm the lack of buffer influence on L-Tyr spectral shape, PARAFAC models were built using the exact same conditions as the models built for insulin. These conditions (constraints, initialization method, convergence criterion, model repetitions), as well as the results from this short study are shown in section 7.18.

The most important result to extract from this study was that the buffer had no influence on Tyr spectral shape, but only on its intensity. The models, even if reproducible, showed components that had highly collinear scores whether the data was normalised or not. Whereas a collinearity in the scores might sometimes be related to a real chemical property of the system or sometimes due to improper sampling<sup>309,493</sup>, here it is symptomatic of data being overfitted as we are working with a very simple system. This simple analysis enabled thus to continue with the analysis of insulin data.

---

<sup>00</sup> The infinity norm is defined for a matrix  $X$  ( $m \times n$ ) by  $\|X_\infty\| = \max(\sum_{j=1}^n |x_{1j}|, \sum_{j=1}^n |x_{2j}|, \dots, \sum_{j=1}^n |x_{mj}|)$ . It is easier to see in the case of a vector  $X$  ( $I \times n$ ),  $\|X_\infty\| = \max(|X_j|)$  for  $j = 1 \dots n$ , that inf-norm corresponds to a simple normalization to the maximum variable. For an emission spectrum, this corresponds to a normalization at the wavelength  $\lambda$  of maximum intensity.

## 5.7 Analysis of insulin UV-Vis data.

UV-Visible spectra from all the insulin samples in the different buffers were measured primarily to verify the concentration of insulin in each buffer, and to ensure reproducibility between the various replicate measurements.

Data were collected at 0.2 and 2 mg/mL, and the corresponding buffer baseline automatically subtracted during the collection. Only three replicate measurements could be made per day at each concentration. To get to the 12 replicate measurements per insulin form, samples were made fresh before analysis and the concentration adjusted so that a minimal deviation was observed between series of replicate measurements (Table 20).

**Table 20: Average (AVG) and standard deviation (STD) of the absorbance at 276 nm for the 12 replicates analysed per insulin form\*. The AVG was converted to mg/mL using a molar extinction coefficient of  $1.0 \text{ cm}^{-1} \cdot (\text{mg/mL})^{-1}$  for 1 mg/mL at 276 nm<sup>76</sup>.**

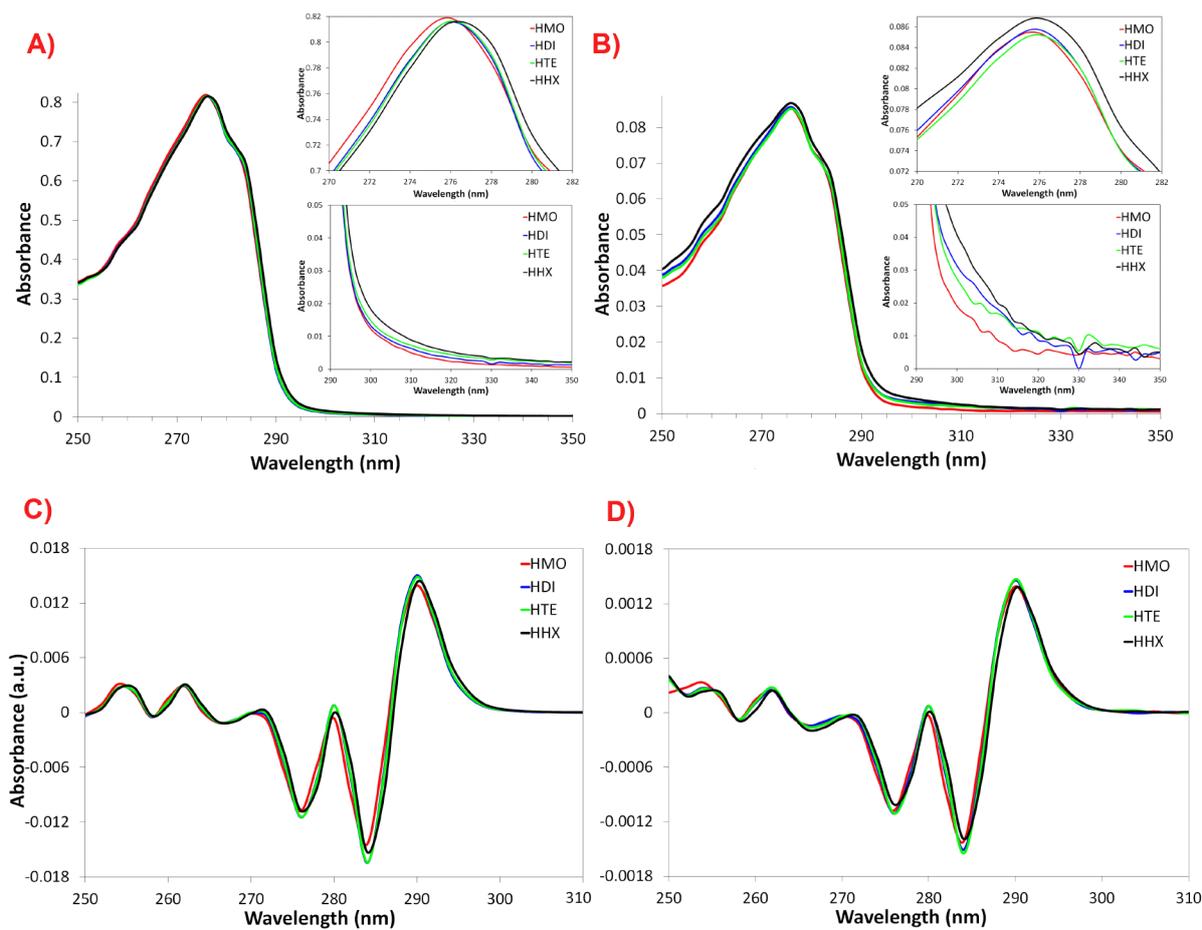
<b>High Conc.</b>	<b>HMO</b>	<b>HDI</b>	<b>HTE</b>	<b>HHX</b>
AVG	0.819	0.816	0.816	0.815
STD	0.007	0.015	0.010	0.008
mg/mL	2.047	2.040	2.041	2.038
<b>Low Conc.</b>	<b>HMO</b>	<b>HDI</b>	<b>HTE</b>	<b>HHX</b>
AVG	0.085	0.086	0.085	0.087
STD	0.001	0.002	0.002	0.003
mg/mL	0.214	0.214	0.213	0.217

\*HMO stands for human monomer, HDI for human dimer, HTE for human tetramer and HHX for human hexamer.

It appeared that the UV-Vis spectra collected for the various insulin forms also contained interesting information to start characterising the association states spectroscopically (Figure 56).

As previously reported<sup>52,473,494</sup>, the absorption maximum is obtained around 276 nm for insulin (Figure 56). Compared to free L-Tyr that showed in all buffers a maximum at 274 nm as it was exposed to polar environments (see Figure 55), the maximum in insulin is thus red-shifted by two nm because some residues like TyrA19 are buried in a more hydrophobic environment<sup>495-497</sup>. At first sight, this shift might not appear significant and even a bit dubious. However, it is known that changes in Tyr environment only produce small absorbance or excitation shifts<sup>494</sup>.

5. Elucidating the intrinsic fluorescence of human insulin and its different association states



**Figure 56:** UV-Vis spectra collected at 2 (A) and 0.2 (B) mg/mL, averaged over the 12 replicate measurements collected for each of the insulin forms (HMO: human monomer, HDI: human dimer and so on). The insets are used to focus on the top and tail of the absorbance peak, enabling to better see the small but real red shift that occurs going from HMO to HHX. The second derivative spectra of the 2 and 0.2 mg/mL samples are also shown in C/D and were obtained without pre-smoothing the data.

NAYA, often used to model Tyr in protein, was for example proven to shift by a maximum of 4 nm (274 to 278 nm) going from a completely polar solvent like water to a non-polar solvent like dioxane<sup>33</sup>. Proteins are expected in fact to show even smaller shifts than NAYA, due to the fact that often several Tyr residues are absorbing simultaneously, each surrounded by a different environment. The crucial point though, is that as observed in many other studies involving class A proteins<sup>496-500</sup>, shifts even smaller than 2 nm are real and indicate profound changes in Tyr environments.

An interesting feature in Figure 56 is that the red shift slightly increases upon formation of dimers and hexamers, which is particularly clear for the 2 mg/mL samples. This shift is very small, which is not surprising considering that in HMO (see section 5.2.1.4), three out of four Tyr residues are already either completely (TyrA19) or partially (TyrB16 and TyrB26) buried

in hydrophobic environments. However this progressive shift of the  $^1L_b$  band to lower energy is real, and is diagnostic of Tyr residues getting surrounded by a more hydrophobic (or non-polar) environment<sup>495</sup>.

Two mechanisms can explain why the Tyr absorption is slightly red-shifted in hydrophobic environments<sup>33,464</sup>. One of them is the appearance of hydrogen bonding between the Tyr and its surrounding environment<sup>461</sup>. In particular, the absorbance is red-shifted only if the phenol group of Tyr acts as a proton donor<sup>464,494</sup>, the extent of shift depending on the hydrogen acceptor strength<sup>33,464</sup>. This aspect of Tyr absorption is explained by quantum mechanics and is related to the contribution of the p-orbital (electron lone pair) of the phenyl oxygen to the HOMO and LUMO, which make them dependent upon excitation on charge transfer occurring from this oxygen to the ring<sup>33,464</sup>. Indirectly, this p-orbital contribution makes both the HOMO and LUMO sensitive to any type of molecular interaction that involve that oxygen, in particular hydrogen bonding. It was found through molecular orbital calculations that the contribution of the oxygen p-orbital to the HOMO and LUMO was lowered if the solvent was a proton donor and increased if the solvent was a proton acceptor. Considering that the contribution of the p-orbital is also stronger for the HOMO than the LUMO<sup>33</sup>, it explains why Tyr absorbance is red-shifted if the solvent (or ligand) is a proton acceptor but would be blue-shifted if the solvent was a proton donor<sup>33,464</sup>.

Hydrogen bonding however only partially explains the red-shift in Tyr spectra. Another crucial parameter is the solvent dielectric effect<sup>33,464</sup>. This effect is quite complex and can be inversed depending on the values of the excited and ground states dipoles. Overall it is shown that if the excited state dipole is smaller than the ground state, an increase of the dielectric effect due to an increased solvent polarity causes a blue shift<sup>33,464</sup>. In the opposite case where the excited state dipole is larger, the dielectric effect causes a red shift. For Tyr, molecular orbital calculations have shown that the excited state dipole is smaller than the ground state dipole, probably due to the larger contribution of the p-orbital of the phenyl oxygen to the HOMO than the LUMO<sup>33</sup>. This has been proven also for *p*-cresol (4-methyl phenol) in various configurations<sup>464</sup>. It is therefore logical that for Tyr, a solvent with a high dielectric constant induces a blue shift in the absorbance whereas a solvent with a low dielectric constant induces a red shift<sup>33,464</sup>.

Considering these two phenomena, the formation of insulin natural oligomers is consistent with a slight spectral red-shift in the absorbance. Regarding the first mechanism (*i.e.*

hydrogen bonding), we know that because the pKa of the phenyl hydroxyl group in Tyr is 10.3, this group remains protonated in the various buffers used (maximum pH is 7.4). This means that when Tyr residues become buried in the hydrophobic dimer and hexamer interfaces, they can get close to potential hydrogen donor/acceptors but can only act as a proton donor, which thus make sense with a red-shift.

Regarding the second mechanism (*i.e.* the solvent dielectric effect), we know that only a solvent with a low dielectric effect such as non-polar solvent can result in a shift to the red. In a protein however, the microenvironment of a residue (especially if the residue is buried) is not only defined by the solvent but also by the other amino acids that surround it. In this case it is more correct to consider the protein dielectric constant rather than the solvent one alone when characterizing the residue's microenvironment. A paper has been published recently that examined the dielectric "constant" of proteins<sup>501</sup>. What they found is that the dielectric values vary a lot between protein domains. In particular, hydrophobic domains were found to have a low dielectric constant while exposed domains had higher values. Seeing a red-shift upon the formation of insulin dimers and hexamers is thus consistent with Tyr getting buried in hydrophobic domains.

In addition to the spectral shift, it is also interesting to note the presence of shoulders or structural elements in the absorbance spectra of insulin which were absent in the free L-Tyr spectrum (Figure 55). We know that these structural elements are not due to the buffer as none of the free L-Tyr spectra presented this kind of detail; instead they are present for all insulin forms (Figure 56). One of the main elements is seen at 258 nm and is due to the absorbance of Phe residues<sup>52,502</sup>. Another less pronounced element is seen at 264 nm and is also related to Phe residues<sup>502-503</sup>. Note that there can also be disulfide bridges that absorb at 260 nm.

The second piece of information that can be extracted from the absorbance spectra is related to the light scatter by the protein. It is known<sup>504-505</sup> that as light scatter from a solution increases, it can be observed as an increase in the measured absorbance in the near-UV or Vis region (typically 350 nm) where the light is not absorbed but solely scattered by the protein. A method called the aggregation index (AI) is based on this and derives a ratio (see Equation 58) that is used to determine the degree of aggregation of proteins<sup>504-505</sup>. This method is applicable here because the different insulin association states were collected at the same concentration. But in order to have all samples to be exactly comparable, we normalized all spectra to one

(infinity norm). The results of the AI values for the different association states of insulin are shown Table 21.

**Equation 58**

$$AI = \frac{A_{350}}{A_{280} - A_{350}} \times 100$$

**Table 21: Average and standard deviation obtained for the AI calculated for each insulin form across the 12 replicate measurements, at 2 and 0.2 mg/mL.**

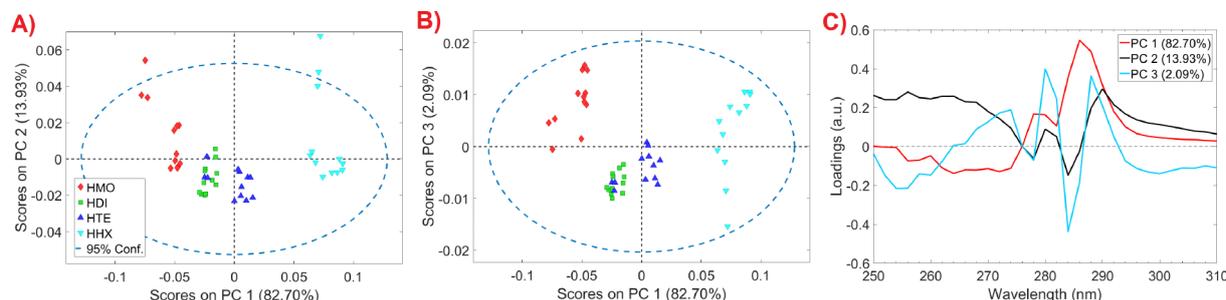
<b>High Conc.</b>	<b>HMO</b>	<b>HDI</b>	<b>HTE</b>	<b>HHX</b>
AVG	0.20	0.23	0.28	0.35
STD	0.07	0.06	0.05	0.07
<b>Low Conc.</b>	<b>HMO</b>	<b>HDI</b>	<b>HTE</b>	<b>HHX</b>
AVG	0.41	1.00	1.12	0.87
STD	0.13	0.39	0.52	0.20

For the high concentration samples (2 mg/mL), even if the large STD can result in overlapping AI values (*e.g.* HMO/HDI), an increase in AI going from monomeric to hexameric buffer can still be observed, which indicates an increase in particle size. For the low concentration samples, the results are however less clear which is due to the higher amount of noise present in the spectra (equivalent inset window in graph B). At low concentration, the light scattered by the sample ceases to be dominated by the protein scatter and becomes more sensitive to the presence of dust, scratch on the cuvettes etc. Combined with the intrinsic loss in SNR due to the concentration being lowered, this explains the limited results obtained at 0.2 mg/mL.

Even if the spectral shifts are very small and that a consequent amount of noise is present, it is still worth trying to apply some chemometrics analysis on these UV-Vis spectra to see if a possible separation of species can be achieved. To do so, all samples were normalised (infinity norm) to one and the spectra were reduced to a region between 250 and 320 nm to decrease the amount of noise being modelled. Normalization was done to remove every potential influence of raw intensity on the models. However due to the good reproducibility of sample concentrations (Table 20), it was expected that non-normalized data would give very similar results. The dataset is of course bilinear (48 samples  $\times$  31  $\lambda_{ex}$ ) and suitable to models like PCA or MCR. Both will not however explain the data in the same way.

A two component PCA model was built first using the high concentration data (Figure A- 39). However as was expected from such overlapping data, PC1 dominates the entire variance and explains over 99.9% of the data. To try and better discriminate the insulin species,

the data was mean-centred. This pre-treatment removes the mean spectrum from all data and helps targeting the very small but real chemical variations (Figure 57, A/B/C).



**Figure 57: (A/B/C) Scores scatter plots and loadings for the 3 components PCA model built using the mean centred 2 mg/mL absorbance spectra.**

The model developed on mean-centred data explained over 98.7% of the data. Apart from the overlap of replicates R10-11-12 of the human tetramer (HTE) with HDI, the model showed that the insulin species can be separated based on very small difference in their absorbance spectra. The separation is done along PC1 (82.7% of the data) that is characterised in the spectral loading by a peak centred  $\sim 286$  nm with a shoulder around 278 nm. This peak relates to the spectral region in the absorbance spectra (Figure 56, A) where the aggregation generated a slight red-shift, which starts around 276 nm up until  $\sim 310$  nm. This region is in fact a decreasing spectral slope that is shifted because the absorbance peak itself is shifted<sup>495</sup>, but also because scattered light is present in the peak tail (Figure 56, A both insets).

PC1 seems thus to separate the insulin forms based on spectroscopic features. The points are however not very clustered (Figure 57A) which is a consequence of mean-centring that enables one to drag out and look at smaller variations in the spectra. The separation for PC2 vs PC3 was not as good. This plot is shown in Figure A- 40, along with the Hotelling's  $T^2$  statistics versus the Q statistics<sup>PP</sup> that is used to detect the presence of outliers.

We saw that the model failed to separate HTE R10-12 with HDI, which mean that HTE R10-12 might be made mainly of dimers. Considering the nature of the tetrameric state, this is not surprising. As explained earlier in the chapter, the tetrameric state is ambiguous and considered to be a transient state between the dimeric and hexameric states, which is a point confirmed by the scatter plots (A/B) as the HTE samples lie somewhere between HDI and

<sup>PP</sup> The two statistics Hotelling's  $T^2$  and Q are commonly used to look at outliers in PCA analysis. The Hotelling's  $T^2$  calculates the sum of the normalized squared scores for each sample, allowing to measure their variation *within* the model. The Q statistics is a measure of lack of fit per sample, which is calculated as the difference between the data point and its projection on the model. Q measures therefore the variation *outside* the model. Of course, a high value in each statistics means either that the sample is different or that the model is not explaining it well.

HHX. The HTE samples are however much closer to HDI and if we consider that the tetrameric state is in fact a loosely bound construct/equilibrium between dimers and hexamers, then one could conceive that every small difference in the sample preparation (*e.g.* longer time left on the bench) could potentially switch the equilibrium towards the dimeric state. This is an important fact that should be remembered for the rest of this chapter, emphasizing how difficult it might be to separate HTE from HDI in particular.

A similar model using the low concentration data failed in separating the insulin states (see Figure A- 41 in Appendices). It failed because even with mean centring, the separation is still largely influenced by the presence of scatter and noise (see loadings in Figure A- 41, D below 286 nm), and we saw earlier that at low concentration the scatter ceases to be dominated by the protein scatter. We tried to eliminate the scatter tail part by selecting the absorbance values from 250 to 290 nm hoping it would help the model to focus on eventual shifts in the absorption peak, but the separation is not improved. But the model failed also because the quality of the data is too poor (low resolution and high level of noise). The 3 PCs model explained 98.63% of the data and is shown in Appendices, Figure A- 42.

One general issue with PCA is that the spectral loadings can be hard to interpret as they do not relate to a pure chemical species, which is a consequence of the orthogonality imposed between components. Understanding PCA loadings can already be tricky with raw data, but the use of mean-centring can make everything worse as shown in Figure 57 or Figure A- 42. Eventual contributions of Phe around 258/264 nm can maybe emerge from the loadings in Figure A- 42-D, but it is only hypothetical. MCR on the other hand can extract pure components, that are meaningful with the use of the appropriate constraints like non-negativity. The detail of this analysis can be found in section 7.19, in Appendices. The MCR model only managed to separate the insulin species at high concentration, and thus confirmed the PCA results

Therefore, this study shows the limits of UV-Vis analysis for characterising insulin aggregation, at least using curve resolution techniques and this kind of data quality/resolution. Based on very small difference in the absorbance spectra, a separation can be achieved on the 2 mg/mL dataset, even if the data quality is poor (low SNR and resolution). The components extracted however did not correspond to pure Tyr contributions. To obtain such components but also to further explore and perhaps separate the 0.2 mg/mL samples based on resolved spectral properties, we move on to fluorescence.

## 5.8 PARAFAC modelling of insulin monomer emission

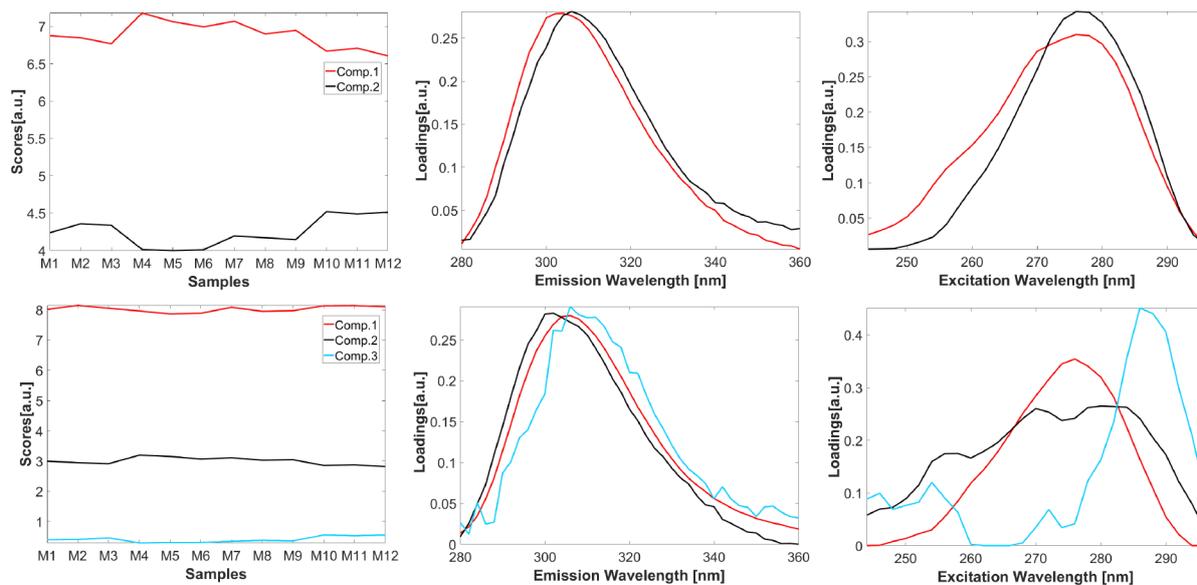
As a starting point in the modelling of insulin MDF data, it seems reasonable to start by modelling only the simpler of the association states *i.e.* monomeric insulin. The PARAFAC models (Table 22, Figure 58) were built using the 12 replicate measurements collected at 0.2 mg/mL, pre-treated following the steps described previously.

With an absorbance below 0.09, this dataset does not need IFE correction which is safer than having to transform/modify the data. The data were normalised to 1 (infinity norm) and the PARAFAC models built on EEM<sub>||</sub> using the convergence criteria and constraints mentioned previously. The split-half analysis<sup>99</sup> is also added at this stage as an additional criterion to judge the quality/robustness of the PARAFAC decomposition.

**Table 22: Summary of the fit captured by each component for the normalized EEM<sub>||</sub> PARAFAC models of the 12 replicate measurements of the insulin monomer. The CC, total variance captured by the model and results from split-half analysis are also indicated.**

	<b>EEM<sub>  </sub> 1 comp.</b>	<b>EEM<sub>  </sub> 2 comp.</b>	<b>EEM<sub>  </sub> 3 comp.</b>
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	276/304	276/304	276/306
<b>Fit model (%)</b>	100	72.39	87.54
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	276/306	280/302
<b>Fit model (%)</b>	/	27.61	12.22
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	/	286/306
<b>Fit model (%)</b>	/	/	0.23
<b>Variance explained (%)</b>	99.97	99.98	99.98
<b>CORCONDIA (%)</b>	100	99.16	-1729.64
<b>Split-half (half, %)</b>	100	95.9	36.2
<b>Split-half (random, %)</b>	100	99	33.8

<sup>99</sup> The concept behind this method is quite simple and says that a PARAFAC model is robust if the model built on the entire data is identical to models built on subsets of that data<sup>506</sup>. The name “split-half” means literally that independent models are run on both halves of the data and the loadings compared to the global model. The more alike the loadings are (in %), the more robust the model is. This method is of course very sensitive to the order in which the samples are arranged in the data. Assigning for example dissimilar samples to both halves will result in a difficult split-half validation. For the 12 monomer replicate measurements analysed here, the order has no importance as the samples are all meant to be identical. The split can be done either by cutting the data in half therefore preserving the original sample order, or by forming two halves from a random selection of samples that is considered even more robust<sup>289</sup>. Both methods were run in the thesis. There was not much sense to do a split-half analysis on the free L-Tyr data as there was only three replicates per buffer. For the insulin analysis however, this analysis is interesting and can also be useful for determining the appropriate rank of a model.

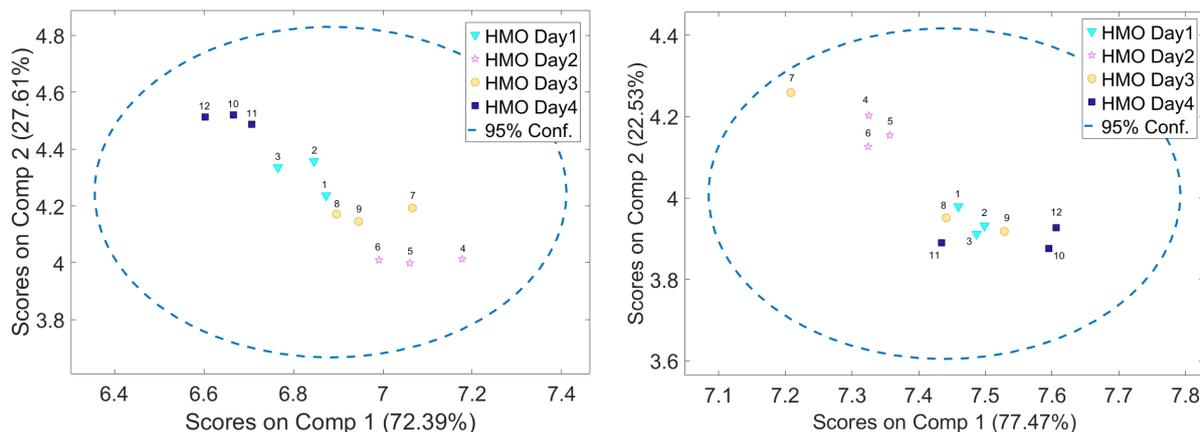


**Figure 58: PARAFAC models built using the normalized pre-treated EEM<sub>||</sub> data consisting of 12 monomer replicate measurements, collected at 0.2 mg/mL. The scores and spectral loadings are shown for the two (top row) and three (bottom row) component models.**

One first comment from these models considering the shape of the spectral loadings and the values in Table 22, is that the PARAFAC model with three components is clearly overfitting the data. The repetition of models is also highly unstable but not shown here.

On the other hand, the fit with two components is good. The loadings are compatible with Tyr residues, with excitation maxima obtained at 276 nm and emission maxima that lie between 304 and 306 nm. The shape of the excitation spectra for the first component that represents >70% of the EEM<sub>||</sub> model is a bit odd however, even if the shoulder seen ~ 256 nm could be explained by the contributions of Phe to Tyr spectra via FRET. The other attractive features of this model are the high CC and split-half results (Table 22), as well as a high reproducibility over multiple runs as seen in Figure A- 46 and Table A- 17.

The scatter scores plot (Figure 59, left) shows samples that are clustered by day of data collection. There is however no specific order in the days and finding an explanation to this separation is difficult. It is possible that the model separate the HMO series based on their slight difference in oligomer population, but there is noise and residual scatter issues to consider. More importantly, the components scores show a certain degree of correlation/collinearity, which is maybe less clear than in the L-Tyr case, but that is still an important warning about potential over-modelling. (Figure 59, left).



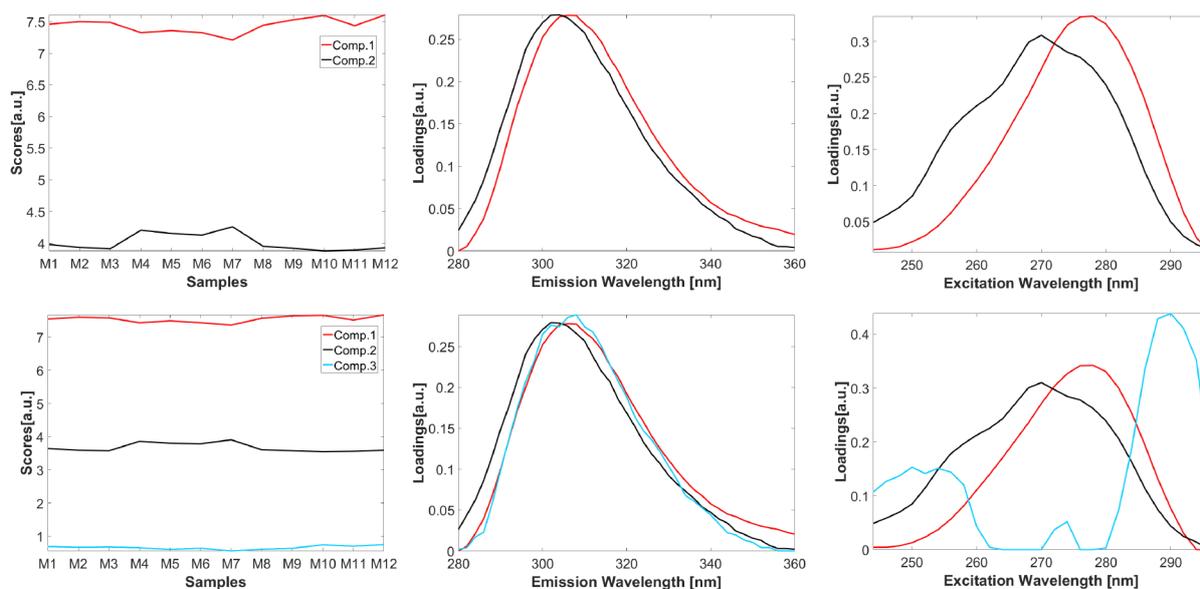
**Figure 59: Scores scatter plot (Comp.1vsComp.2) obtained for the 2 components PARAFAC model built on the normalised  $EEM_{||}$  (left) and  $EEM_{\perp}$  (right) datasets consisting of 12 insulin monomer replicate measurements collected at 0.2 mg/mL.**

To further analyse and investigate the potential validity of these components, the same procedure was performed in the  $EEM_{\perp}$  data. The details of the models obtained are shown in the Table 23 and Figure 60 below.

**Table 23: Summary of the fit captured by each component for the normalized  $EEM_{\perp}$  PARAFAC models of the 12 replicate measurements of insulin monomer. The CC, total variance captured by the model and results from split-half analysis are also indicated.**

	$EEM_{  }$ 1 comp.	$EEM_{  }$ 2 comp.	$EEM_{  }$ 3 comp.
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	276/304	278/306	278/306
<b>Fit model (%)</b>	100	77.47	80.33
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	270/304	270/302
<b>Fit model (%)</b>	/	22.53	19.05
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	/	290/308
<b>Fit model (%)</b>	/	/	0.62
<b>Variance explained (%)</b>	99.96	99.98	99.98
<b>CORCONDIA (%)</b>	100	-199.1	-2027.1
<b>Split-half (half, %)</b>	100	97.1	76
<b>Split-half (random, %)</b>	100	97.9	71.3

Similar to the  $EEM_{||}$  case, the three component model can be discarded and only the two component model preserved. If we look at Figure 58 (top row) and Figure 60 (top row), the spectral loadings are similar. One of the main differences however is that the Comp.1 which had a structured excitation in the  $EEM_{||}$  model, is now Comp.2 in  $EEM_{\perp}$ . The shape of this component is also more suspicious, especially along the excitation where the shoulder at  $\sim 258$  nm seem to have a greater weight on it compared to what was observed for  $EEM_{||}$ .



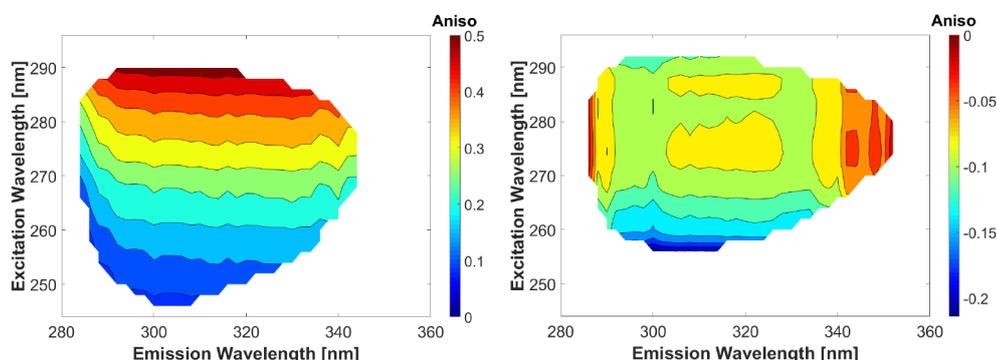
**Figure 60: PARAFAC models built using the normalized pre-treated  $EEM_{\perp}$  data consisting of 12 monomer replicates, collected at 0.2 mg/mL. The scores and spectral loadings are shown for the two (top row) and three (bottom row) component models.**

One of the main photo-physical elements that was absent from free L-Tyr in solution is the Phe to Tyr FRET seen at  $\lambda_{ex}=258$  nm. Considering that FRET is a depolarising process<sup>53,78,457</sup>, it is reasonable to assume that the Phe contribution should be higher in  $EEM_{\perp}$  compared to  $EEM_{\parallel}$  which is apparently what we can see in the excitation loading of the black component (Figure 60, top row). On its own, this would be a significant result for ARMES because the contribution of Phe in protein fluorescence is generally difficult to observe<sup>28</sup>. But this is also what betrays the validity of these models, at least if one wants to resolve the Tyr contributions. In fact, what happens is that by trying to resolve the monomeric insulin fluorescence, the PARAFAC model which always tries to explain the data in a trilinear way, uses these small variations in the excitation shape to create a pseudo component around it, that unfortunately does not relate to the real Tyr contributions. The different clues leading to that conclusion are shown below and changing model parameters such as the initialisation method (using *e.g.* orthogonalized random numbers) did not change the result:

- (i) For both  $EEM_{\parallel}$  and  $EEM_{\perp}$ , the second component (resolved alongside the component with a shoulder) is always free from any structure in the excitation, which proves that all that information has already been used to model some pseudo component around it.
- (ii) Considering the photophysical behavior of Tyr in the absence of ground or excited state proton transfer, a much stronger overlap between Tyr residues should be seen as Tyr is almost insensitive to the polarity of its environment. We saw previously that a 2

to 4 nm shift in maximum excitation is generally observed between an exposed/buried Tyr, which is far from what is observed especially in the  $EEM_{\perp}$  model.

- (iii) The way the components are swapped between  $EEM_{\parallel}$  and  $EEM_{\perp}$  is a clear indication about the influence of the spectral shoulders on the components that are being resolved. In  $EEM_{\parallel}$  where the shoulder is smaller, the model cannot really dissociate them from Tyr resulting in a spectral shape that is not too distorted. In  $EEM_{\perp}$ , the shoulder contribution is strong enough so that the model tries to represent it on its own. This results in a component that explains much less data variance and that is responsible for a smaller part of the model fit (22.53% versus 72.39%). This also results in a spectral shape that is much more distorted (and consequently a negative CC). At this point, one can already foresee the issues that may arise during the PARAFAC resolution of  $EEM_{\perp}$  versus  $EEM_{\parallel}$ .
- (iv) Lastly, one way to look at the validity of the components is to plot their respective *aniso*-MDF plots which is one of the advantages of ARMES. As shown in Figure 61, the plot obtained for Comp.2 does not make sense, neither in terms of shape nor in terms of values. As explained previously, if a pure Tyr residue was resolved, we would expect to get an anisotropy that would be constant along the emission and that would increase from negative to positive values along the excitation, corresponding to light absorbed respectively by the  $^1L_a$  and the  $^1L_b$  states, but emitted by  $^1L_b$ . In that regard, the shape obtained for Comp.1 makes much more sense as horizontal contours are obtained along the emission. However even if the anisotropy increases along the excitation, the values obtained do not make sense as it exceeds 0.4 in the top part of the plot and does not reach negative values below  $\sim \lambda_{ex} = 250$  nm.



**Figure 61: Recomposed *aniso*-MDF plots for Comp.1 (left) and Comp.2 (right), averaged over the 12 monomer insulin samples. Comp.1 and Comp.2 are called after the components order obtained during the  $EEM_{\parallel}$ -PARAFAC model.**

Therefore, the problem is more complex than it appears. The big issue encountered is that the insulin data we are trying to fit is highly non-linear. The first reason for the non-linearity is the presence of internal FRET, and the second is that the components we are trying to resolve are extremely overlapped. Therefore, these components are highly collinear which generates what is known in chemometrics as rank deficiencies<sup>rr</sup>.

In MDF, we are dealing with three modes *i.e.* concentration, excitation, and emission. As long as the fluorophores are non-interacting, PARAFAC is the most suited method and probably the most employed to model the latent trilinear structure of EEM data. Given the appropriate number of factors to be modelled, PARAFAC extracts without ambiguity the pure concentration, excitation and emission profiles of all fluorophores, respecting full ranks in all modes. However when linear dependencies are observed in one or several modes, the underlying PARAFAC structure is perturbed and the model might fail to resolve the data properly<sup>299,300</sup>, which is what happened with the modelling of monomeric insulin. This depends of course upon the degree of deviation from the trilinearity. Linear dependencies in the EEM concentration mode were for example observed when mixed fluorophores co-varied exactly<sup>493</sup>, or during a standard addition experiment, where PARAFAC failed to separate the analyte of interest from an interference spectra that emerged due to matrix effects<sup>507</sup>. However, dependencies can also be observed in the EEM spectral modes and can be present simultaneously in more than one mode. In particular, complex photophysical processes that introduce interaction between fluorophores like FRET<sup>144,299-300</sup> or excited-state proton transfer<sup>508</sup> are believed to create dependencies in at least two modes, namely the excitation and emission modes. The presence of EEM with such dependencies is rare in the literature as mentioned by Chen<sup>144,508</sup> because intermolecular FRET usually requires mM concentrations generating severe IFE that are hard to correct. For that reason the authors considered either to study excited state proton transfer<sup>508</sup> instead of FRET, or bypass the concentration issue by confining both donor and acceptor in micelles so that transfer would occur even when the

---

<sup>rr</sup> The notion of rank deficiency was very well explained by Bro *et al.*<sup>493</sup>, using a simple two-way example:  $M + N \rightarrow P$ . This reaction is monitored over time by collecting spectral data and it is considered that the three species M, N and P are each characterized by a different spectrum. Let us now consider that the two reactants have the same initial concentration. This leads to concentration profiles for M and N that are strictly identical over time or said otherwise perfectly correlated. In that case then, only two “perceived” profiles are obtained in the concentration domain (one for M and N, one for P) whereas the rank in the spectral domain is full with one spectrum per species (M, N and P all have a different characteristic spectrum). In this example there is therefore rank deficiency in the concentration domain, and this illustrates perfectly how certain phenomena, hidden behind the data, can generate different ranks between the concentration and the spectral domain.

concentrations of D and A are low<sup>144</sup>. The case of insulin is different as the energy transfer is internal to the protein, meaning that it is independent from the concentration.

In theory, the presence of collinear factors in one of the modes prevent the PARAFAC model from uniquely determine these factors in the other modes<sup>493</sup>. The factors that are not affected by collinearity are on the other hand uniquely determined, leading to the notion of partial uniqueness. This property is still not well understood, especially if more than one mode present simultaneously rank deficiencies<sup>493</sup>. The reality however is quite different from the theory, since experimental data always has inherent noise that perturbs the pure linear dependencies between factors. Indeed, by adding part of the noise in the factor loadings, the perfect collinearity that was observed until then is broken and the non-uniqueness is lost. Said differently a PARAFAC model, unless applied to noise-free data, will always generate a unique solution that is not rank deficient, even if in theory it should not. This phenomenon called “surface uniqueness” was important to mention as it explains how PARAFAC can still manage to recover reproducibly components that are very collinear (as we will see in 5.10.1 for example).

This leads naturally to PARALIND, which is one of the most important alternatives to PARAFAC to handle data with linear dependencies. What this algorithm precisely does is to inform the PARAFAC model about the presence of dependencies, so that even if noise is present, the model can adopt a latent structure that fits with the theory. The use of such algorithms improved in particular the recovery of information in the concentration modes and enables better predictions than PARAFAC<sup>493,507</sup>. However, one of the major disadvantages of PARALIND is that it does not provide unique answers for the factors that are collinear, and it proposes instead multiple solutions in the spectral modes that may not always make sense<sup>493</sup>. The other major inconvenience of that method is that it requires an *a priori* knowledge of the data if one wants to obtain the right model. This *a priori* knowledge is needed to fill manually the dependency matrix (or dependency matrices) that inform the model about the exact dependencies between factors and requires of course a deep knowledge of the chemical/system that is being modelled.

Along with the non-uniqueness of the solution, this need for *a priori* information is the main disadvantage of all the other methods that have been proposed to handle rank deficiencies. The use of Tucker2<sup>509</sup> or restricted Tucker3 models<sup>299-300,508</sup>, were for example proposed as alternatives to PARAFAC for modelling cases where fluorophores exchanged energy via

FRET. However, these methods suffer either from rotational ambiguities or partial uniqueness, and require an *a priori* knowledge to feel the core array of the Tucker model<sup>510</sup>. This is in a sense almost identical to PARALIND except that 0 is inserted in the core array in order to remove interactions that are chemically meaningless, improving the interpretability and validity of the model.

With insulin, we do not have this *a priori* knowledge. We know that the molecule contains four Tyr residues, but we do not know if they all contribute to the fluorescence emission, either in the monomeric or worse again in the other associated states. We also do not know which residues are exchanging energies, except that FRET extent is important in insulin. Our approach is therefore more exploratory than the studies<sup>144,299-300,508</sup> mentioned above where the authors, by titrating an acceptor molecule with an increasing amount of a donor, knew exactly which FRET species they wanted to resolve. The FRET species is characterised by having the same excitation spectrum as the donor, and the same emission spectrum as the acceptor. Due to this property, it means that the rank of the excitation and emission modes is one less than the concentration mode, as the FRET species does not have its own unique spectra. This is easy to anticipate and efficiently modelled by PARALIND or restricted Tucker3 models. Even better models can be obtained using hard-modelling approaches (*e.g.* hard trilinear decomposition HTD) that fit the multivariate data with a previously known model based on chemical equilibrium or kinetic laws<sup>299-300</sup>.

When there is no *a priori* knowledge about linear dependencies, there are two common approaches to try to break them and resolve the fluorescence. One of them that worked for example in case of 2-D data is to add samples of a different composition/nature with the rest of the dataset and augment the matrix in the mode where the rank deficiency is observed (concentration or spectra)<sup>511-512</sup>. The other one is to design the right data collection. As we were not able to resolve the fluorescence of monomeric insulin, our idea was to use the fact that in its native state, three out of four Tyr residues are located on the surfaces responsible for insulin oligomerization. Therefore, by independently engaging these residues in the formation of dimers and hexamers, we hope to give enough variability in the data to enable PARAFAC to break these dependencies and resolve the various Tyr emitters.

This way to overcome rank deficiencies is thus fairly unique to insulin, and is possible due to the key location of Tyr residues but also because the protein fold<sup>ss</sup> is preserved in the different pH used. Considering the photo-physics of Tyr that generates, in the absence of excited state proton transfer a near absence of spectral shifts with the change in Tyr micro-environment, we can expect to have highly collinear loadings in at least the excitation and emission modes. The collinearity of the factors should however not be perfect as each resolved Tyr should in theory have their own independent excitation and emission spectra as they all have a different microenvironment. In this regard and in the absence of *a priori* knowledge, we needed an exploratory tool to go through the insulin data, and PARAFAC appeared to us as a more suited method than PARALIND. In fact, PARAFAC may be the only practical approach when we have to face an unknown system. Chen *et al.* particularly shared this point of view, and proved in their study that the result from PARAFAC matched almost perfectly the theoretical predictions about the distribution of species in a system exhibiting FRET, both in terms of concentration and spectral shape<sup>144</sup>.

In terms of scores, we should expect changes mainly between the monomer/dimer and the dimer/hexamer samples. Sudden changes in the scores should be seen between them as this is where the environment of Tyr changes the most. The scores should not vary collinearly however because of the many other quenching mechanisms that affect each Tyr residue, meaning that even if there is transfer of energy, the amount emitted might not be equal to quantity of energy received.

## **5.9 Spectroscopic analysis of the insulin ARMES data at 0.2 and 2 mg/mL, or how to dissect the measurement.**

Before looking at the PARAFAC resolution of insulin fluorescence, we first wanted in this section to study in detail the entire ARMES measurement and see if any other useful tool could be extracted that would help to further characterize the proteins.

### **5.9.1 Utilization of the 1<sup>st</sup> order Rayleigh scatter band**

The first element that comes in mind when protein aggregation is mentioned is of course the increase in particle size, and one possibility to characterize it in fluorescence is to use the Rayleigh scatter intensity. For example, it is common to excite proteins at 280 nm and measure

---

<sup>ss</sup> Protein fold here means that the tertiary structure of the monomer insulin is preserved at all pHs and during its natural oligomerization.

the scattered light at 280 nm and the fluorescence emission intensity at 340 nm<sup>513-514</sup>. By calculating the intensity ratio ( $I_{280}/I_{340}$ ), an information is obtained that is related to the extent of protein aggregation. This method was called aggregation index (AI)<sup>513</sup> or aggregation rate (AR)<sup>514</sup> depending of the source reference. We decided here to call it AI-F in order to avoid the confusion with AI calculated previously using the absorbance spectra.

With MDF measurements, the calculation of AI-F is very easy as all the required information is contained in the same EEM measurement. We calculated this ratio for the 2 and 0.2 mg/mL samples using the EEM<sub>||</sub> (Table 24). The wavelength at 340 nm was replaced by 306 nm that correspond to the maximum fluorescence emission of insulin (Figure 62).

**Table 24: Average and standard deviation obtained for the AI-F calculated for each insulin form across the 12 replicate measurements, at 2 and 0.2 mg/mL.**

<b>2 mg/mL</b>	<b>HMO</b>	<b>HDI</b>	<b>HTE</b>	<b>HHX</b>
AVG	0.044	0.046	0.053	0.094
STD	0.003	0.003	0.004	0.007
<b>0.2 mg/mL</b>	<b>HMO</b>	<b>HDI</b>	<b>HTE</b>	<b>HHX</b>
AVG	0.530	0.494	0.480	0.774
STD	0.046	0.055	0.051	0.111

The AI-F did not work for the low concentration samples, for the same reason as what was said during the calculation of AI in section 5.7. While the results for the 2 mg/mL are better, they are not that robust either, particularly if one wants to separate monomers and dimers, as both values are largely overlapping if we consider their margin of error. The reason for that failure is related to the influence that the buffer has on the fluorescence of Tyr. Even if this affects less insulin (Figure 63) than it does for free L-Tyr in solution (Figure 55) (as residues in insulin are being isolated from the solvent upon aggregation), it is significant enough to not be able to trust the AI-F. Instead, we want to have a tool that is more robust than the AI-F, and that can give information about the increase in particle size independently of the experimental conditions used. The idea was to look more in detail in the 1<sup>st</sup> order Rayleigh scatter band, that is in most PARAFAC/MCR-EEM studies discarded as it ruins the trilinearity of the EEM data.

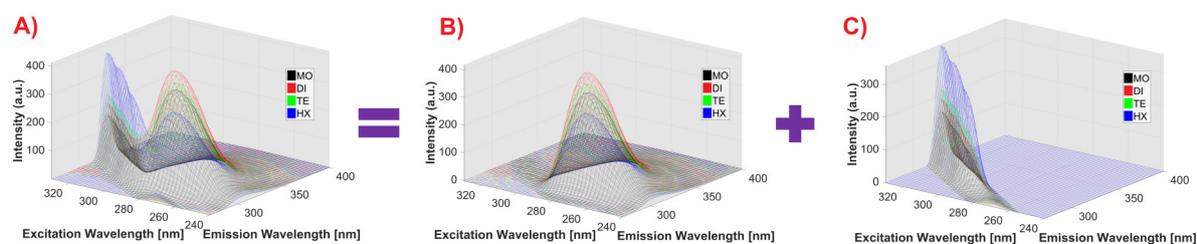
The light that is scattered by a particle, is not scattered identically depending on the size of the particle compared to the wavelength of the incident light. For a uniform and spherical particle, Mie<sup>190</sup> developed a theory that related the intensity of the scattered light with particle

size. This theory holds for particles having sizes less than the wavelength of incident light. Proteins and especially insulin have diameters that are much smaller than the wavelength of incident light (typically  $\ll 1/10$ ), which allows one to simplify the Mie theory<sup>189</sup> and enter the Rayleigh scatter regime. In that case, the light scattered by a particle of diameter  $d$ , at a distance  $R$ , and in the direction  $\theta$  can be expressed following Equation 59<sup>515</sup>.

**Equation 59**

$$I = I_0 \cdot \frac{1 + \cos(\theta)^2}{2 \cdot R^2} \cdot \left(\frac{2 \cdot \pi}{\lambda}\right)^4 \cdot \left(\frac{n^2 - 1}{n^2 + 2}\right)^2 \cdot \left(\frac{d}{2}\right)^6$$

In this equation  $n$  is the ratio between the refractive index of the medium and the refractive index of the absorbing material,  $I_0$  and  $I$  are the intensity of the exciting light and of the scattered light respectively,  $\lambda$  is the wavelength of excitation. It appears then quite clearly that the intensity of the Rayleigh scattered light from a particle is largely influenced by its size, as it is proportional to the sixth power of  $d$ . Of course, when the problem is extended to a dilute protein solution and not a single particle, the total amount of scattered light that is measured comes from all the particles. Its intensity is then dependent on the weight-averaged molecular weight of the particles and on the concentration of the solution, which is what is typically measured by classical or static light scattering (SLS) measurements. As an off-line measurement, it is employed frequently in formulation development to assess the degree of protein aggregation<sup>307</sup>. Its employment in that field is facilitated by the fact that the formulation concentrations are normally kept constant, making the aggregation detection easier. What is interesting is that a conventional right angle ( $90^\circ$ ) fluorometer can also act as a light scattering photometer<sup>307-308,516</sup>, as long as the excitation and emission are set to the same wavelength and that the molecule of interest does not absorb at this wavelength. One aspect of ARMES is the collection of an enhanced 1<sup>st</sup> order Rayleigh scatter band in the EEM<sub>||</sub> measurement. The idea is then to separate both fluorescence and scatter as shown Figure 62, and work independently on both to get complementary information to further characterize protein solutions.



**Figure 62:** Illustration of the way in which ARMES (A) can bring simultaneously (after pre-treatment) a pure fluorescence signal (B) useful for multivariate models, and a pure scattered light (C) useful to monitor the increase in particle size.

We therefore isolated the 1<sup>st</sup> order Rayleigh scatter band using PARAFAC modelling as explained in the previous chapter. We then calculated the volume under the Rayleigh scatter curve ( $V$ ), that is equal to:

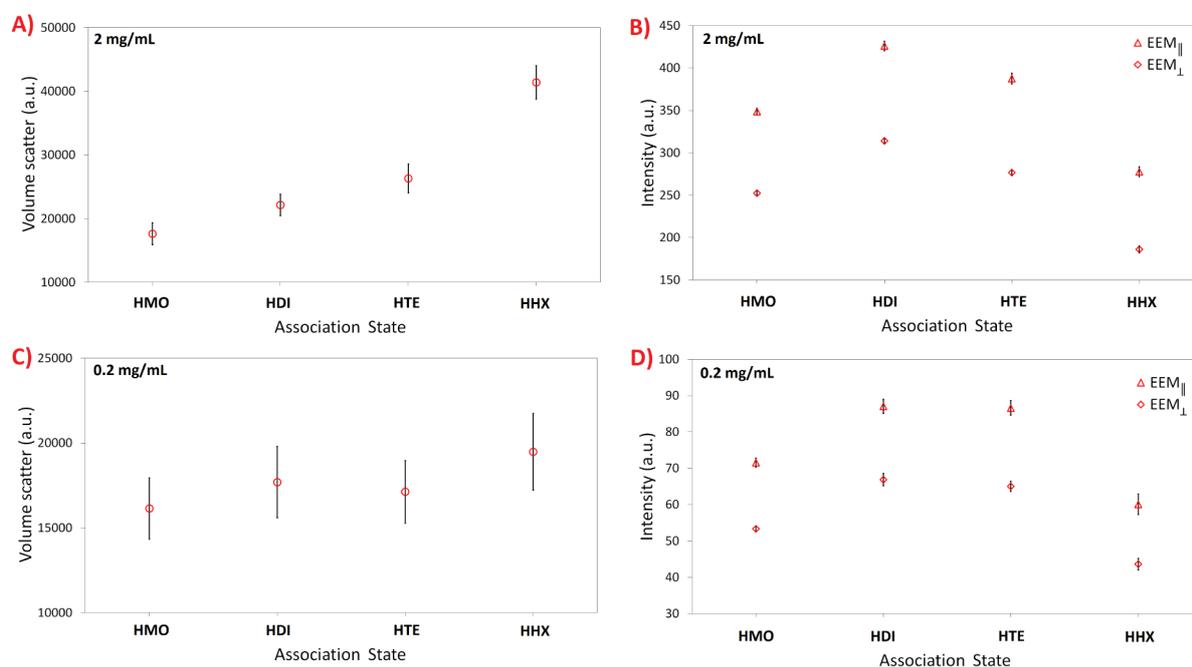
**Equation 60**

$$V = \int_{\text{ex}} \int_{\text{em}} I_{\lambda_{\text{ex}}\lambda_{\text{em}}} d\lambda_{\text{ex}} d\lambda_{\text{em}}$$

With the data collected being discrete (2 nm path interval in both excitation/emission), the volume can be calculated as the sum of all variables (Equation 61). The average values plus errors are shown Figure 63, together with the maximum fluorescence intensity from the raw measurement, for each insulin form.

**Equation 61**

$$V = \sum_{\text{ex}} \sum_{\text{em}} I_{\lambda_{\text{ex}}\lambda_{\text{em}}} \Delta\lambda_{\text{ex}} \Delta\lambda_{\text{em}}$$



**Figure 63:** In A and C is shown the average volume ( $\pm$  STD) calculated under the 1<sup>st</sup> order scatter band of the EEM<sub>||</sub> measurement, at 2 and 0.2 mg/mL respectively. In B and D is shown the average ( $\pm$  STD) maximum fluorescence measured for the blank subtracted EEM<sub>||</sub> and EEM<sub>⊥</sub>, at 2 and 0.2 mg/mL respectively. All values were calculated for each insulin form (12 replicate measurements).

Because we collected the insulin data at the same concentration in all buffers, comparing the scatter volumes (Figure 63) directly relates to the particle size rather than just being influenced by the concentration. The use of the entire band volume as opposed to a single wavelength is advantageous in the way that it reduces the noise associated with the value. We also plan in a future study to try using this band for multivariate modelling and see if useful information can be extracted from it.

As seen previously using AI or AI-F, the scatter at 0.2 mg/mL does not seem to be diagnostic of the size of insulin in solution. The value increases slightly going from HMO to HDI and HHX but a large error is associated with these values. This again is not surprising as most of the scatter is related to the cuvette/dust/etc. In fact, four semi-quartz cuvettes were used during that study. All of them have a different intrinsic scatter and they were used randomly during the data collection. Considering that the large error seen in Figure 63-C (or A) is constant between insulin species, it is more than likely the use of cuvettes with different intrinsic scatter that generated it. The results obtained at 0.2 mg/mL could thus be easily improved if *e.g.* cleaner/newer cuvettes were used or if the same cuvette was used across all measurements. If it is not possible, a control should at least be operated to know which cuvette was used for which measurement.

Insulin is generally formulated at 100 IU/mL<sup>517</sup>, corresponding to ~ 3.5 mg/mL. And at this concentration level, using the 2 mg/mL samples, we proved in Figure 63-A that the method is working well with a clear increase going from HMO to HHX. The values obtained in Figure 63-A are interesting because the scatter going from HMO to HDI or from HDI to HHX is not doubled or tripled. In fact, only non-interacting proteins in solution should give a scatter that is the sum of the individual contributions. Therefore a smaller scatter obtained going from HMO to HDI or HDI to HHX is an indication of protein interaction<sup>516</sup>. It does not seem that the scatter volumes are however varying linearly with the increase in particle size. We mentioned earlier that in solution as opposed to a single particle, the intensity of the scattered light depends also upon the concentration of the solution and the weight-averaged molecular weight of the particles. This relation is called the Debye-Zimm relation and can be approximated for macromolecules of molecular weight < 10<sup>6</sup> g.mol<sup>-1</sup> by the following expression<sup>518</sup>:

**Equation 62** 
$$\frac{Kc}{R_{\theta}} = \left( 1 + \left( \frac{16\pi^2 R_g^2}{3\lambda^2} \right) \cdot \sin^2 \left( \frac{\theta}{2} \right) \right) \cdot \left( \left( \frac{1}{M_r} \right) + 2Bc \right)$$

In this equation  $c$  is the concentration of the solution,  $R_g$  the radius of gyration of the protein,  $M_r$  the weight-average molecular weight,  $\lambda$  the measurement wavelength,  $\theta$  the angle between the scattered and the incident light,  $B$  a virial coefficient used to represent the non ideality,  $K$  an experiment constant.  $R_{\theta}$  is the Rayleigh excess ratio, and is defined if the light is polarized, as the ratio between the intensity of the excess scattered light (once the blank is removed) and the intensity of incident light. Considering that the latter is very stable in the Cary Eclipse fluorometers and that the blank scatter seems to be in major part due to the cuvette

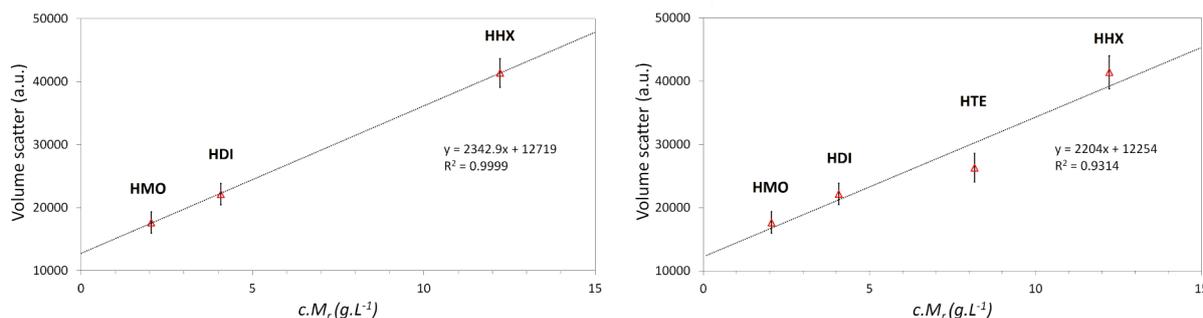
(Figure 63-C), we did the approximation that  $R_\theta$  could be directly replaced by the volume  $V$  calculated above.

As suggested by Brahma *et al.*<sup>516</sup>, this equation can be simplified. First with a concentration of 2 mg/mL (= 0.00035 M) and  $B$  being in mL.mol.g<sup>-2</sup>, the term  $2Bc$  can be ignored. The values of  $R_g$  for the different insulin states were obtained from the work of Nielsen *et al.*<sup>72</sup>, who measured at 2 mg/mL using SAXS the various insulin states in the exact same experimental (buffers) as here. Using  $R_g$  values of 11.6, 14.9, 17.8, and 19.8 Å going from HMO to HHX, the term  $16\pi^2 R_g^2 / 3\lambda^2$  is equal respectively to 0.0011, 0.0019, 0.0027 and 0.0033. As we work with a 90° angle fluorometer,  $\sin(\theta/2)^2$  is equal to 0.5 meaning that the entire term at the end of the first bracket is  $\ll 1$ . The previous equation can therefore be approximated to the simple relation below:

### Equation 63

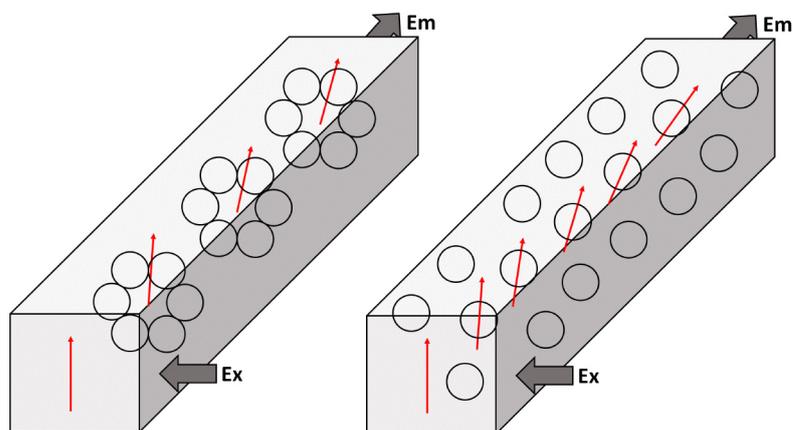
$$R_\theta = KcM_r$$

Without having to go as far as calculating the Stokes radii as suggested by Brahma, this equation can be used by replacing  $M_r$  with the molecular weight of human monomer insulin, dimer etc. This is equivalent to considering the macromolecular solution as ~ homogeneous which should be the case as each buffer favours one insulin form. The results for the 2 mg/mL samples are shown in Figure 64. It is clear from the left plot that the monomers, dimers, and hexamers are perfectly aligned and are well described by this law. Adding the tetramers on the right led to more discrepancy even if the correlation coefficient is not bad ( $R^2 = 0.933$ ). Considering that HMO, HDI, and HHX are the well characterised states and that presence of HTE has always been tenuous (see Figure 57 or Figure A- 44 plus comments), this could be seen as a proof that HTE as opposed to the rest of the insulin states, cannot be considered as homogeneous. It is thus tempting to use the left curve as a calibration curve and estimate the HTE  $M_r$  from the scatter measured. By doing so and taking on board the reasonable assumption that HTE is mainly a mixture of dimers and hexamers, a  $M_r$  estimated at 16,497 g.mol<sup>-1</sup> leads to 79% dimers and 21% hexamers which is plausible with the scores scatter plots from Figure 57 or Figure A- 44. We cannot unfortunately say with certainty if this result is real or not, but it would explain why the “tetrameric” state presents so many spectral parameters that are closer to the dimeric state than anything else.



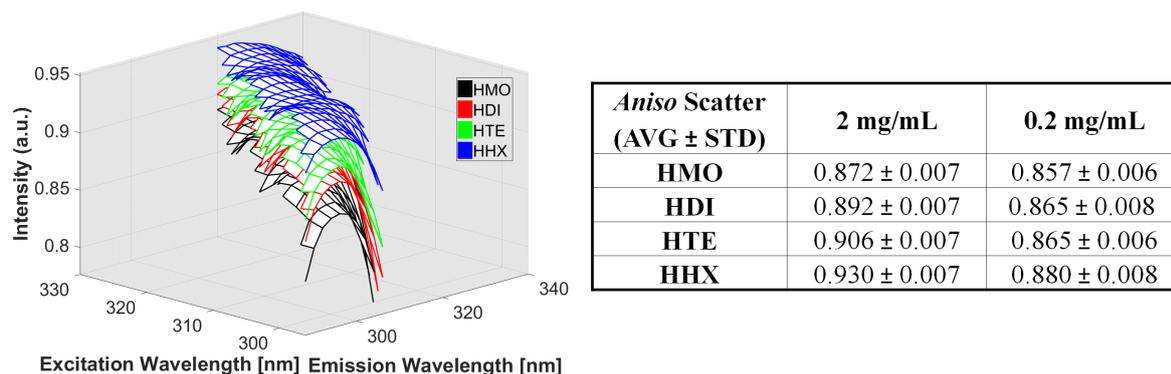
**Figure 64: Linear regression between the volume of the 1<sup>st</sup> order Rayleigh band of the EEM<sub>||</sub> measurement, and the product of the concentration by the molecular weight of the associated protein. HTE is omitted on the left but added to the regression on the right.**

The use of the 1<sup>st</sup> order Rayleigh scatter band of the EEM<sub>||</sub> measurement provides very valuable information, and as expected this method is more robust than for example the use of AI or AI-F. However, we can go further and use an aspect of ARMES which is the anisotropy of the scattered light. In the ideal case of a dilute suspension of pure scatterer, the scattered light should have an anisotropy of 1. However when multiple scattering events are occurring, this light is depolarized<sup>78</sup>. To illustrate this, let us consider the case of two identical volumes in which the same concentration of insulin was dissolved. If the insulin is associated as a hexamer rather than a monomer, it is obvious that the scattered light will have less probability of encountering a particle and be re-scattered in the hexameric case rather than the monomeric one, as shown in Figure 65 below.



**Figure 65: This illustration shows how the light scattered by a particle can be depolarized if multiple scattering events occur. Considering the case of two identical volumes in which the same concentration of insulin was dissolved, the scattered light has less chance to be re-scattered in a solution of hexamers (left) than a solution of monomers (right).**

Therefore, at a constant protein concentration and in a fixed volume, the extent of depolarization of the scattered light should be related to the association state of the molecule.



**Figure 66: (left) Overlap of the average scatter *aniso*-MDF spectra (10% threshold) calculated over the 12 replicate measurement samples, in each insulin form. The table associated (right) summarizes the values where a 5% threshold had to be used to avoid the abnormal low anisotropy regions.**

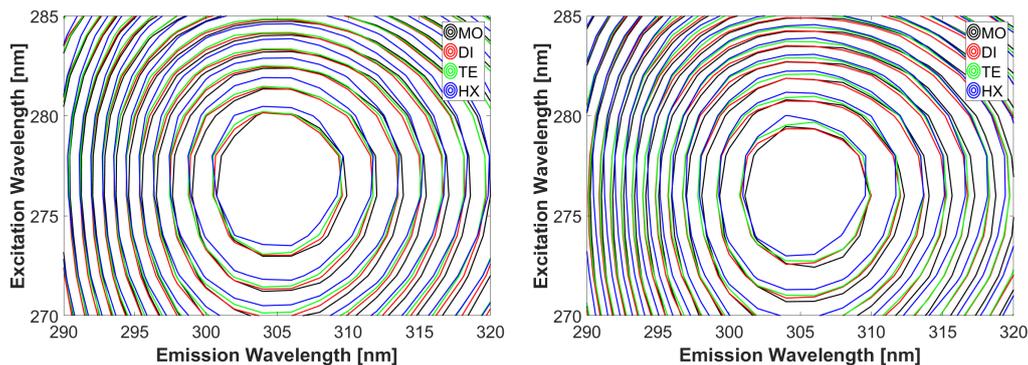
The average *aniso*-MDF spectra obtained on the 2 mg/mL scatter band (Figure 66) prove indeed that at constant concentration, the scattered light is less depolarized if the particle size increases. The values at 0.2 mg/mL also increase slightly, with the advantage to show this time a low error across the 12 replicate measurements. Being able to measure the level of scatter depolarization is therefore useful and provides information that is complementary to the study of the 1<sup>st</sup> order Rayleigh scatter band in the EEM<sub>||</sub> measurement.

### 5.9.2 Appearance of slight shifts in the fluorescence

We can now move on to the pure fluorescence part of the measurement (Figure 62-B). Looking at the EEM<sub>||</sub> and EEM<sub>⊥</sub> fluorescence peak at 0.2 mg/mL (Figure 67), a red shift of about ~ 0.5/1 nm was observed along the excitation with the increase in particle size<sup>tt</sup>. This shift is directly related to the slight red-shift that was previously seen with the absorbance data (section 5.7). This shift is characteristic of Tyr residues being progressively buried in a hydrophobic environment. The emission on the other hand is constant and centred at 305 nm similarly to what was observed by Bekard *et al.*<sup>454</sup> when they monitored bovine insulin fibrillation. This data seems thus to confirm that the excitation spectra of Tyr are more sensitive to a change in environment than their emission.

Because MDF measurements combine the emission and excitation information, one can expect this type of measurements that combine more dimensions<sup>131</sup> to be naturally more suited to resolve insulin fluorescence than only “classically” looking at the fluorescence emission.

<sup>tt</sup> The red shift is also observed at 2 mg/mL, but it is shown in the last part the chapter as IFE also have an impact on the spectral shapes.



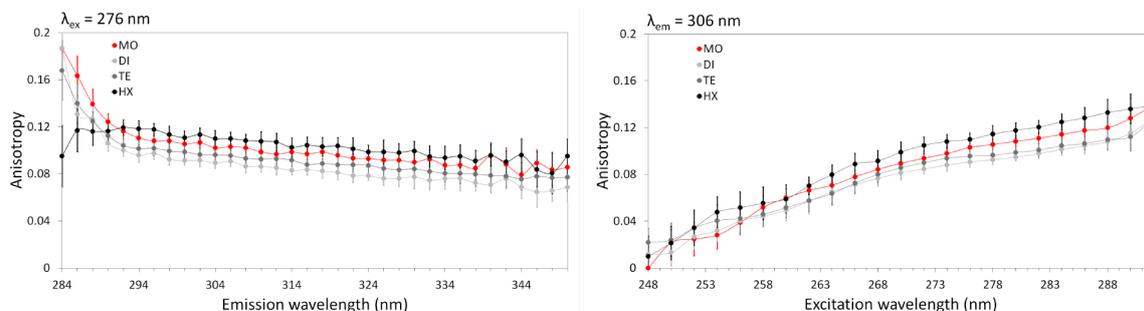
**Figure 67: Overlap of the normalized-averaged  $EEM_{\parallel}$  (left) and  $EEM_{\perp}$  (right) fluorescence signal obtained for all samples in all buffers at 0.2 mg/mL.**

The degree of overlap in Tyr emission is however still very high in both measurements, but we hope that the slight shifts seen in the MDF data created by the change of Tyr environment upon aggregation, may help towards the variability assumption required for successful PARAFAC models.

It was also mentioned earlier in this chapter the possibility of ionization of the phenolic group of Tyr in the ground or excited state. Considering the pH used in this study (Table 4 in section 2.4.1) and the pKa of 10.3 for Tyr in the ground state, deprotonation of the hydroxy group is in theory not possible. However, because the pKa drops between  $\sim 4 - 5$  in the excited state, an emission from tyrosinate is possible in the hexameric buffer (pH 7.4). The fluorescence intensity was thus plotted for  $\lambda_{ex/em} = 295/345$  nm and as shown in Figure A- 47, no signal was found for tyrosinate in the different buffers used. Di-tyrosine can also emit but at a range that is outside of the one used for the data collection ( $\lambda_{ex/em} = 315/420$  nm).

### 5.9.3 Analysis of the 2-D *aniso* and *aniso*-MDF data

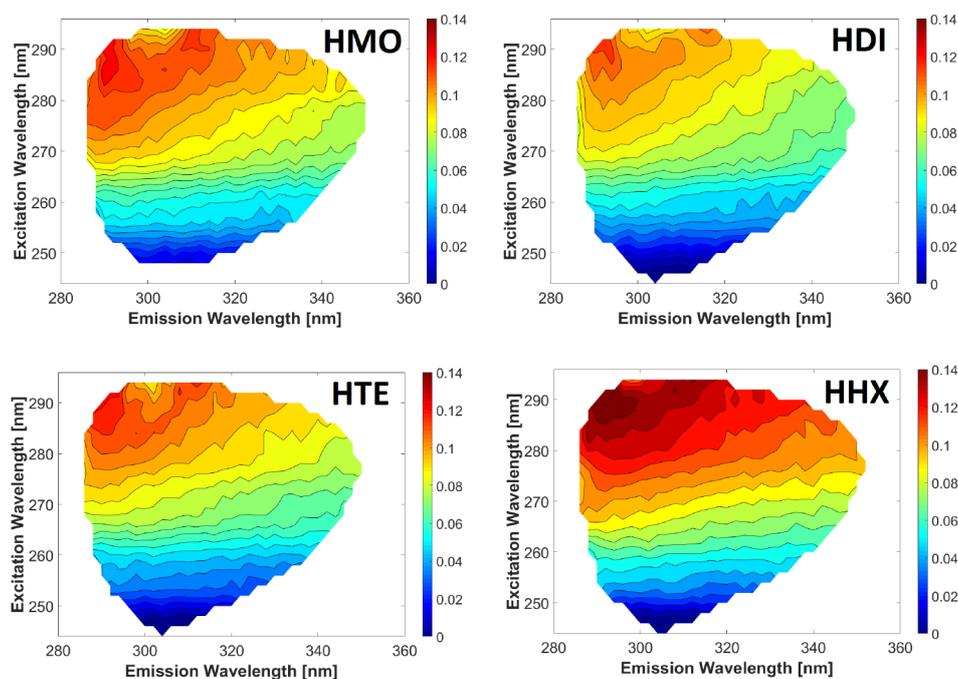
We started looking at the anisotropy data with simple 2-D plots (Figure 68). A typical increase along the excitation wavelength and decrease along the emission is observed, which is typical for class A proteins<sup>519</sup>.



**Figure 68: Average anisotropy values ( $\pm$  STD) obtained for each insulin form over all the 0.2 mg/mL samples. The data is shown along the emission ( $\lambda_{ex} = 276$  nm, left) or along the excitation ( $\lambda_{em} = 306$  nm). These wavelengths correspond to Tyr maximum fluorescence.**

It was surprising however to see the anisotropy dropping slightly going from the monomeric to the dimeric form. Unfortunately, we could not find any specific insulin references to validate these values. One possible cause for that decrease can be the Tyr homo-FRET, which is a well-known mechanism for fluorescence depolarization upon oligomerization<sup>520-521</sup>. Indeed between the different association states studied, dimerization is accompanied by the largest structural changes<sup>17</sup>. It brings TyrB16, PheB24, and TyrB26 from one monomer into close contact with their antiparallel-facing homologues. This generates shielded aromatic cages where extensive homo-FRET can occur, as the Tyr are separated by only  $\sim 5$  Å. At this point, 3 of the 4 insulin Tyr are buried and will remain in an unchanged environment upon formation of tetramers or hexamers as the dimer is their sub-unit. The formation of dimers is probably also accompanied by the largest change in energy redistribution among Tyr residues. Taking the dimer anisotropy value as a reference, higher-order oligomers composed of dimers should then see their intrinsic anisotropy increase, which is what we see for the tetramers and hexamers in Figure 68. This theory seems thus to make sense.

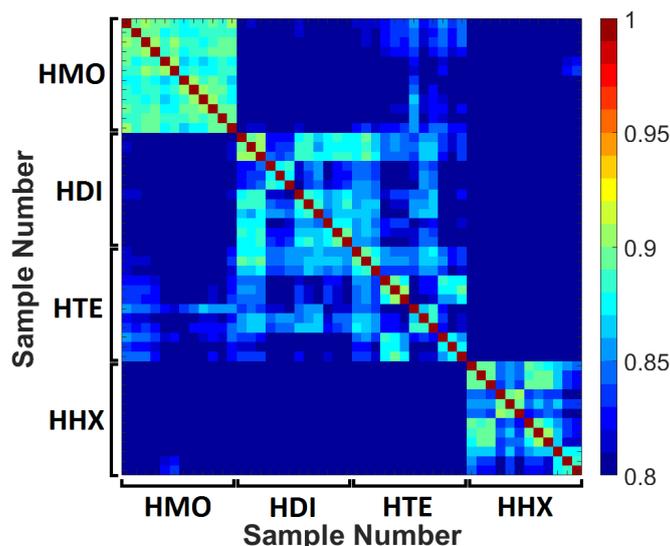
The main drawback of the 2-D plot is that it is not very diagnostic, especially for the excitation. For example, at  $\lambda_{ex} = 265-275$  nm where the intensity is high, monomers and dimers are hard to distinguish. Instead more and better information can be obtained if we look at the *aniso*-MDF spectra for each insulin form (Figure 69).



**Figure 69:** Average *aniso*-MDF spectra calculated for each insulin form over the 12 replicates collected at 2 mg/mL. The colour code corresponds to the anisotropy values.

These spectra, calculated from the raw measurements post-Rayleigh scatter correction, clearly show that each plot is unique to each insulin form, indicating that the contribution of the various Tyr residues to the fluorescence and their respective anisotropy must vary upon aggregation. The other interesting aspect of these plots (compared to *e.g.* Figure 61-left) is the presence of a diagonal trend of decreasing anisotropy that is clearly shown by the *iso-aniso* contours in the plots, and which indicates that more than one Tyr participates in the fluorescence of insulin in each associated state.

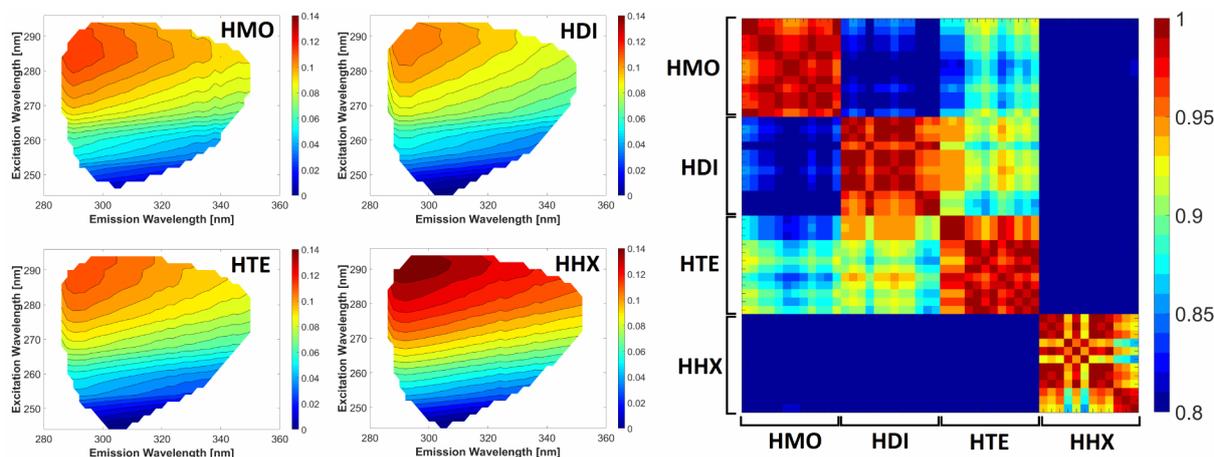
The idea is to use these *aniso*-MDF spectra to produce a method that can be easily used to assess protein aggregation state. This tool uses the similarity index (SimI) and the data region showing the most variability between *aniso*-MDF spectra which was obtained by calculating the difference spectra between the *aniso*-MDF plot of HMO and the other insulin forms (see Figure A- 48 for more details). The best region was however different depending if either HDI, HTE, or HHX were subtracted from HMO. A compromise was thus necessary, and we found adequate to select the spectral region spanning  $\lambda_{\text{ex}} = 268 - 284 \text{ nm}$  and  $\lambda_{\text{em}} = 294 - 332 \text{ nm}$ .



**Figure 70: SimI calculated using the *aniso*-MDF spectra of each samples (2 mg/mL) for a selected spectral region ( $\lambda_{\text{ex}} = 268 - 284 \text{ nm}$  and  $\lambda_{\text{em}} = 294 - 332 \text{ nm}$ ). Each sample was successively taken as the reference during the calculation, explaining the square obtained.**

The results obtained with 2 mg/mL samples are however not striking (Figure 70). The issue is that the *aniso*-MDF spectra are highly sensitive to noise and residuals from the Rayleigh scatter correction. This was not obvious in Figure 69 as the data shown were averaged over twelve measurements. Denoising the HH and HV spectra prior to calculation of the G factor, helps to remove part of the noise, however residual scatter persists especially in the VV or  $\text{EEM}_{\parallel}$  dataset. We therefore denoised both  $\text{EEM}_{\parallel}$  and  $\text{EEM}_{\perp}$  using the same approach as

for the G factor (*e.g.* 1 component PARAFAC model with non-negativity). To remove purely the noise/residual scatter and avoid the alteration of the spectra, it was crucial to do the denoising on each insulin form separately as the small spectral shift seen between them would have otherwise been eliminated. The discrimination between species is much improved as shown in the figure below.



**Figure 71: (Left) Average *aniso*-MDF spectra obtained for each insulin form at 2 mg/mL. (Right) Corresponding SimI calculated on the selected region.**

Therefore, the method seems to work well and enables discrimination of the various insulin forms without any ambiguity. The SimI values within each insulin form are also very high ( $> 0.95$  with few exceptions) leaving little doubt about the high similarity between the measurements once the noise and residual scatter are removed. The method was reproduced on the low concentration samples and the results obtained before denoising (Figure A- 49) are much improved when the denoising approach is applied (Figure A- 50). The samples can also be separated which is good considering the low SNR in these measurements. It is interesting to note that the iso-aniso contours obtained at 0.2 mg/mL are slightly different from the high concentration ones. They preserve the iso-aniso diagonal trend as well as the changes upon aggregation, but their anisotropy values are slightly higher. This is normal and explained by the presence of IFE at 2 mg/mL that depolarizes the fluorescence emission.

The method seems thus to be viable and robust as it worked at 2 mg/mL (without the need for correction) and at 0.2 mg/mL. Also because the denoised *aniso*-MDF spectra appeared as a pretty clear diagnostic of insulin aggregation state, we could consider using them for classification models like PLS-DA. Before considering this work for a later publication, of course the noise and low SNR issue will have to be addressed first by collecting better quality data (*e.g.* through longer integration times or by averaging multiple measurements). Ultimately

though, this will result in a very simple and robust tool to carry out efficient predictions on the association state of insulin.

To conclude this part, we can say that a careful analysis of the ARMES measurement has brought a lot of useful information to characterize insulin. Many of the tools and analysis discussed are new. Some could be used as analytical methods on their own, but the strength of the ARMES measurement is to integrate all these aspects into one rapid and non-invasive measurement. The advantage of having different spectroscopic effects in a single measurement is of course being able to use them to cross-validate what is happening in the solution. The only important piece that is missing is robust models based on the fluorescence part of the measurement.

### **5.10 Chemometric analysis of ARMES data (0.2 mg/mL): *towards resolving the complex Tyr overlap in insulin fluorescence***

In order to build robust models and try to resolve the fluorescence of insulin, it is preferable to avoid the use of methods like IFE corrections which could have adverse effects on the ARMES spectral shape (see next section for more details). Therefore, all the models presented here were developed on the low concentration samples, which is much safer at least to develop what we could call “academic” models. The high concentration data was kept for the next section where we want to show that models, perhaps destined for industrial applications, can also be done.

The 12 replicate measurements collected in each insulin associated state were split in two series of 6 replicate measurements (*i.e.* Series 1: HMO R1-6, HDI R1-6, THE R1-6, HHX R1-6. Series 2: HMO R7-12 and so on...). This was done so that a meaningful split-half analysis could be done using either the half as a subset or using a random selection of samples. All the models were otherwise built using the parameters defined in section 5.5, and the data was normalized to one for the various reasons mentioned previously.

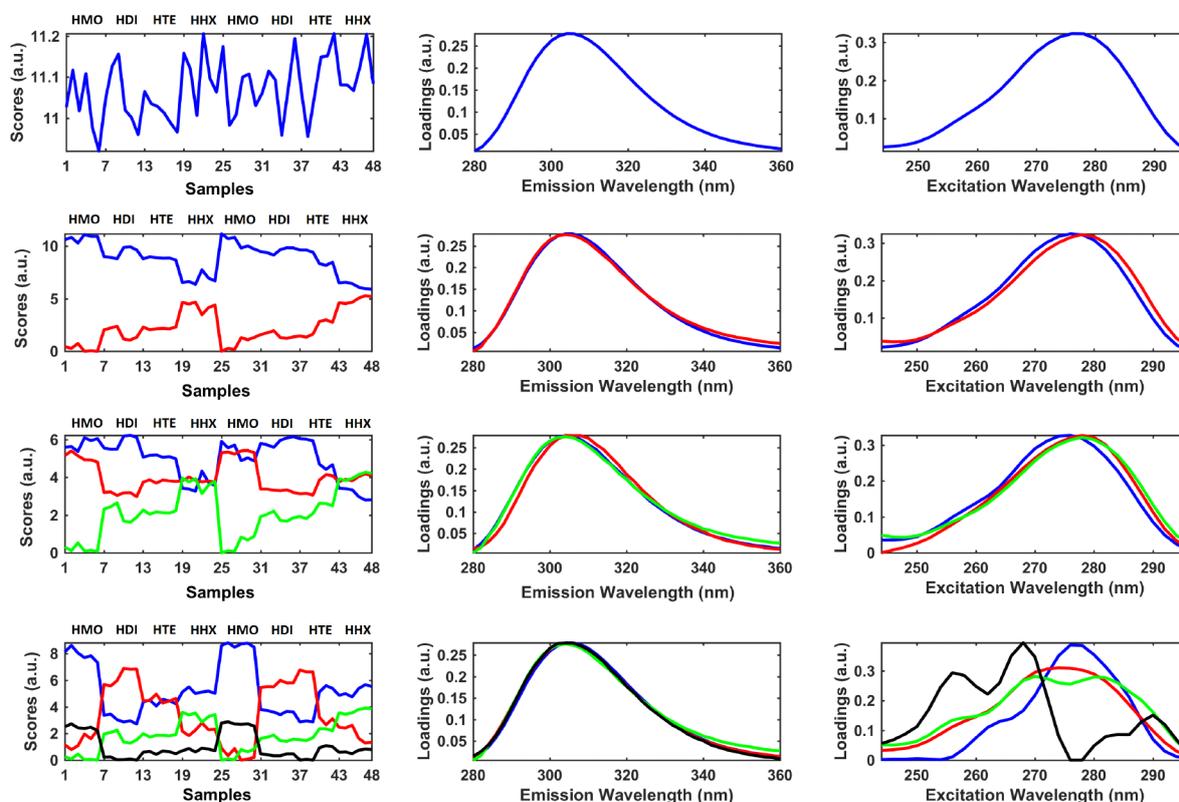
#### **5.10.1 EEM<sub>||</sub>**

One crucial point to address in the modelling of the insulin data is to find how many components we believe can be extracted from PARAFAC. As we are dealing with non-trilinear data, we already know that usual tools like CC will not be reliable. Considering that there are 4 Tyr residues in insulin and that it is very unlikely that Phe are resolved, we know however

that four is the maximum theoretical number of components to extract. Series of models were built to fit the data successively with one to four components. The scores and loadings are shown Figure 72 while some of their important characteristics are summarized in Table 25.

**Table 25: This table sums up the CC, total variance captured, and results from split-half analysis for the one to four component PARAFAC models, built using the normalized insulin EEM<sub>||</sub> data at 0.2 mg/mL.**

	1 Comp.	2 Comp.	3 Comp.	4 Comp.
Variance explained (%)	99.95	99.97	99.98	99.98
CORCONDIA (%)	100	-69.26	-3391.89	-459772.35
Split-half half (%)	100	99.8	93.3	51.6
Split-half random (%)	100	99.8	90.8	44.2



**Figure 72: Scores (left) and loadings (middle + right) of the PARAFAC models built using the pre-treated and normalised 0.2 mg/mL EEM<sub>||</sub> insulin data. The models were fitted with an increasing number of components, starting from one (top) and going up to four components (bottom). Comp.1/2/3/4 are respectively in blue/red/green/black.**

Looking at these models but also based on common sense, several arguments tip the scale in favour of three components:

- First, by working with the associated states of insulin that engage respectively TyrB16/TyrB26 in dimerization and then Tyr A14 in hexamerization, only two major

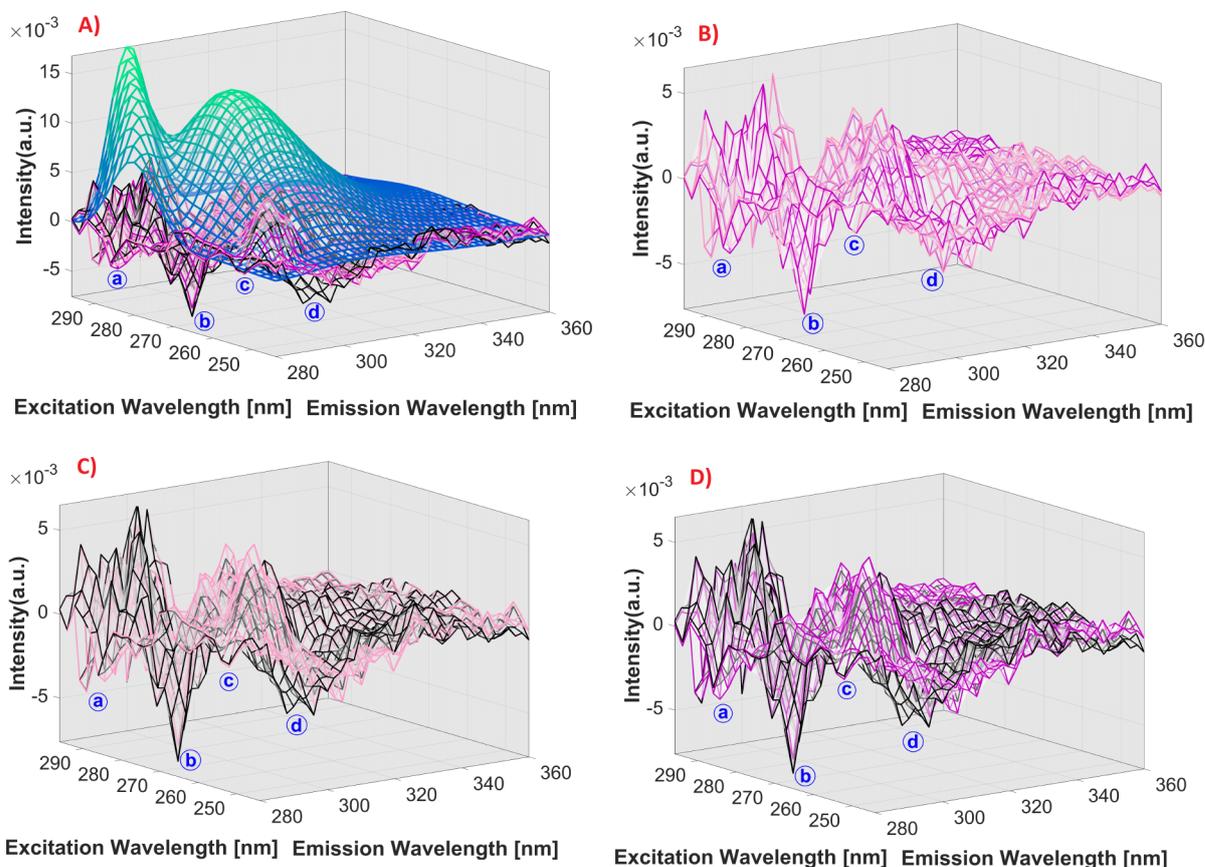
changes in structure are produced engaging Tyr residues. Added to TyrA19 that is located in the hydrophobic core of each monomer, there should thus be 3 contributions that should emerge from the models. Also due to the antiparallel arrangement of monomers to form dimers and the creation of aromatic cages between B16<sub>monomer1</sub>-B26<sub>monomer2</sub> and vice versa, it is logical that these two components will only form one factor.

- Second compared to free L-Tyr in solution (Figure A- 36) or the modelling of the monomer samples (Figure 58), the loadings extracted here with three components correspond to Tyr residues and are strongly overlapped as we expected. Going up to four components however leads to the loss in the correct loadings shape, while going down to two components lead to scores that are the mirror image of each other (*i.e.* perfectly collinear). The three components solution seems therefore to be the most correct. This impression is also reinforced by the fact that the exact same observations were found for the EEM<sub>⊥</sub> (Figure A- 53).
- Then if we look at the residuals of each model, it also seems that the correct number of components is three. Because the residuals represent for each sample the portion of the signal that was left behind during the modelling, a careful analysis must be performed to know if any important part of the analyte signal was omitted. The average residuals for the one to three components PARAFAC models are shown in Figure 73 below. They are overlapped in A with the averaged spectra of all EEM<sub>∥</sub> spectra (not Rayleigh corrected), that had to be scaled as the residuals with an intensity  $< 5 \cdot 10^{-3}$  represent less than 1/200 of the fluorescence signal (intensity = 1). For clarity purposes, the four components model was not added as this possibility was already ruled out due to the wrong excitation loadings obtained.

The first comment which can be made is that for all models, the residuals have negligible intensities compared to the fluorescence signal, and what is mostly left behind seems to be either noise or some residual signal that is present along the 1<sup>st</sup> order Rayleigh band (a) and the juxtaposed Raman band (b). This is good and is expected from a model that manage to filter out the bad data elements to only preserve the pure fluorescence information.

A second point is that even if the residuals from all models are highly similar, some slight difference can be observed. The one (purple) and two (pink) component models are the ones showing the most similar residuals (Figure 73, B). This agrees with the perfect scores collinearity seen in the two component model (Figure 72) and is proving that going

from one to two components is not explaining more the data; it only splits the component extracted from the one component model into two.



**Figure 73:** Average residuals calculated across the 48 samples for the one to three components PARAFAC models, overlapped in A with the scaled-averaged spectra of all EEM<sub>||</sub> (green, not Rayleigh corrected), and shown by pairs in B/C/D for the one/two, one/three and two/three component models respectively. The residuals for the one and two component models are shown in purple and pink respectively, while the residuals for the three component model are shown in black.

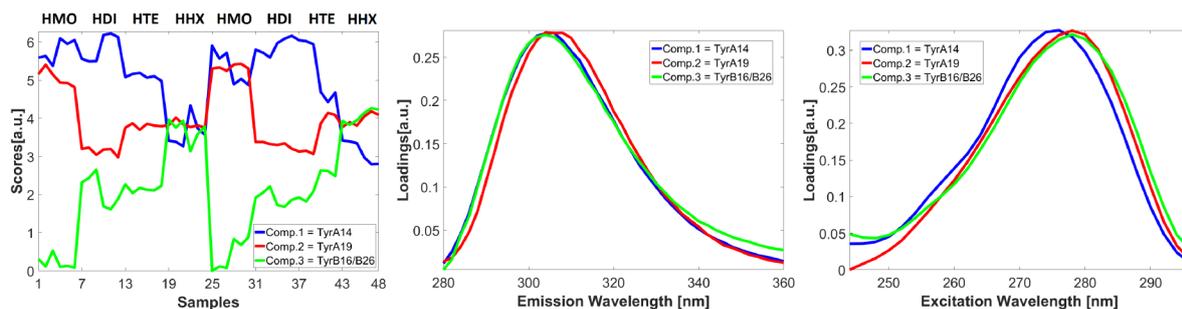
Adding a third component however induces some slight changes in the average residuals that are probably beneficial to the retrieval of sound components. In particular looking at (a) and (b), it appears that a higher proportion of the noise left along the 1<sup>st</sup> order Rayleigh band and the Raman band post-correction, is removed (Figure 73, C/D). A very small portion of the fluorescence signal is maybe lost in (c), but as one can see, it is lowered when three components is used as opposed to one and two. Finally, in (d), it is interesting to see that only the three-component model manages to filter out this portion of the signal that corresponds to the beginning of the <sup>1</sup>L<sub>a</sub> band, and that we do not particularly want to model as it might include more source of variances that can harm the right decomposition

of the overlapped Tyr. Overall, the residual analysis seems thus to confirm that the three-component model gives the best PARAFAC answer.

- The split-half analysis also confirms that point, with a percentage of similarity between subset loadings that stays above 90% whether a half or random subset is used. It is only with four components that this percentage drops, with the same behaviour seen for the EEM<sub>⊥</sub> models (Table A- 20).
- The values obtained for the CC are also informative. As opposed to well behaved samples where CC is a clear diagnostic about the appropriateness of the model, the reliability of such parameter is lost when difficult cases are studied. For example, samples that covary or that are highly correlated in the spectral modes will result in a model with a negative CC, which yet does not mean that the components resolved are not correct. For such samples, Murphy *et al.*<sup>289</sup> warned in particular about the possibility of serious underfitting if one relies solely on the CC results. Instead when appropriate loadings profiles and good split-half results are obtained, a sudden drop in CC is almost a good sign as components start to span non-trilinear variation in the data.

Chen *et al.*<sup>508,144</sup> for example obtained negative CC during the PARAFAC modelling of FRET systems. Apart from advising the use of PARAFAC over PARALIND in the absence of *a priori* knowledge, they proposed if all the other model parameters are correct, to use the retrieval of negative CC as a tool to indicate the presence of special linear dependencies in the samples. It is particularly interesting to note in their case that PARALIND had difficulties to equalize the perfect fit that was obtained by PARAFAC, for which the only problem was the large and negative CC (-55200). This corresponded however to a model with linear dependencies in three modes<sup>508</sup>. For a model having “only” linear dependencies in both spectral modes<sup>144</sup>, a value of -971 was obtained for CC which is in the same order of magnitude as the value obtained here with three components (-3391.9). This three-component solution seems thus to be the most correct, considering that the data present multi-modes dependency.

For all these reasons, we believe that the model can extract three real components. Below are presented the scores/ loadings (Figure 74) and some important characteristics (Table 26) of the best model obtained out of 10 repetitions, using the best fit model as initialization method.



**Figure 74: Scores (left) and loadings (middle + right) of the three component PARAFAC model built using the 0.2 mg/mL EEM<sub>||</sub> insulin data.**

**Table 26: Summary of the fit captured by each component for the three component PARAFAC model built using the normalized insulin EEM<sub>||</sub> data. The CC, total variance captured by the model and results from split-half analysis are also indicated. Each component is also assigned a specific Tyr residue.**

	<b>EEM<sub>  </sub> 3 components</b>	<b>Residue assignment</b>
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	276/304	TyrA14 ( <i>Y14A</i> )
<b>Fit model (%)</b>	53.47	
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	278/306	TyrA19 ( <i>Y19A</i> )
<b>Fit model (%)</b>	34.05	
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	278/304	TyrB16 + TyrB26 ( <i>Y16B</i> + <i>Y26B</i> )
<b>Fit model (%)</b>	12.48	
<b>Variance explained (%)</b>	99.97	/
<b>CORCONDIA (%)</b>	-3391.33	/
<b>Split-half half (%)</b>	93.3%	/
<b>Split-half random (%)</b>	90.8%	/

Looking at the scores and loadings in Figure 74 and as expected from Tyr photo-physics (in the absence of ground/excited state proton transfer), the spectra resolved in emission and excitation are highly overlapped. The solution obtained is thus very different from what was obtained using the monomer data alone (Figure 58) and suggests that the model succeeded in breaking the linear dependencies. This is characterised by a negative CC that we said in the same order of magnitude as other FRET studies that showed collinearity in two modes<sup>144</sup>. Also from the components that are resolved, we can see that Comp.3 shares the same excitation loading as Comp.2 but the same emission loading as Comp.1 which is characteristic of a resolved FRET species<sup>144,299-300</sup>.

Concerning the spectral shifts, Figure 74 proves that slightly more shift is seen along the excitation than the emission, which agrees with what we said in sections 5.4 or 5.9.2 but also

with some early PARAFAC models on NAYA fluorescence in presence of various solvents and ligands<sup>464</sup>. We already explained in section 5.7 the meaning of spectral shifts in Tyr excitation spectra in terms of change in its microenvironment<sup>33,464</sup>. However we did not mention yet the meaning of slight shifts in Tyr fluorescence emission. Even if the latter is mostly insensitive to its environment<sup>461</sup>, the same authors that used NAYA to model Tyr behaviour in proteins, proved that a slight red shift can appear in Tyr emission if the solvent has a low dielectric constant and high hydrogen bond acceptor strength<sup>33,464</sup>. Tyr emission can thus be slightly red-shifted in strongly hydrophobic environments (as observed for example by Poveda *et al.*<sup>497</sup>) and if Tyr is hydrogen-bonded to carboxylate groups<sup>464</sup> (or other good proton acceptors) which is often the case when Tyr is buried in the protein's hydrophobic core and close to other residues<sup>522</sup>.

Using this information and some of the details developed in section 5.7, we tried then to assign each component to a specific Tyr residue of insulin. Looking at the loadings in Figure 74, the model potentially extracted a hydrophobic Tyr (red shifted, Comp.2), a solvent exposed Tyr (blue shifted, Comp.1) and another Tyr component (Comp.3) that has the same characteristics as the FRET species. We believe that Comp.2 is TyrA19, Comp.1 probably TyrA14, and Comp.3 probably TyrB26 + TyrB16. The reasons for this assignment are given below.

First, it is important to mention that the assignment is a difficult and uncertain task considering that we could not find any other literature study to confirm our results. The other point that makes it difficult is that according to the inter-Tyr distances in insulin (Figure 49<sup>uu</sup>) and referring to the 9-16 Å range usually observed for Förster distance (Table 19), every Tyr residue are in close enough for FRET to occur. One book chapter, written by Longworth<sup>522</sup>, advanced some interesting hypothesis about the intrinsic fluorescence of insulin. Based on the calculation of the natural lifetime<sup>vv</sup> of insulin and tyrosine, the author concluded that only two Tyr residues were contributing equally to the fluorescence of insulin. This information is interesting as it matches the fact that (only) Comp.1 & Comp.2 have non-zero and equal scores across the monomer samples (see HMO, Figure 74-left).

There are several reasons explaining why Comp.2 is probably TyrA19. As shown in Figure 49 and explained in section 5.2.1.4, TyrA19 is the only residue part of the hydrophobic

---

<sup>uu</sup> For clarity, it is good to remind that in Figure 9, the abbreviation Tyr is replaced by Y, Phe by F, and Cys by C.

<sup>vv</sup> The “natural lifetime” is identical to  $\tau_r$ , the radiative lifetime, introduced in section 1.2.3. The radiative lifetime can be calculated from the ratio between the measured lifetime and the quantum yield of the molecule.

core of the monomer. TyrA19 is thus in a region of low dielectric constant but is also surrounded by neighbouring residues, meaning that hydrogen-bonds to a carboxyl or amide group from either the peptide backbone or neighbouring residues can easily be established. In fact, it was suggested that TyrA19 was hydrogen bonded to the carbonyl group of GlyA1<sup>329,523</sup> as well as being in van der Waals contact with IleA2 (see section 5.2.1.4). For that reason, TyrA19 must be Comp.2 as both excitation and emission maxima are slightly red-shifted (Figure 74). In terms of intensity, the literature is quite contradictory about the fluorescence of Tyr residues buried in protein hydrophobic cores. For some authors, these residues should be totally quenched due to the formation of complexes with other residues or peptide backbone<sup>524</sup>. For others, buried Tyr should be the main source of fluorescence as observed in the case of Histone H1<sup>525</sup>. In fact, it must surely be different for each protein case depending on the position of the fluorophore in respect to potential quencher elements.

We believe for example that for TyrA19, several different photo-physical processes control its emission. Due to its proximity with the disulfide bridge CysA6-CysA11 (especially in the UK molecule 2 conformation, Figure 49), we think that TyrA19 might in fact be partially quenched<sup>473</sup>. However we also think that because this residue shows a slightly red-shifted excitation, it must act as energy trap for other Tyr residues. Typically TyrB26 (Figure 49), by being partially exposed in the monomeric state and separated by only  $\sim 9$  Å from TyrA19, probably transfers most of its energy to the latter<sup>522</sup>. The other Tyr residues (TyrB16 and TyrA14) are on the other hand a bit further out ( $\sim 16$  and  $14$  Å respectively) and will have thus a lower FRET rate. It was suggested by Longworth that TyrB16 was also transferring its energy to TyrA19<sup>522</sup>, unfortunately we cannot be sure of it.

What is certain however, is that upon dimerization, the environments of TyrB16 and TyrB26 change drastically as both residues are buried in the hydrophobic region between the monomers. Such a change must have consequences on the fluorescence emission of both residues, in particular one might expect a decrease in their singlet energy level as both residues get engaged in a variety of interaction with facing residues (see section 5.2.2 for more details). The intermolecular distances between Tyr residues from anti-parallel monomers become also much shorter than *e.g.* the distance separating TyrA19 to TyrB26 in each monomer sub-unit (Figure 49). For these reasons, it is highly probable that upon dimerization, the energy that was transmitted to TyrA19 while in the monomeric form ceases to flow as far in the molecule, and instead stays located in the monomer/monomer interface. In that case, we should see upon dimerization a decrease in Comp.2 scores (TyrA19) and a sudden rise in the FRET species *i.e.*

Comp.3 (TyrB16 & TyrB26), which is exactly what was obtained by the model (see Figure 74, left).

So far, the model is thus in good agreement, with a pretty sensible explanation of what could happen photo-physically upon insulin oligomerization. To complete this assignment, we still have to relate Comp.1 (that has not been assigned yet) to a specific Tyr residue. The only one that has not been used yet is TyrA14.

We know that Comp.1 is related to a Tyr that is exposed to a polar environment, due to excitation and emission maxima that are slightly blue shifted. In that regard, TyrA14 is indeed a perfect match as this residue is exposed to the solvent in the monomeric and dimeric state (see section 5.2.1.4) and stays outside of the molecule even upon hexamerization<sup>327,329</sup>. We thought that TyrB16, by being relatively distant from TyrA19 and being partially exposed, could also be a potential good candidate. However, TyrB16 is already engaged in dimerization and one can see that the scores of Comp.1 are only dropping upon the formation of hexamers. Considering that TyrA14 is the only Tyr residue to be directly engaged in the formation of hexamers (see section 5.2.2), associating Comp.1 to TyrA14 would thus appear more natural. The difficulty though is to explain why Comp.1 scores drop while Comp.3 scores increase. The fact that both scores are varying simultaneously in opposite directions (and that Comp.3 is the FRET species) would suggest that another FRET process exists. This process seems to be disturbed by the formation of hexamers, and we propose two scenarios that could explain it:

- As seen in section 5.2.2, TyrA14 forms with PheB1 and their facing dimer homologues an aromatic cage exposed on the surface of the hexamer. Both TyrA14 are in close contact<sup>327,329</sup>, and might exchange energy via FRET. One of these residues act thus as the donor and is quenched (Comp.1) while the acceptor (part of Comp.3) is enhanced.
- Another possibility is to consider TyrB16 as acceptor from TyrA14 energy. Indeed from all the remaining Tyr residues present in insulin except TyrA14, TyrB16 is the only one to have its environment perturbed not only upon dimerization, but also further upon hexamerization<sup>329</sup>. Its phenyl ring in particular, by being more buried, is slightly less accessible in the hexamer than in all other forms<sup>329</sup>. We could thus presume that the formation of hexamers strengthens the hydrophobic behaviour of the region located between monomers, which contains TyrB16 & TyrB26 and end up trapping TyrA14 energy. The distance between TyrA14 and B26 is a bit high  $\sim 18 \text{ \AA}$  but this is within matches the upper limits of Tyr-Tyr Förster distances.

Both scenarios are plausible and could in fact form a single FRET mechanism if the transfer of energy is total in the aromatic cage (Tyr-Tyr < 6 Å) and that this energy is only partly transmitted (longer distance ~ 18 Å) to TyrB16.

Nevertheless, assigning Comp.1 to TyrA14 seems appropriate and another interesting element, present in the excitation loadings, further supports this view. Looking at Figure 74 (right), Comp.1 shows a clear shoulder at 258 nm that is more pronounced than for the other resolved factors. This shoulder is a consequence of Phe to Tyr FRET<sup>52,458,526</sup> and means that the Tyr residue representing Comp.1 must be in close contact to a Phe residue to receive its energy. As shown in Figure 49, TyrA14 is separated by only ~ 5.7 Å from PheB1 and would thus match this requirement. The issue is that other tyrosine residues such as TyrB16 that is only 7.3 Å away from PheB24 or TyrA19 (in UK molecule 2 conformation) 5.8 Å from PheB25, are also close to Phe residues. The difference however is that TyrA14 is always in a 1:1 stoichiometry with a Phe residue, even upon formation of the aromatic cage in the hexamer, while TyrB16 for example form a cage in the dimer with two Tyr residues for a single Phe residue<sup>ww</sup>. The FRET contribution of Phe to Tyr should therefore be stronger for TyrA14, which is another argument for assigning Comp.1 to TyrA14.

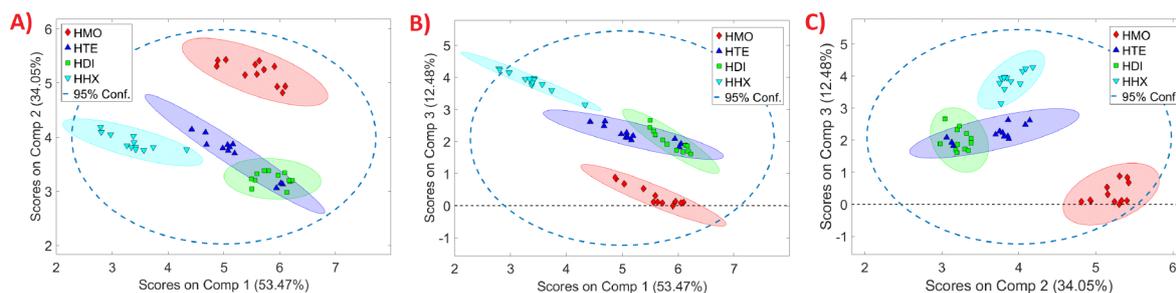
Overall, the PARAFAC model developed and presented in Figure 74 is thus very interesting. Even if care must be taken due to the poor SNR/data resolution and the small spectral changes observed, it seems that we were able to separate buried and exposed Tyr residues in insulin. The model also managed to uncouple the Phe contributions to each specific Tyr residues, which to the best of our knowledge was never obtained before using steady-state fluorescence. This is different but also more powerful than observing the Phe-to-Tyr FRET in proteins containing a single Tyr for multiple Phe residues<sup>458,526</sup>. The model was also able to resolve the FRET species (Comp.3), thus proving the presence of internal FRET in insulin, but also showing that multiple FRET mechanisms might exist simultaneously. In fact, it seems that each important step in insulin natural oligomerization (*i.e.* dimerization or hexamerization) that result in a drastic change in Tyr microenvironment, led to a different repartition of the energy across the Tyr residues. This change in energy distribution is manifested by changes in the intensity or scores of all residues including the FRET species. What is particularly interesting is that these changes occur upon dimer and hexamer formation, supporting the conclusions

---

<sup>ww</sup> TyrA19 does not form aromatic cages. However it is only in UK molecule 2 conformation that TyrA19 is close to PheB25, but it is also in this conformation that TyrA19 is close to the disulfide bond (and probably more quenched). It is thus logical to expect a lesser contribution of Phe to TyrA19 than for TyrA14.

made in section 5.9.3 and the fact that the *aniso*-MDF spectra mostly changed for these two species.

By breaking the linear dependencies, we also developed a model that can clearly discriminate the insulin forms as shown in Figure 75. This is a huge improvement compared to the models developed using the 0.2 mg/mL absorbance data (see section 5.7) where we could not separate the species. Because this model was based on data normalised to one, the degree of variation seen in the scatter scores plot is small. The fact that the model still manages to separate the species without having to multiply the scores by the normalization factors, is thus even more interesting as it proves that the separation can be done purely based on spectral shape changes (more robust). As one can see, the 12 replicate samples of HMO and HHX are clearly separated from the rest of the samples. However HDI and HTE are overlapping, due to the replicate samples HTE R7-R9. This is thus quite similar to what was seen in section 5.7 with the HTE R10-R12 samples at 2 mg/mL and confirm the ambiguous and unstable nature of the tetrameric state, which can easily reverse to the dimeric state.



**Figure 75: (A/B/C) Scatter scores plot obtained for the three component PARAFAC model, built using the normalised EEM<sub>||</sub> insulin oligomers data and containing 12 independant replicate samples (four series of three) per insulin form.**

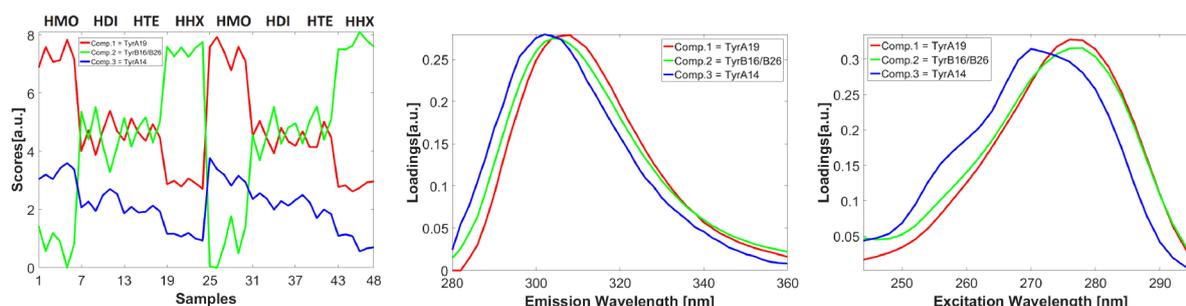
To go further, classification models can be built using the PARAFAC scores as shown previously by Lenhardt *et al.*<sup>527</sup>. The authors used the scores to build a PLS-DA model that could for example classify and predict with 100% success rate if the honey was real or fake. Another model also enabled to predict with very high success rate the honey's botanical origin. Such models were not developed yet on the insulin data, because the issue of the data quality (low SNR/resolution) should be addressed first. But based on the PARAFAC results, we can presume that successful models can be built to predict if the insulin sample analysed is either mainly monomeric, dimeric, or hexameric.

Finally, to conclude on the EEM<sub>||</sub> data, it was crucial to ensure that the model developed was robust and reproducible. As shown in Figure A- 51 and Table A- 18, it appeared that upon

10 repetitions, the model was highly reproducible as all solutions almost perfectly overlapped. To complete the analysis, we also compared the answer obtained using the best fit model (BFM) as initialization method with other initialization methods such as Alternating TriLinear Decomposition (ATLD), Singular Value Decomposition (SVD) and Random Orthogonalized values (ROV). It appeared again that all the answers obtained overlapped almost perfectly as shown in Figure A- 52 and Table A- 19, signifying that the model is highly robust. This result also means that the global minima is probably achieved, because the ALS algorithm converged towards the same solution even when it started from several different starting points (ROV).

### 5.10.2 EEM<sub>⊥</sub>

The same methodology was applied on the pre-treated and normalized EEM<sub>⊥</sub> insulin data (0.2 mg/mL). Successive PARAFAC models were built using one to four components (Figure A- 53, Table A- 20) and for the same reasons, a three component model was chosen. The scores/loadings and some important model characteristics are shown Figure 75 and Table 27.



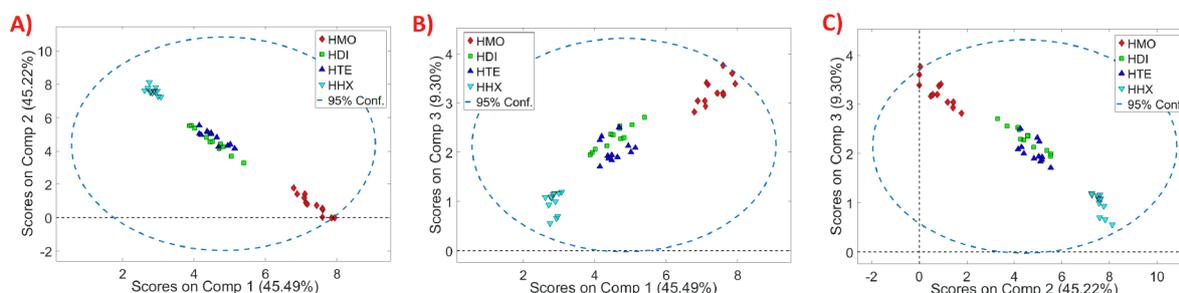
**Figure 76: Scores (left) and loadings (middle + right) of the three component PARAFAC model built using the 0.2 mg/mL EEM<sub>⊥</sub> insulin data.**

**Table 27: Summary of the fit captured by each component for the three component PARAFAC model built using the normalized insulin EEM<sub>⊥</sub> data. The CC, total variance captured by the model and results from split-half analysis are also indicated. Each component is also associated to the Tyr residue it represents.**

	EEM <sub>⊥</sub> 3 components	Residue assignment
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	276/308	TyrA19 ( <i>Y19A</i> )
<b>Fit model (%)</b>	45.49	
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	278/304	TyrB16 + TyrB26 ( <i>Y16B</i> + <i>Y26B</i> )
<b>Fit model (%)</b>	45.22	
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	270/302	TyrA14 ( <i>Y14A</i> )
<b>Fit model (%)</b>	9.30	
<b>Variance explained (%)</b>	99.98	/
<b>CORCONDIA (%)</b>	-12958.74	/
<b>Split-half half (%)</b>	99.0%	/
<b>Split-half random (%)</b>	99.0%	/

As one can see, the characteristics of this model are different than the model developed for the EEM<sub>||</sub> data. The Comp.3 that used to have the spectral characteristics of resolved FRET species, is different this time. We tried to preserve the same colour code as the one used for the EEM<sub>||</sub> model. However, because the components are slightly different, it was difficult to be sure of their identity. Due to the shift on the excitation/emission, we assumed that TyrA19 was Comp.1 (instead of Comp.2 in the EEM<sub>||</sub> model) while TyrB16/B26 was Comp.2 (scores are almost at zero for monomer samples) and TyrA14 was Comp.3.

Clearly it seems that the model failed in correctly resolving the Tyr residues. First of all the FRET species is lost, and if we look at the scatter scores plot (Figure 77), the components seem to have a high degree of collinearity, especially Comp.1 and Comp.2 that explain > 90% of the data.



**Figure 77: (A/B/C) Scatter scores plot obtained for the three component PARAFAC model, built using the normalised EEM<sub>⊥</sub> insulin oligomers data and containing 12 independent replicate samples (four series of three) per insulin form.**

Different reasons can explain this failure. As seen in section 5.8, it is quite possible that the PARAFAC model, which tries to explain the data in a trilinear way, is once again negatively influenced by the presence of a strong shoulder in the excitation spectra due to Phe-to-Tyr FRET. This shoulder is stronger in EEM<sub>⊥</sub> than in EEM<sub>||</sub> because FRET depolarizes the fluorescence emission. As seen in the EEM<sub>||</sub> model, there is very little variation in terms of spectra between exposed and buried Tyr. Any other spectral feature that generate more variation will thus be picked up by the model and used to make a component. This has the consequence of ruining the resolution of Tyr residues, and it is probably what happened here. It would explain in particular why TyrA14, that was Comp.1 before, is now Comp.3 and represents only 9% of the model fit. It would also explain why this component is more blue shifted with an excitation shape more distorted, but also why the FRET species disappeared.

A second reason which make the EEM<sub>||</sub> case more successful, might also simply be due to the fact that even if Tyr spectral shape is not very affected by changing environments, the

intensity and anisotropies are on the other hand very sensitive. During the natural oligomerization of insulin, it is thus quite reasonable to think that the anisotropies of each Tyr residue strongly change, which was in fact already suggested and supported by the dissimilar *aniso*-MDF spectra that characterised each insulin form (section 5.9.3). This generates changes in the polarized emission and might be the key to resolving Tyr emission, giving a reasonable explanation about why the polarized emission (EEM<sub>||</sub>) was more successful.

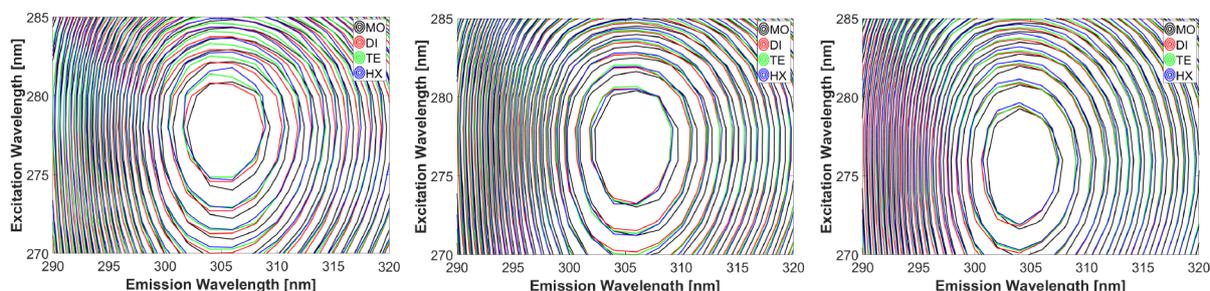
In any case, even if the EEM<sub>⊥</sub> model was reproducible (Figure A- 54, Table A- 21), the components that were extracted are not correct and their collinearity (responsible for the very negative CC) would tend to suggest over-modelling. Using the components resolved in the EEM<sub>||</sub> and EEM<sub>⊥</sub> models, there is thus currently no hope to extract *aniso*-MDF spectral shapes that are coherent with the resolution of pure Tyr contributions (data not shown). With the PARAFAC exploration of the EEM<sub>||</sub> dataset and the resolution of the FRET species, we however now have the advantage to know that essentially two contributions can be extracted in the spectral modes while three contributions are present in the concentration mode. This could thus be modelled by an appropriate PARALIND model, which could also be applied to the EEM<sub>⊥</sub> data. In that regard, there is thus some hope to extract meaningful components in EEM<sub>⊥</sub> and thus meaningful *aniso*-MDF spectra for the components. However as mentioned previously, the data quality issue will have to be addressed first which we plan to do in the near future.

### **5.11 Models using 2 mg/mL samples: non-IFE corrected data for separating highly concentrated solutions**

The 2 mg/mL insulin samples, characterised by absorbance values above 0.8 (Figure 56), suffer from severe IFE. Using an absorbance-based approach (ABA, section 1.3.3), we tried to correct the ARMES data. However as shown in Figure A- 55, the correction by ABA lead to a spectral shape that is distorted with the appearance of a shoulder at  $\lambda_{em} \sim 284$  nm. The spectral shape obtained seems in fact to be overall too influenced by the absorbance spectral shape, that for example also present a shoulder at 284 nm. According to Equation 20, the correction factor by ABA were larger for shorter wavelengths, explaining why it is mostly in this wavelength range that the spectral shape is the most distorted. ABA thus failed to correct the data and cannot be trusted. Instead, we want to show that it is possible to build PARAFAC models that can separate insulin species without having to correct for IFE. The first reason to do so is because the insulin data must be normalised to one (to abolish the pH buffer influence),

meaning that the recovery of true intensities by IFE correction is for us meaningless. Instead we are more interested in spectral shape changes, and we saw in Figure A- 55 that the closer shape to data not affected by IFE (0.2 mg/mL) were obtained without IFE correction. The second reason is because there must be cases in industry where insulin is either used or produced in bioreactors at concentrations even higher than 2 mg/mL<sup>517</sup>, and where the IFE correction will not be possible. In this kind of cases, being able to discriminate between insulin associated forms is important and it is more interesting if it can be done in a straight forward manner, avoiding intermediate measurements and extra correction steps.

We thus looked at the fluorescence data and realised that with the increase in insulin size and even in the presence of IFE, the  $EEM_{\parallel}$  showed a larger red-shift than  $EEM_{\perp}$  or a normal EEM (Figure 78).



**Figure 78: Overlap of the normalized-averaged  $EEM_{\parallel}$  (left),  $EEM_{\perp}$  (middle), and normal EEM (right) fluorescence signal obtained for all samples in all buffers at 2 mg/mL.**

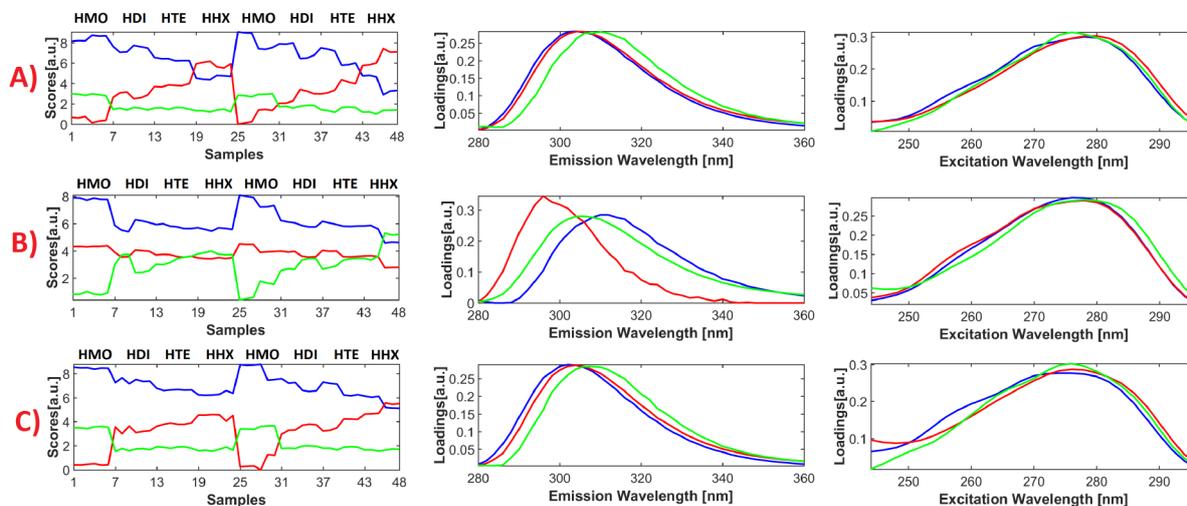
The spectral variability, necessary for good PARAFAC models, seemed therefore to be enhanced in the  $EEM_{\parallel}$  data compared to the other forms of EEM measurement. A sensible explanation to this effect would be to consider once again the fact that Tyr intensity and anisotropies are very sensitive to changes in the microenvironment. During insulin oligomerization, we saw that the environment of specific Tyr residue is highly affected. Their anisotropy is thus also affected, which generate changes in the polarized emission, and thus more variability in  $EEM_{\parallel}$ <sup>xx</sup>.

In any case, this property seen in Figure 78 makes the comparison between PARAFAC models build using  $EEM_{\parallel}$ ,  $EEM_{\perp}$  and EEM IFE-contaminated data an interesting case. The models were built following the same rules as described in section 5.10.1. Once again, we built

---

<sup>xx</sup> It also may be possible that IFE are differently affecting the EEM dataset depending on the direction that is given to the light during the measurement. In particular, the phenomenon of photo-selection may influence how the light is reabsorbed by the fluorophore or other non-fluorescent absorber. This would thus generate an IFE effect that is different and maybe lower in  $EEM_{\parallel}$  as opposed to *e.g.*  $EEM_{\perp}$ , but this is just a hypothesis.

three-component models because this is the limit at which reasonable spectral shapes, CC (except for  $EEM_{\perp}$ ) and split half results are obtained. The details of the models are shown in Table 28 and Figure 79 below.



**Figure 79:** Scores (left) and loadings (middle + right) of the three component PARAFAC models built using the 2 mg/mL (A)  $EEM_{\parallel}$ , (B)  $EEM_{\perp}$  and (C) EEM normalised insulin data.

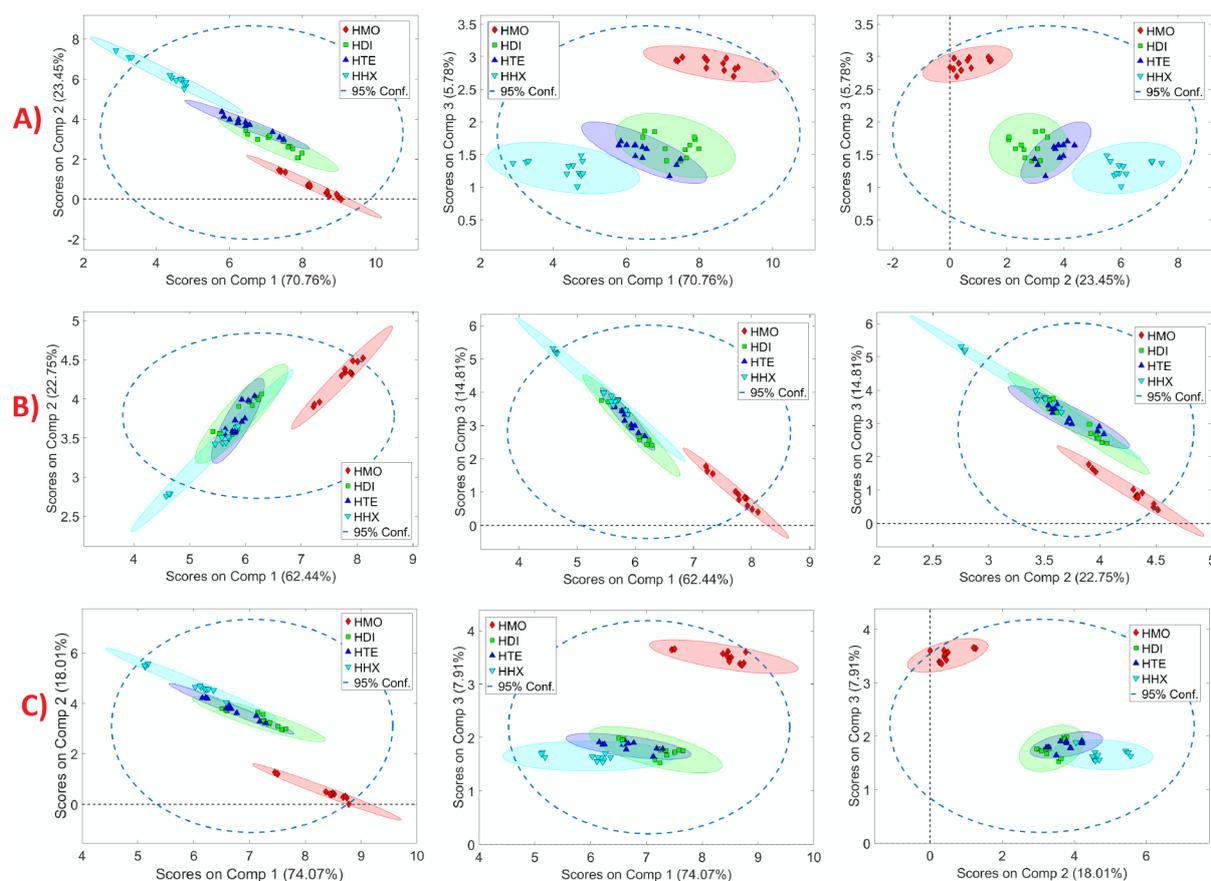
**Table 28:** Summary of the fit captured by each component for the three component PARAFAC models built using the 2 mg/mL  $EEM_{\parallel}$ ,  $EEM_{\perp}$  and EEM normalised insulin data. The CC, total variance captured by the model and results from split-half analysis are also indicated.

	$EEM_{\parallel}$ 3 comp.	$EEM_{\perp}$ 3 comp.	EEM 3 comp.
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	278/304	276/310	274/302
<b>Fit model (%)</b>	70.76	62.44	74.07
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	280/304	278/296	276/304
<b>Fit model (%)</b>	23.45	22.75	18.01
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	276/310	278/306	276/306
<b>Fit model (%)</b>	5.78	14.81	7.91
<b>Variance explained (%)</b>	99.99	99.99	99.99
<b>CORCONDIA (%)</b>	-1795.66	-39231.57	-1194.54
<b>Split-half (half, %)</b>	93.0	70.4	55.0
<b>Split-half (random, %)</b>	89.4	72.3	58.0

Due to the presence of IFE, there was no point here to try assign a component to a specific Tyr residue, even if there is a certain similarity with the 0.2 mg/mL models (at least for the  $EEM_{\parallel}$  and EEM models). If we look at Figure 79, the  $EEM_{\parallel}$  and EEM models seem to give a similar answer except that the  $EEM_{\parallel}$  model is also characterised by a more negative CC but

also very satisfying split half results. The model developed using the  $EEM_{\parallel}$  is thus more robust which make sense if more spectral variability is observe in the data.

Looking at the scatter scores plot (Figure 80), only the  $EEM_{\parallel}$  model was able to clearly separate the HMO samples from HHX and HDI/HTE. The components that have been extracted also do not have collinear scores, which is not the case for example with  $EEM_{\perp}$  (see Comp.1 and Comp.3). We might thus consider the  $EEM_{\parallel}$  data to not be over-modelled and the components extracted to be “real” (at least for a model contaminated by IFE).



**Figure 80: Scatter scores plot obtained for the three component PARAFAC models, built the 2 mg/mL (A)  $EEM_{\parallel}$ , (B)  $EEM_{\perp}$  and (C)  $EEM$  normalised insulin data. The data contained 12 independant replicate samples (four series of three) per insulin oligomer.**

The  $EEM_{\parallel}$  model was also reproducible (Figure A- 56, Table A- 22), even if more variability is seen compared to the models developed on the 0.2 mg/mL data that did not suffer from IFE.

## 5.12 Conclusions

We have seen in this chapter that ARMES, through different aspects of the same measurement, offered multiple ways to characterise either qualitatively or quantitatively the soluble non-covalent oligomeric states of human insulin.

This type of aggregate are known to not be suitable for conventional SEC<sup>18</sup> and are also difficult to characterise spectroscopically<sup>17</sup>, which indicates the need for a better analytical method. By splitting the ARMES measurement into pure scatter and fluorescence information, we have shown that separately the 1<sup>st</sup> order Rayleigh scatter band, the scatter depolarization, and the denoised *aniso*-MDF spectra were able to discriminate/identify the aggregation state of insulin in solution. The simple use of SimI with denoised *aniso*-MDF spectra, for example enabled us to discriminate the 12 replicate samples collected independently over several weeks and at low (0.2 mg/mL) and high (2 mg/mL) concentrations. The polarized 1<sup>st</sup> order Rayleigh scatter band on the other hand, proved to be more useful at high concentration, which is very significant considering that insulin is formulated at concentrations  $\geq 2$  mg/mL<sup>517</sup>. Of course, refinements can be done to make it suitable for lower concentrations, and in general the quality of the signal (higher resolution, less noise) can be improved. The strength of the method though was proven, as we have shown that ARMES offers a fast, non-invasive, robust, and inexpensive measurement, with multiple ways to probe what is going on in the sample.

The other major result of this chapter was to show that it was possible for the first time, to resolve the intrinsic Tyr fluorescence of insulin, despite the presence of multi-mode dependencies due to internal FRET. The crucial parameter that enabled to break these linear dependencies was the location of insulin's tyrosine residues in the regions involved in dimer and hexamer formation. The successful PARAFAC models suggested in particular that we could separate buried and exposed Tyr residues in insulin, as well as observing different Phe to Tyr contributions. The FRET species was also resolved, which proved that internal FRET occurred in insulin to a significant extent, but also to build robust models for discriminating the small soluble insulin aggregates. These results are thus very promising, even though they need to be confirmed using better quality data (higher SNR and spectral resolution). This is however not currently feasible with scanning-based spectrometers that we have available in-house.

Overall this chapter demonstrated the potential of ARMES for the non-invasive analysis of peptide solutions.

## 6. Conclusions

The aim of this thesis was to apply ARMES to insulin, in order to resolve its intrinsic emission and enable the discrimination of small soluble aggregates in solution. The goal was achieved and required, at first, to increase the ARMES spectral measurement range (Casamayou-Boucau *et al.*<sup>15</sup>). The method needed also to be validated, which was done by accurately extracting, from mixtures of non-interacting fluorophores, the spectral and anisotropy information of individual components (Casamayou-Boucau *et al.*<sup>16</sup>).

ARMES is being developed for applications in biopharmaceutical manufacturing and the analysis of proteins. In the first ARMES studies, Groza *et al.*<sup>13</sup> were for example interested in measuring and extracting the protein signal (BSA) from a complex fluorescent background of small molecules (yeastolate). The authors were also the first to combine factor based chemometrics like MCR with polarized-MDF measurements, leading to the creation of ARMES<sup>14</sup>. They observed that multi-fluorophore proteins (BSA/HSA) generated complex *aniso*-TSFS spectra, that were diagnostic of protein structure and stability. One drawback in these studies was the use of TFP which absorbed light below 290 nm, preventing accurate measurement of intrinsic protein fluorescence from Tyr and Trp. The first objective in this thesis was thus to improve the ARMES hardware by replacing the TFP with polarizers that had a better UV transmittance. This was solved by using broadband WGP, which had to be assembled in pairs to correct for their weak efficiency in the important 250 to 400 nm range. The efficacy of these new filters was validated with solutions of perylene and EB/PB. The results were in excellent agreement with the literature<sup>165,171</sup>, and confirmed the accuracy of dWGP anisotropy measurements over the most important wavelength regions required for protein analysis by intrinsic ( $\lambda_{ex} < 300$  nm) and extrinsic ( $\lambda_{ex} > 300$  nm) fluorescence.

One collateral consequence of using UV transmitting filters was to eliminate the spectral reshaping previously observed in the TSFS-ARMES studies<sup>13-14</sup>. While this distortion initially compensated for the lack of trilinearity in TSFS data, the fact that we were now able to collect the full polarized fluorescence signal had direct consequences for the data structure. In order to investigate this issue, two mixtures of non-interacting fluorophores (no IFE, no FRET) were used: a major/minor mixture first (perylene in glycerol), and a mixture of fluorophores with similar sizes (EB/PB), overlapping emission spectra but different anisotropies. The absence of IFE<sup>59</sup> and FRET<sup>299-300</sup> was critical to ensure data tri-linearity and minimize unwanted

depolarization effects. The first result showed that using dWGP, the non-trilinearity of TSFS data was restored<sup>311</sup>, making CWA unsuitable for data analysis by MCR. Instead, SWA with the appropriate constraint yielded acceptable results, at least on the perylene data. Overall, however, MCR modelling of TSFS data was less efficient than the modelling of EEM data which was, for example better able to cleanly resolve the impurity in the perylene solutions. A pre-requisite for the latter was efficient removal of the 1<sup>st</sup> order Rayleigh scatter, which was achieved by modelling<sup>222</sup> rather than interpolation<sup>306</sup>. Using the appropriate constraints, the EEM-MCR model enabled the recovery of component anisotropy quite accurately:  $\pm 4\%$  for relatively large magnitude anisotropies (*e.g.* Perylene S<sub>1</sub> and EB) but only  $\pm 20\%$  for the lower magnitude anisotropy ( $< 0.06$ ) due to noise and relative measurement error. While it could be rectified by improving the SNR, the method proved to work, and the two sample systems used provided a simple and robust way to validate the measurement and chemometric data analysis elements of ARMES. This was also the first time to show how to accurately model component anisotropy using MCR from data collected using TSFS and EEM measurement methods.

Having improved and validated the ARMES hardware and chemometric methods, the second objective of the thesis could be addressed, which was to characterize and discriminate the small soluble oligomers of insulin using their intrinsic fluorescence. Without going into the details of the results obtained (section 5.12), such discrimination was possible and brought us the ability to resolve insulin intrinsic fluorescence. Even if the SNR and spectral resolution are not ideal, one critical aspect of the study was to prove that ARMES offered in a single measurement, multiple approaches for characterizing insulin soluble oligomers, either quantitatively or qualitatively. This was interesting, but also required, because these aggregates are for example unsuitable for standard SEC analysis<sup>18</sup>, and are known to be spectroscopically difficult to characterize<sup>17</sup>.

This work developed on insulin was also highly relevant to the current trend in protein manufacturing. Indeed in a recent report addressed to protein manufacturers<sup>451</sup>, the FDA asked industries to track the presence of subvisible aggregates in therapeutic products and encouraged, along with several authors<sup>9,445,437,441</sup>, the use of orthogonal methods to complement SEC. Ideally, these methods should be fast, inexpensive, sensitive, non-destructive, and non-invasive which are all inherent attributes of ARMES.

Furthermore in the last 15 years, there has also been a growing interest towards the development of PAT<sup>528</sup>, in order to help improve process understanding and improve the

manufacturing control of biologics. A recent review highlighted in particular that spectroscopy based PAT, once combined with advanced data pre-treatment and chemometrics, were emerging as promising tools for monitoring downstream processes<sup>529</sup>. In that regard also, this insulin study was relevant because being able to assess in near-real time insulin's oligomerization would be very useful in *e.g.* purification or formulation steps.

Apart from the multiple probe aspect of the method, another facet of this thesis work was to show that robust chemometric models (here PARAFAC) could be built on the ARMES data, which would help to further characterize/discriminate the oligomers. Generally, these types of PARAFAC-EEM models are difficult to build for protein systems due to the presence of internal FRET, that generate multi-mode dependencies<sup>144,299-300</sup>. In the case of insulin, these linear dependencies could be broken due to the key location of Tyr residues in the regions responsible for dimer and hexamer formation, and enabled buried and exposed Tyr to be resolved, as well as the FRET species. One very important aspect of being able to break these dependencies (not mentioned in section 5.12) is to now possess a unique and powerful way to monitor and look at other important molecular mechanisms that can affect insulin.

Insulin can for example form insoluble fibrils when exposed to destabilising conditions<sup>71,367,391</sup>. These fibril structures can be, for certain proteins, very harmful to human health and responsible for well known pathologies such as Alzheimer's and Parkinson's diseases<sup>530-531</sup>. One aspect of amyloidogenic peptides is that they are believed to all share the same fibrillation process<sup>401</sup>. Even if the mechanism is not yet fully understood, it is commonly accepted that soluble oligomers first dissociate into monomers<sup>71,367,400,478</sup> before forming a partially unfolded molecule<sup>71,367,391</sup> which can then stack together to form a nucleus, protofibrils, and ultimately mature fibrils. Being able to detect the presence of unfolded monomer species, or simply being able to monitor the partial unfolding, would thus be of great interest.

In terms of fluorescence however, insulin is known to have an emission that is insensitive to destabilising conditions. It was illustrated in section 5.4 with the work of Bekard *et al.*<sup>454</sup>, but this is also the case when insulin is exposed to an increasing concentration of chemical denaturant (urea or guanidine hydrochloride, GuHCl), where neither the intensity nor the band maximum show much variation<sup>52,532</sup>. Due to the importance of being able to characterise the partially unfolded monomer, we also measured the ARMES data of a monomeric insulin solution (at 0.2 mg/mL) exposed to an increasing concentration of GuHCl. The details are shown in section 7.26, and the stability of the signal in Figure A- 57. While at first glance,

fluorescence does not seem adapted to this study, this section demonstrates that when the data is fitted with the model developed in section 5.10.1, the partially unfolded species can be clearly identified in a denaturant concentration range that matches other literature sources. This result is crucial as it would enable one to resolve, for the first time, the fluorescence of this critical species but also to monitor its formation. The data quality should however first be improved (SNR, resolution) by using, for example, a much faster spectrofluorometer equipped with a cooled charge-coupled device instead of a PMT detector. This would permit, within a suitable time frame (~ minute), the collection of multiple acquisitions with a better integration time/resolution. While this is the case, the results we obtained are already very encouraging.

Similarly, other complex molecular mechanisms that are either still partially unresolved or not well understood, could also be studied with this type of modelling approach once the SNR issue is addressed. These include:

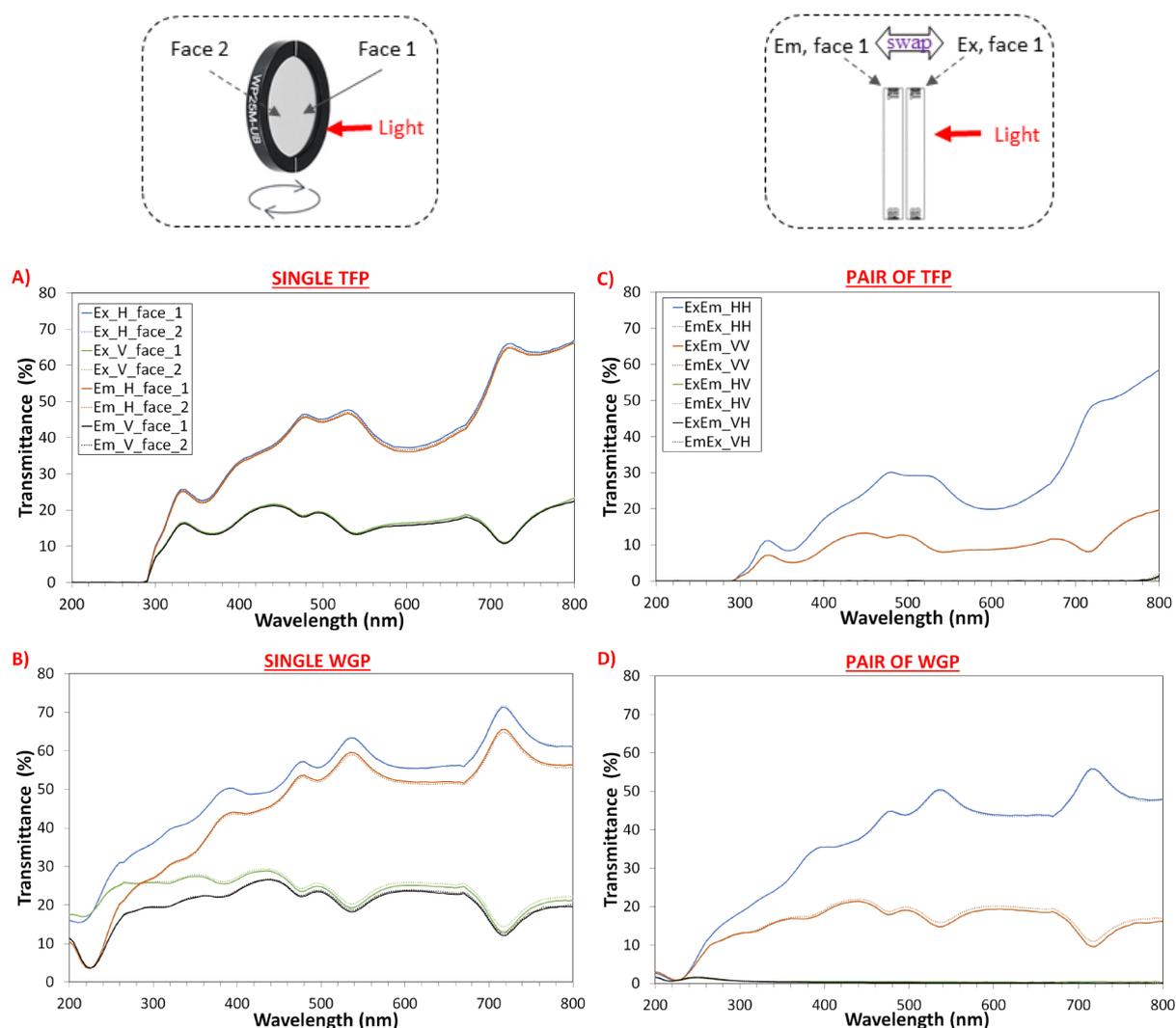
- The separation of human/bovine insulin sources, which would not be achievable by typically DLS or SEC due to their almost identical molecular weight and structure.
- The  $T_6 \rightarrow R_6$  allosteric transition mentioned in section 5.2.1.4.
- Insulin-receptor binding, as it is now clear that many of the hydrophobic residues involved in insulin dimerization are also engaged in receptor binding<sup>328,355-356,533</sup>. The detachment of B-chain C-terminal region (B21-B30) and subsequent rearrangement expose these hydrophobic residues<sup>362,533-534</sup> to form what is called the “classical binding surface” GlyA1, GlnA5, TyrA19, Asn21, ValB12, TyrB16, GlyB23, PheB24, PheB25, TyrB26<sup>408,535</sup>. There is here huge potential to study this association by ARMES, as half of the classical surface residues are fluorophores.

The application of ARMES to insulin intrinsic fluorescence, but also to other peptides, is therefore just beginning.

## 7. Appendices

### 7.1 Further analysis on filters transmittance spectra

Further analyses were carried out on the TFP and WGP, to see if there were any differences between the transmittance spectra of each individual filter faces (Figure A- 1 A,B), and between filter pair configurations where the filter positions were swapped (Figure A- 1 C, D).

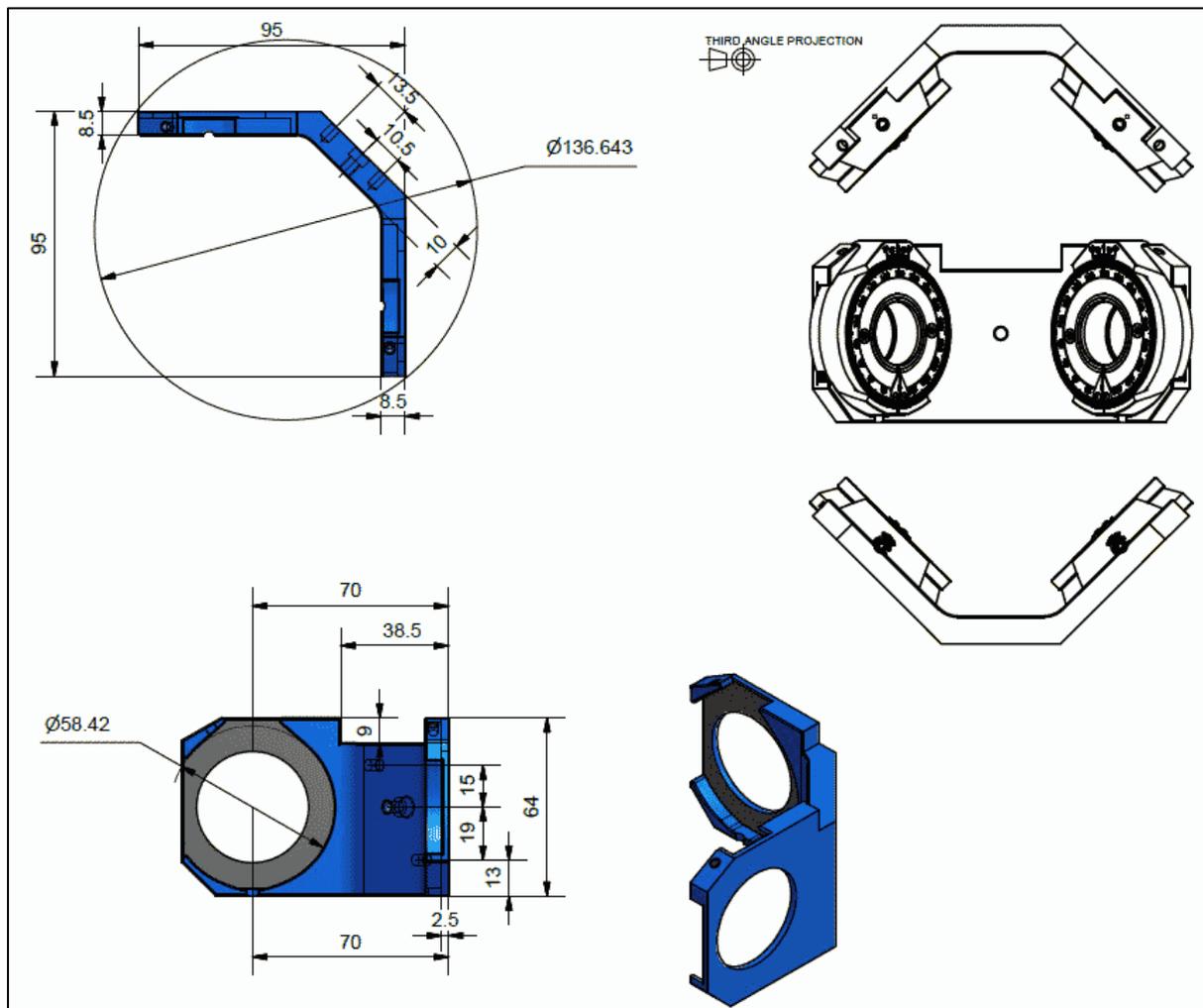


**Figure A- 1: Transmittance spectra of excitation (Ex) and emission (Em) polarizers, collected separately first on both faces (A, B) as well as together by swapping the position of Ex and Em filter (C, D). Both horizontal (H) and vertical (V) position were measured. The centre of the polarizer was aligned normal to the light beam.**

It appeared that for each excitation and emission filter analysed (WGP/ TFP), there was no difference in the transmittance spectrum for both faces. It also appeared that the transmittance spectrum of filter pairs was independent of filters positions ( $ExEm = EmEx$ ).

## 7.2 AutoCAD design for dWGP holder

The first step in getting this AutoCAD design was to use the existing Manual Polarizer Accessory (part no. 00-100761-00) to measure the important distances between the pin holes, the main screw and the centre of the polarizers. Because the rotational mounts used to fix the dWGP were thicker and wider than the mounts holding the TFP, the rest of the holder had to be adapted to house the larger wheels. The distances were calculated so that the centre of the wheel remained in precisely the same location.



**Figure A- 2: Final AutoCAD file used to produce the holder for dWGP. Designed with the help of Patrick Meier, from the Combustion Chemistry Centre (School of Chemistry, NUIG).**

### 7.3 Influence of integration time

Another parameter that was taken into consideration to improve the accuracy of the anisotropy measurements was the integration time. This parameter is basically used to select the duration during which the fluorescence signal is measured. Increasing the integration time will increase SNR, however this comes with a cost in drastically increasing the data collection time. This is of critical importance considering that ARMES will ultimately be applied to monitor various protein structural and stability issues, which require rapid measurements.

The integration time is inter-related to two other parameters: the scan speed and the data interval. The scan speed refers to the number of excitation/emission wavelengths scanned per unit of time ( $\text{nm}\cdot\text{min}^{-1}$ ). In order to ensure high-resolution spectra necessary for chemometric analysis without compromising too much on time of analysis. Groza *et al.*<sup>13</sup> suggested that a data interval of 2 nm associated with scan speed of  $1200 \text{ nm}\cdot\text{min}^{-1}$  should be used. For the same reason, we also chose to fix the data interval to 2 nm which means that an increase in integration time will automatically be compensated by a decrease in scan speed (Table A- 1).

**Table A- 1: Accuracy of anisotropy measurements obtained for  $1\mu\text{M}$  solution of perylene in glycerol, using dWGP ( $25^\circ\text{C}$ ) and various integration times.**

Integration time (s)	Data Interval (nm)	Scan speed ( $\text{nm}\cdot\text{min}^{-1}$ )	$r \pm \text{STD (RSD)}$ $\lambda_{\text{ex}} = 244\text{-}264 \text{ nm}$	$r \pm \text{STD (RSD)}$ $\lambda_{\text{ex}} = 390\text{-}430 \text{ nm}$
0.1	2	1200	$-0.035 \pm 0.006$ (17.14%)	$0.200 \pm 0.003$ (1.50%)
0.2	2	600	$-0.034 \pm 0.005$ (14.70%)	$0.196 \pm 0.003$ (1.53%)
0.5	2	240	$-0.033 \pm 0.003$ (9.09%)	$0.198 \pm 0.002$ (1.01%)

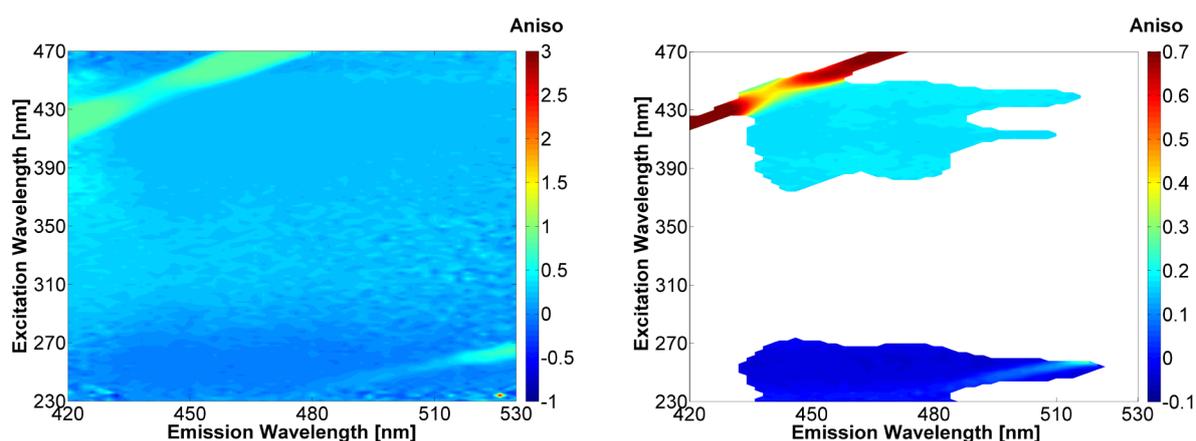
For anisotropy measurements, Agilent recommend in the Cary eclipse software to use a minimum integration time of 0.25s. Looking at Table A- 1 , the RSD decreased as expected for longer integration times but the accuracy of anisotropy values only changed in the last decimal place. The values obtained with 0.5s integration time are even closer to the results from Barkley *et al.* ( $-0.031$  and  $0.197$  respectively for short and long wavelength range). However considering that we will collect multidimensional measurements, an integration time of 0.5s

compared to 0.1s will increase the data collection time by a factor of 5. The gain in anisotropy accuracy is too small to consider this option as worthwhile<sup>yy</sup>.

## 7.4 *Aniso*-MDF spectra, threshold and measurement dispersion

Using either EEM or TSFS and the four polarization settings HH, HV, VH, and VV, it is possible to build a new data matrix that we called an *aniso*-MDF spectra (EEM:  $\lambda_{\text{ex}} \times \lambda_{\text{em}} \times r$  ; TSFS:  $\lambda_{\text{ex}} \times \Delta\lambda \times r$ ) based on the anisotropy value at each point of the space instead of the fluorescence intensity.

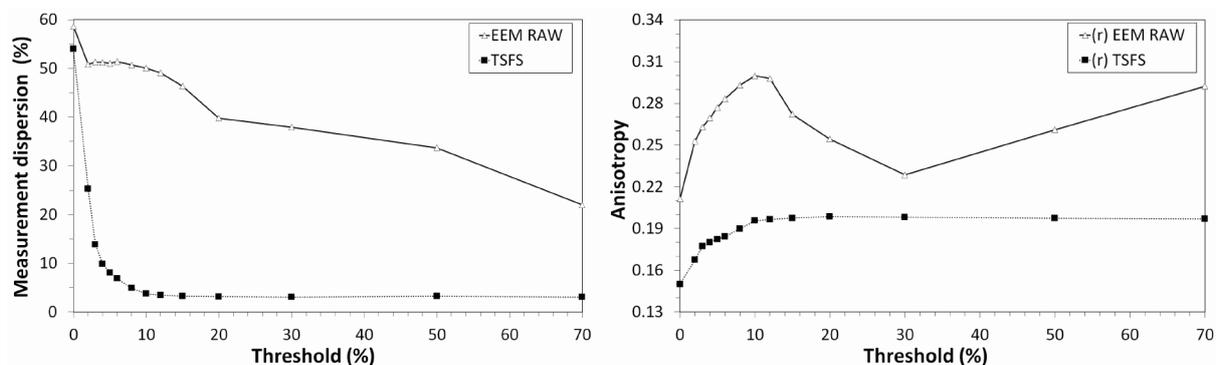
However in the region of low spectral intensity where noise levels are high, the calculated anisotropy values are essentially random and distort the *aniso*-MDF spectra (Figure A- 3, left). In order to get accurate *aniso*-MDF spectra, these artefacts need to be removed by setting a threshold to eliminate low intensity, noisy regions. The threshold is set on the polarization setting that showed the weakest intensity: HH for S<sub>2</sub> ( $\lambda_{\text{ex}} = 230\text{-}340$  nm) and VV for S<sub>1</sub> ( $\lambda_{\text{ex}} = 342\text{-}470$  nm). If a threshold of 10% is applied as in the example below (Figure A- 3, right), it means mathematically that all the points having an intensity below 10% of HH or VV maximum (depending on the spectral region) are removed<sup>14</sup>. In fact, the points are set as NaN in MATLAB which mean that they are not considered as numbers anymore, thus improving the accuracy of the depicted anisotropy plots.



**Figure A- 3: Anisotropy-EEM contour plot of 1 $\mu$ M perylene in 100% glycerol shown (left) without spectra threshold and (right) with a 10% threshold of the HH scan for  $\lambda_{\text{ex}} = 230 - 340$  nm, and VV scan for  $\lambda_{\text{ex}} = 342 - 470$  nm. Data collected using dWGP, in triplicate at 25  $^{\circ}$ C.**

<sup>yy</sup> While this statement is true in chapters 3 and 4, we believe however that for protein analysis where the SNR is low, this statement should be revised because a longer integration time is crucial to obtain better quality data (as outlined in the final conclusions).

The polarized EEM spectra of Perylene are contaminated by both 1<sup>st</sup> and 2<sup>nd</sup> order Rayleigh scatter (Figure A- 3), and this is particularly noticeable in the *aniso*-MDF spectra because of the very high *r* values. The measurement dispersion, that represents how much *r* values vary within a selected region, is affected by the presence of Rayleigh scatter (Figure A- 4 left), scatter that also affects the final *r* value averaged across that region (Figure A- 4 right).



**Figure A- 4: (left) Plot of the measurement dispersion in *r* values vs threshold and (right) average *r* values vs threshold for the EEM (triangle) and TSFS (square) measurements. The data were obtained from 1  $\mu$ M perylene in 100% glycerol, collected using dWGP in triplicate at 25  $^{\circ}$ C. Calculations were done using the raw measurements across the S<sub>1</sub> perylene band.**

The idea is that because the measurement dispersion is very sensitive to the presence of Rayleigh scatter, a good correction method should reduce EEM *r* dispersion close to the values obtained for TSFS one (which are free from Rayleigh scatter). The best threshold value can then be selected for EEM-corrected data, by choosing carefully the option that gives the lowest measurement dispersion (ensuring accurate *r* values) while preserving the highest *r* region.

## 7.5 Slight modifications of Rinnan's method to deal with 2<sup>nd</sup> order Rayleigh scatter

Below are shown the modifications (in red) that had to be implemented in the flumod function developed by Rinnan *et al.*<sup>222</sup>, to handle the 2<sup>nd</sup> order Rayleigh scatter. These were done at the step where the algorithm is building up the matrix (*X<sub>r</sub>*) that is used afterwards to model the scatter. Instead of saying to the algorithm to find the values around the 1<sup>st</sup> order region ( $\lambda_{em} = \lambda_{ex} \pm$  scatter width), we simply told it to do the same but where  $\lambda_{em} = 2 \times \lambda_{ex}$ , corresponding to the 2<sup>nd</sup> order region. This function was saved as flumod2.

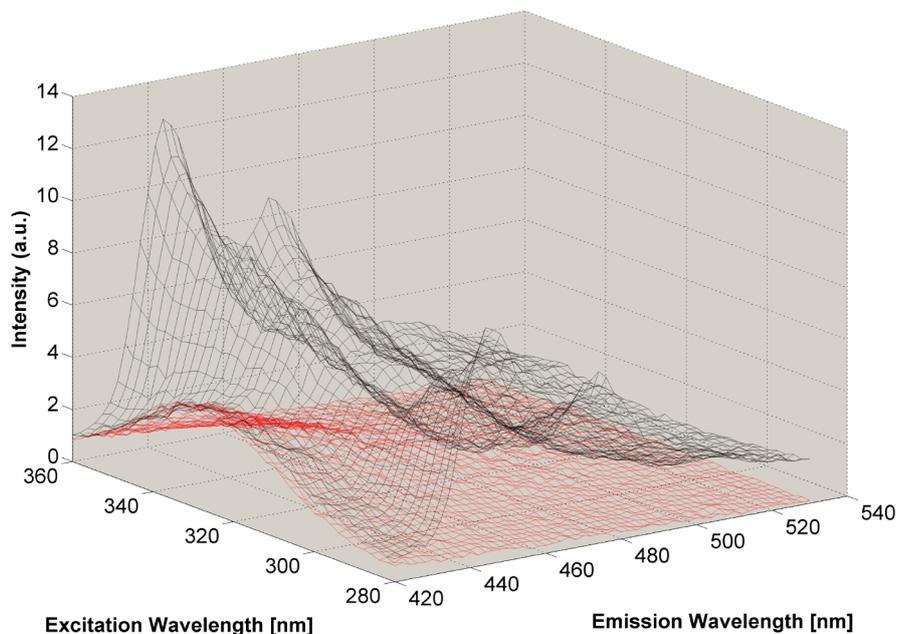
```
Xr=ones(size(X,1),wid*2+1,length(Ex))*NaN;
for i=1:length(Ex)
    k=find(Em<=2*Ex(i)+wid & Em>=2*Ex(i)-wid);
    j=find(abs(Em-2*Ex(i))==min(abs(Em-2*Ex(i))));
    m=find(j(1)==k);
```

```

xray=squeeze(X(:,k,i));
pos(i,1)=k;
pos(i,2)=round(Em(k)-2*Ex(i))+wid+1;
Xr(:,pos(i,2),i)=xray;
end;

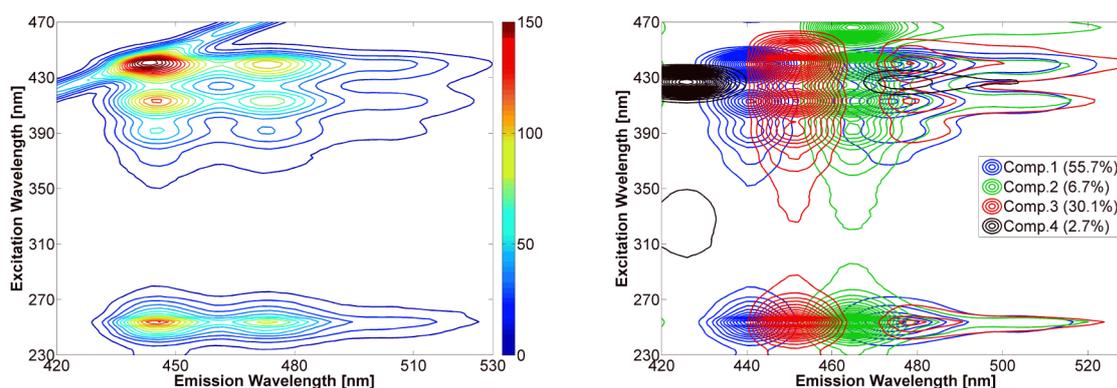
```

## 7.6 Presence of an impurity in the glycerol

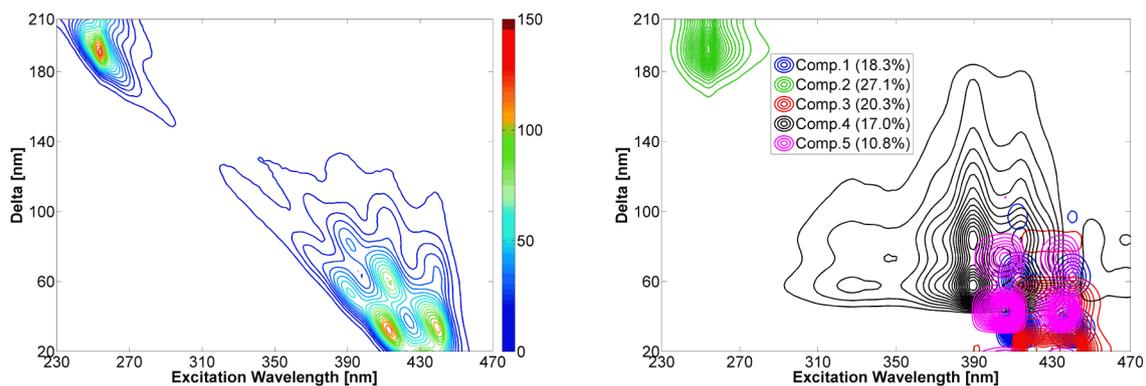


**Figure A- 5: Overlay of HH-EEM spectra of 1  $\mu$ M perylene in glycerol (black) with HH-EEM spectra of pure glycerol (red) in region of interest.**

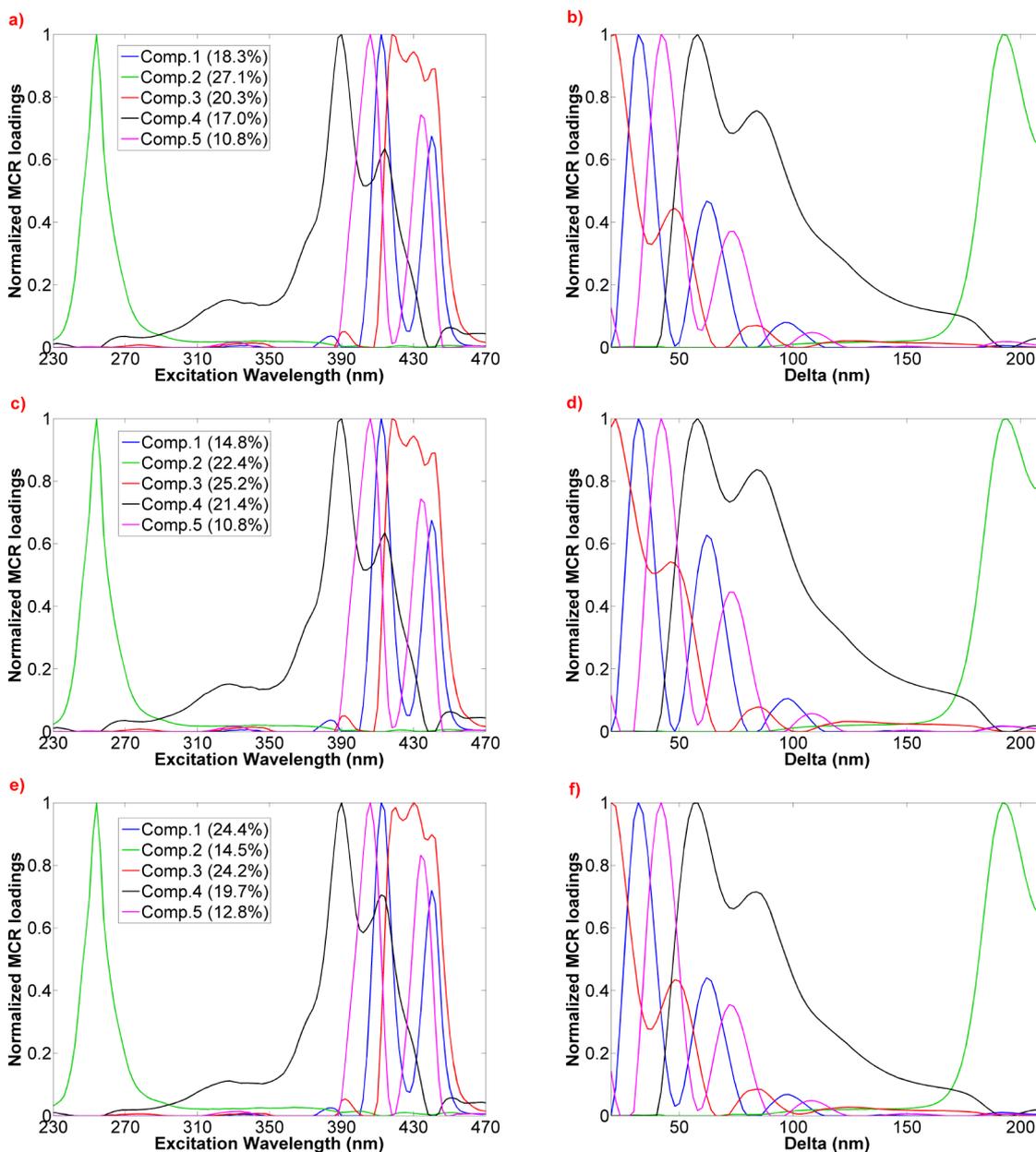
## 7.7 CWA-MCR modelled perylene data

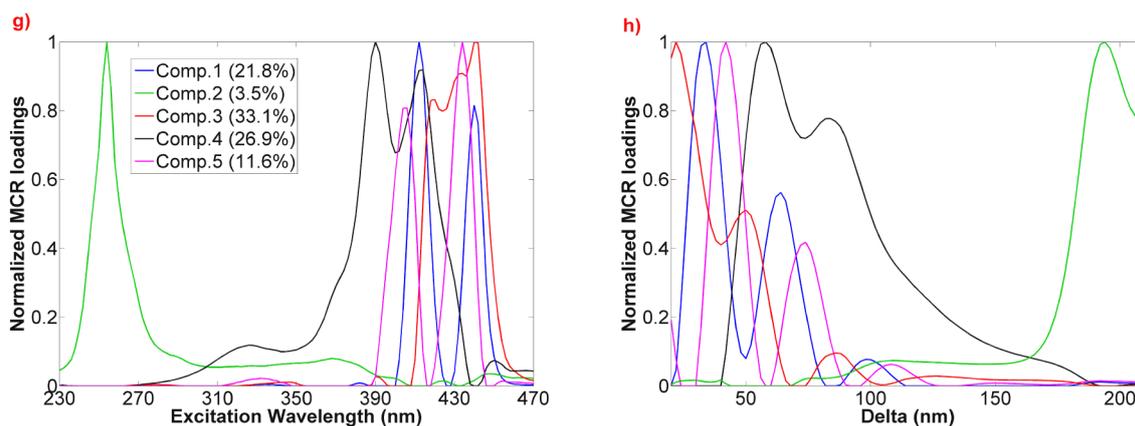


**Figure A- 6: Effect of Rayleigh scatter contamination: (Left) HH-EEM polarized spectra of 1 $\mu$ M solution of perylene in Glycerol (Right) Pure components extracted from MCR model run on raw CWA-HH-EEM dataset, contaminated by 1<sup>st</sup> order Rayleigh scatter. Models obtained using: 1). normalization, and 2). non-negativity constraints.**

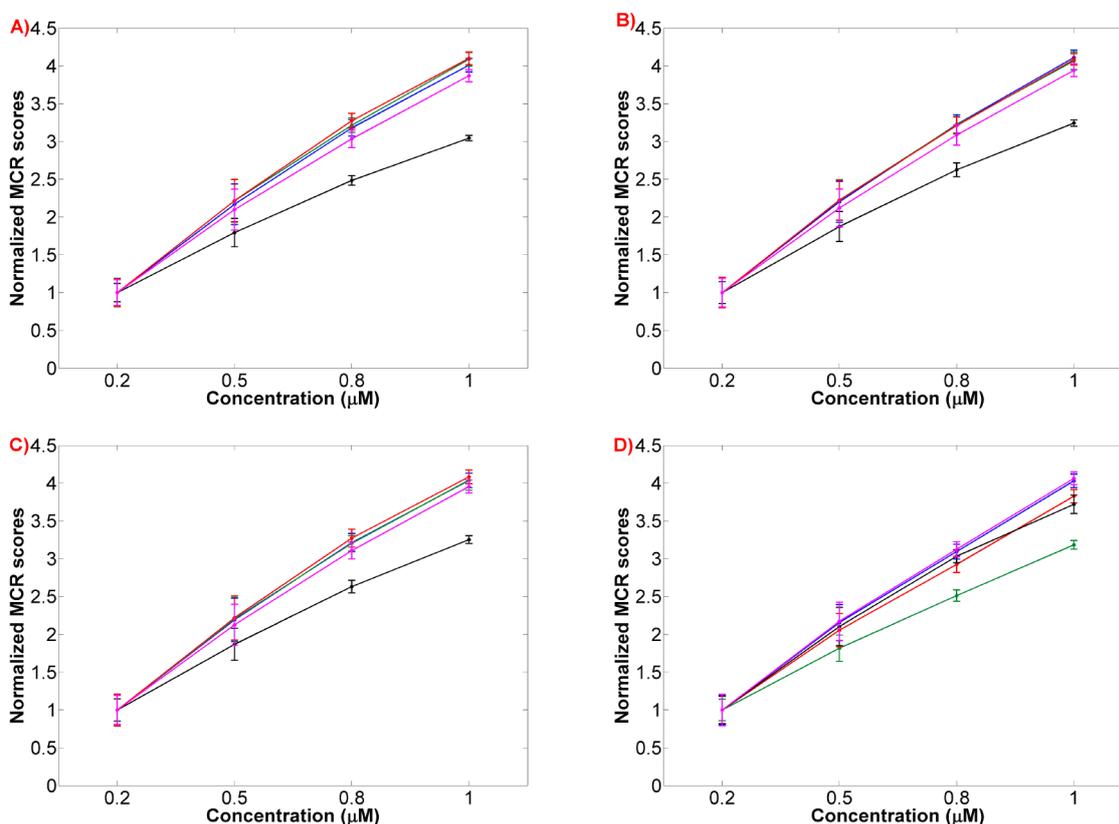


**Figure A- 7: (Left) HH-TSFS polarized spectrum of a 1µM solution of perylene in Glycerol; (Right) Pure components extracted from MCR model run on the CWA HH-TSFS perylene dataset. Model obtained using: 1). normalization, and 2). non-negativity constraints.**

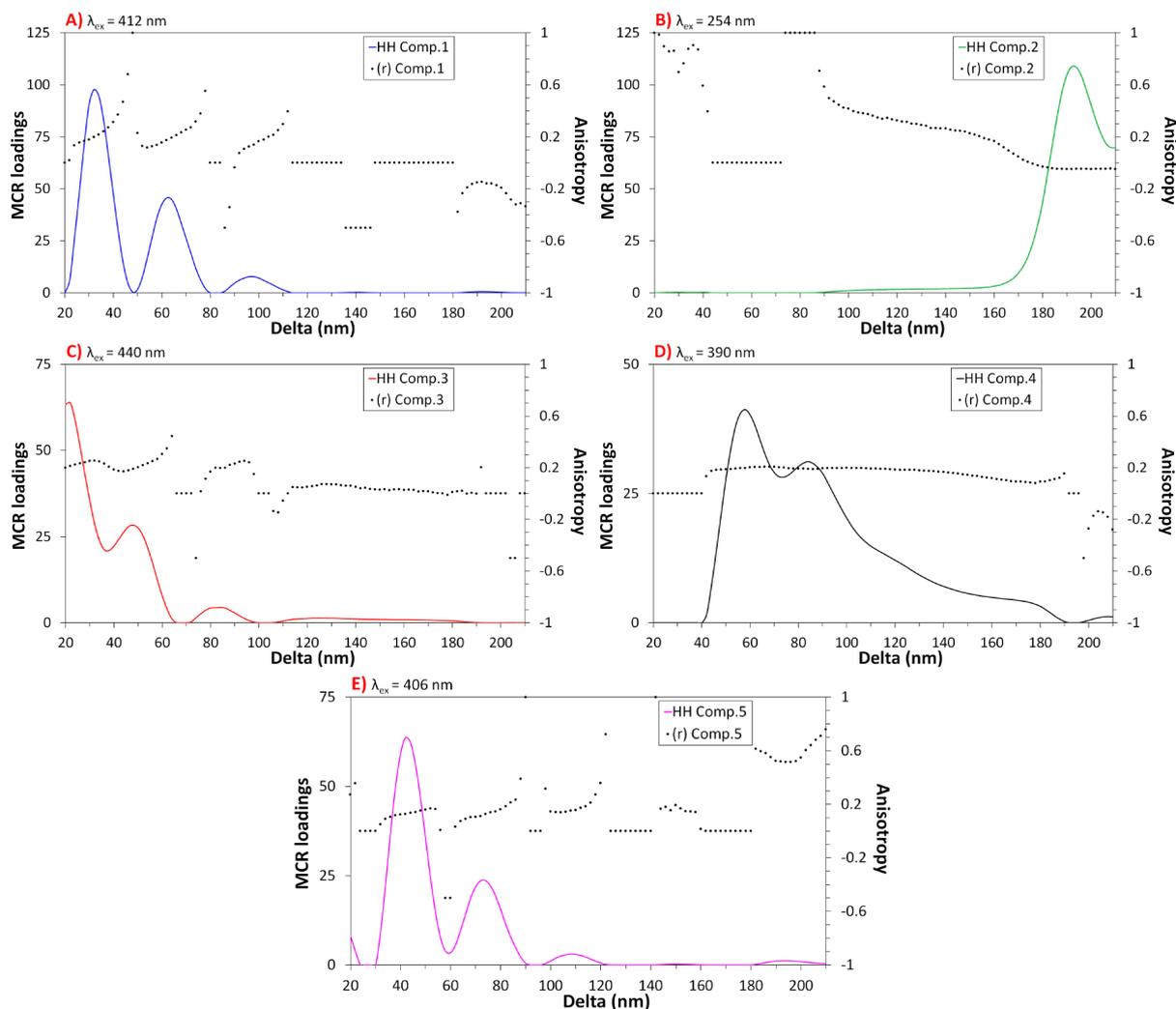




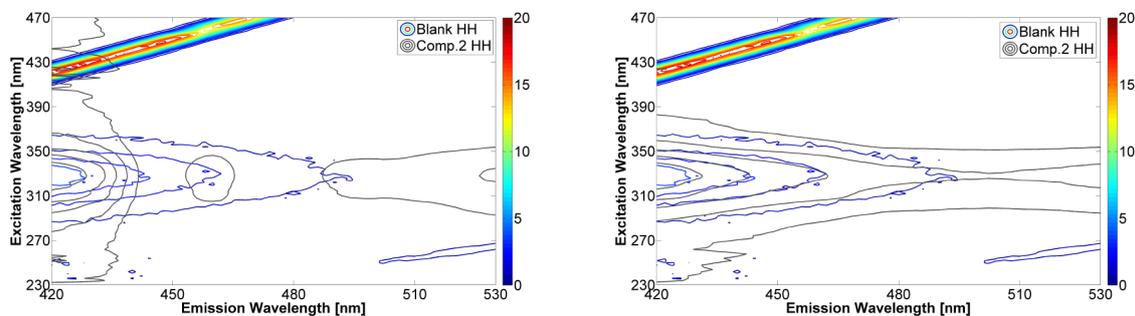
**Figure A- 8: Normalized excitation (averaged across all samples, left) and normalized  $\Delta\lambda$  (right) profiles obtained for each pure component after MCR modelling of the CWA HH-TSFS perylene dataset(s): HH (a/b), HV (c/d), VH (e/f), and VV (g/h) polarized data. Models obtained using: 1). normalization, and 2). non-negativity constraints.**



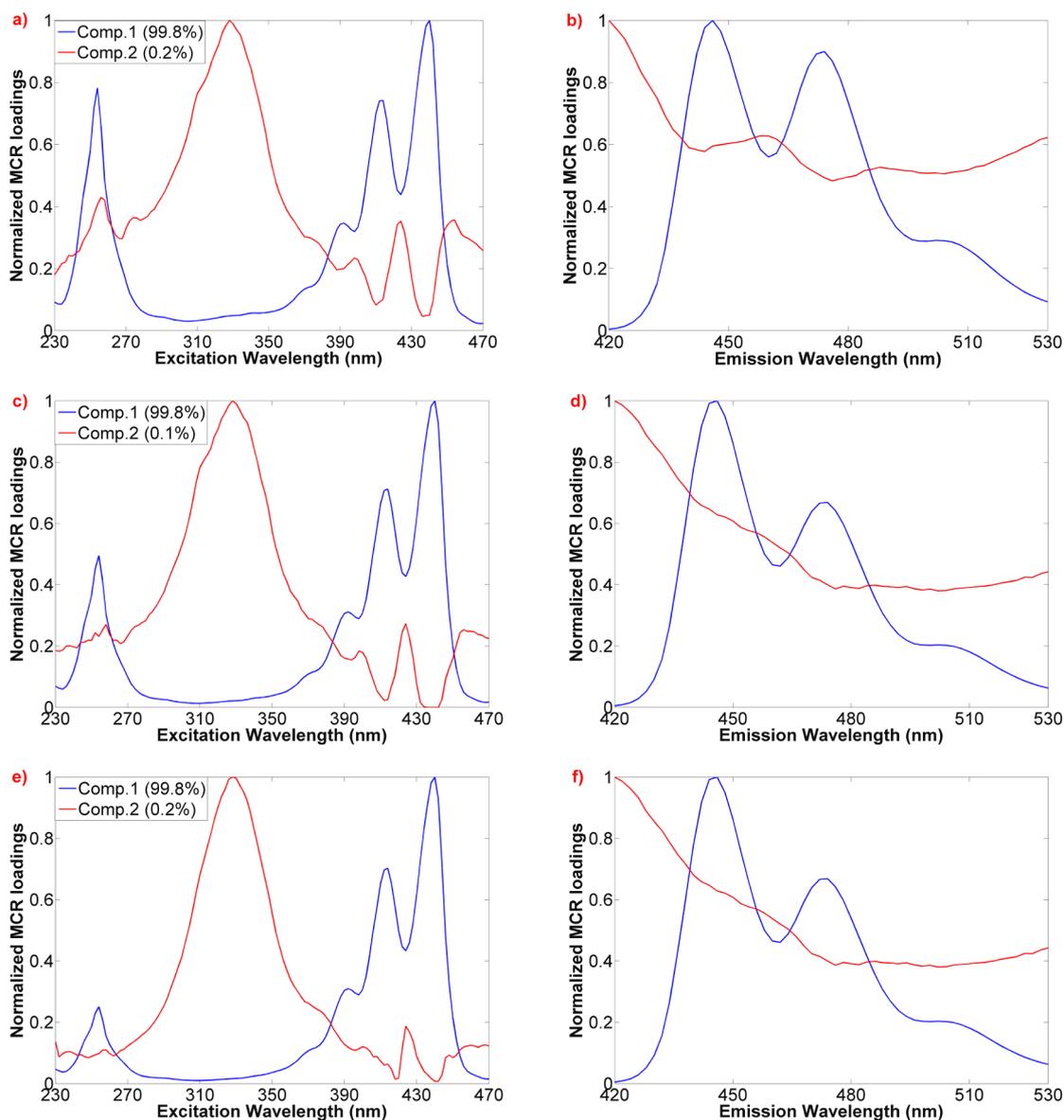
**Figure A- 9: MCR scores (scaled to 1<sup>st</sup> sample) obtained for each pure component after modelling of the CWA HH-TSFS perylene datasets: HH (A), HV (B), VH (C), and VV (D) polarized data. Models obtained using: 1). normalization, and 2). non-negativity constraints.**



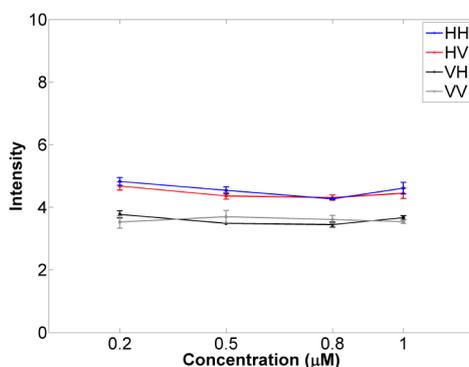
**Figure A- 10: Recalculated anisotropy (dots) for MCR components (CWA-TSFS perylene dataset using 1). normalization and 2). non-negativity), resolved from polarized spectra, overlaid with HH emission generated by MCR (continuous line).  $\lambda_{ex}$  correspond to the wavelength at which the excitation loading was at his maximum for each component, thus used to observe the emission. MCR scores (scaled to 1<sup>st</sup> sample) obtained for each pure component after modelling: HH (A), HV (B), VH (C), and VV (D) polarized data.**



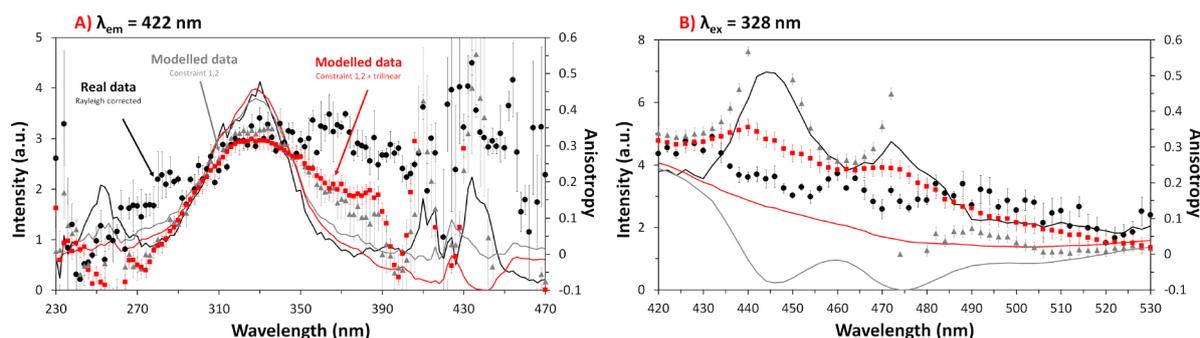
**Figure A- 11: Overlay between HH-EEM contour plot of blank (pure glycerol) and 2<sup>nd</sup> component, resolved using non-negativity and normalization (left), and adding the total trilinear constraint (right).**



**Figure A- 12: Normalized excitation (averaged across all samples, left) and emission (right) profiles obtained for each pure component after MCR modelling of the HV (a/b) VH (c/d) VV (e/f) polarized EEM datasets. Models obtained using: 1). normalization, 2). non-negativity constraints, and 5). total trilinear constraints.**



**Figure A- 13: Fluorescence intensity measured at  $\lambda_{ex}/\Delta\lambda = 328/70$  nm (peak of fluorescent impurity) for HH, HV, VH, VV polarized TSFS data.**



**Figure A- 14: Overlay of HH-EEM perylene spectra with anisotropy data for: A). Excitation spectra observed at  $\lambda_{em} = 422$  nm, and B). Emission spectra observed at  $\lambda_{ex} = 328$  nm. Data generated from a 1  $\mu$ M solution of perylene in glycerol. The plotted anisotropy data points are: black = real data; grey = modelled using constraints 1,2; red = modelled using constraints 1,2 and tri-linearity. The excitation and emission wavelengths correspond to the maximum intensity of the impurity.**

**Table A- 2: MCR model for CWA-TSFS data from Perylene sample set, using: 1). normalization, and 2). non-negativity. Percentage of variance captured by each component (Fit %X<sup>zz</sup>). The model converged below 100 iterations.**

Component number	HH Fit (%X) vs It	HV (%X) vs It	VH (%X) vs It	VV (%X) vs It
	100	100	100	100
1	18.3%	14.8%	24.4%	21.8%
2	27.1%	22.4%	14.5%	3.5%
3	20.3%	25.2%	24.2%	33.1%
4	17.0%	21.4%	19.7%	26.9%
5	10.8%	10.8%	12.8%	11.6%
Sum	93.5%	94.6%	95.6%	96.9%
Lack of fit	20.4%	18.4%	16.7%	13.6%

**Table A- 3: Estimation of rotational ambiguities per component by MCR-BANDS, for each CWA-TSFS model from Perylene sample set using: 1). normalization, and 2). non-negativity.**

Components	Models	$f_n$ initial	$f_n$ max	$f_n$ min	$f_n$ max- $f_n$ min
Comp.1	HH	0.365	0.365	0.365	0.000
	HV	0.319	0.319	0.319	0.000
	VH	0.401	0.401	0.401	0.000
	VV	0.362	0.362	0.362	0.000

<sup>zz</sup> %X is the percentage of variance captured by each component = sum-squared signal relative to the total signal in the data

Comp.2	HH	0.444	0.444	0.444	0.000
	HV	0.392	0.392	0.392	0.000
	VH	0.399	0.399	0.399	0.000
	VV	0.446	0.446	0.446	0.000
Comp.3	HH	0.384	0.384	0.384	0.000
	HV	0.415	0.415	0.415	0.000
	VH	0.309	0.309	0.309	0.000
	VV	0.403	0.403	0.403	0.000
Comp.4	HH	0.351	0.351	0.351	0.000
	HV	0.383	0.383	0.383	0.000
	VH	0.360	0.360	0.360	0.000
	VV	0.265	0.265	0.265	0.000
Comp.5	HH	0.280	0.280	0.280	0.000
	HV	0.272	0.272	0.272	0.000
	VH	0.291	0.291	0.291	0.000
	VV	0.145	0.145	0.145	0.000

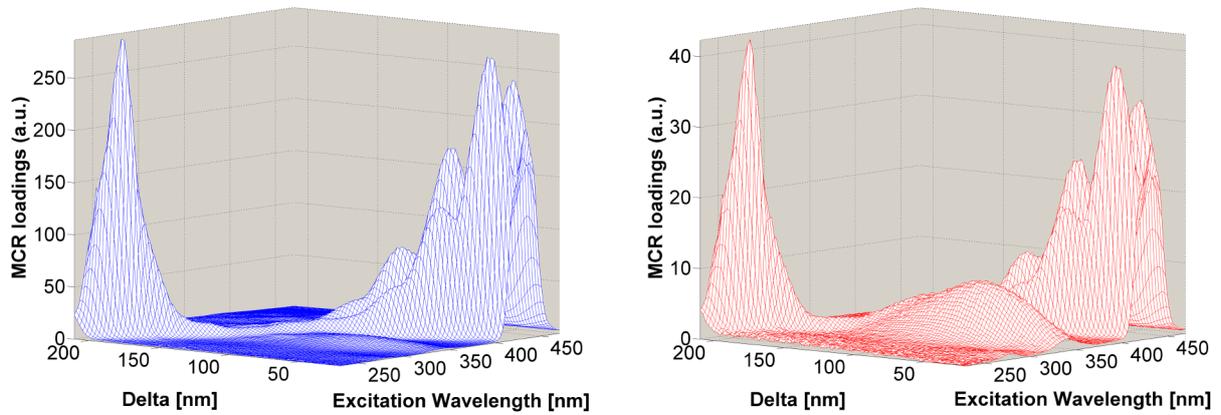
**Table A- 4: Estimation of rotational ambiguities per component, for each polarized-EEM model from Perylene sample, using: 1). normalization, and 2). non-negativity.**

1,2	Component 1 (~99.8%)				Component 2 (~0.1%)			
	HH	HV	VH	VV	HH	HV	VH	VV
$f_n$ initial	0.996	0.995	0.997	0.997	0.032	0.031	0.026	0.028
$f_n$ max	0.996	0.995	0.997	0.997	0.050	0.053	0.041	0.041
$f_n$ min	0.985	0.983	0.988	0.990	0.032	0.031	0.025	0.028
$f_n$ max- $f_n$ min	0.011	0.012	0.009	0.007	0.018	0.022	0.015	0.013

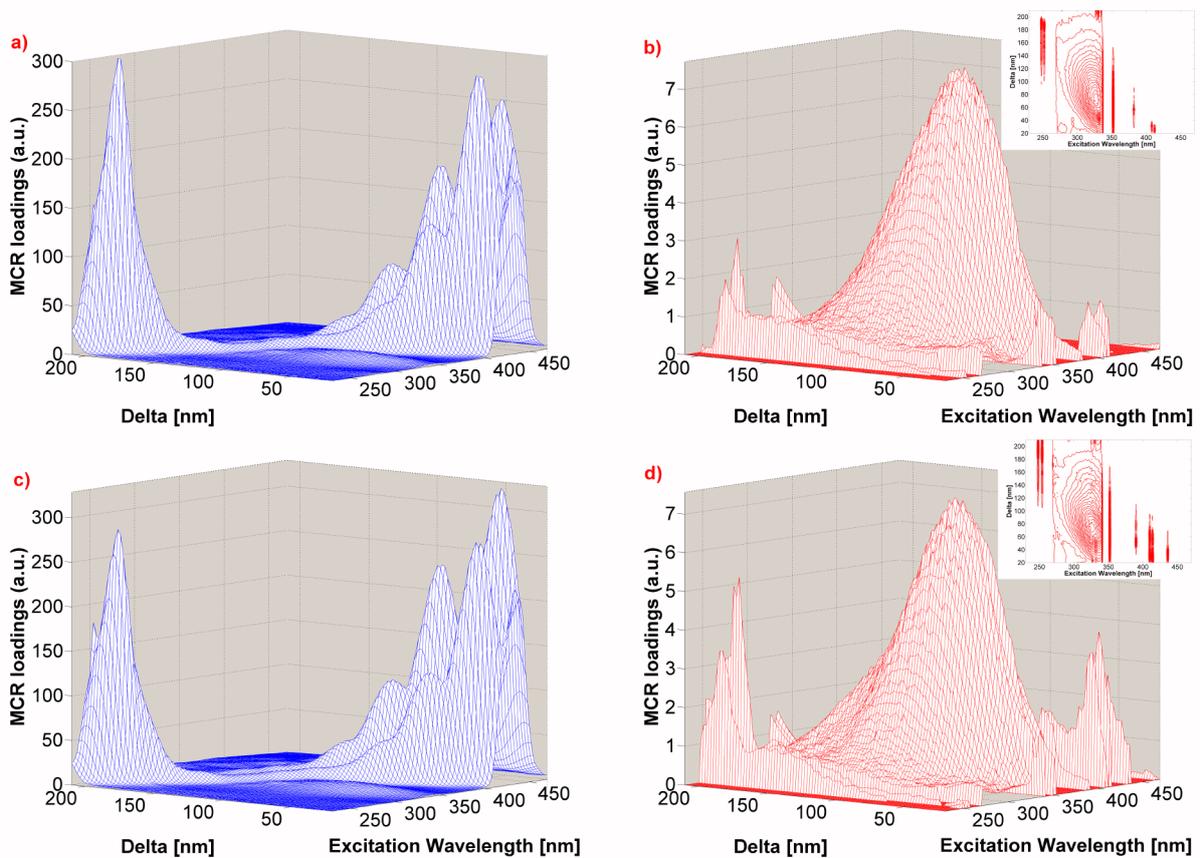
**Table A- 5: Real and modelled (from EEM-MCR, constraints 1,2) averaged  $r$  values calculated for each concentration, for  $\lambda_{ex/em}$  ( $S_2$ ) = 240-270/430-520 nm and  $\lambda_{ex/em}$  ( $S_1$ ) = 370-460/430-520 nm.**

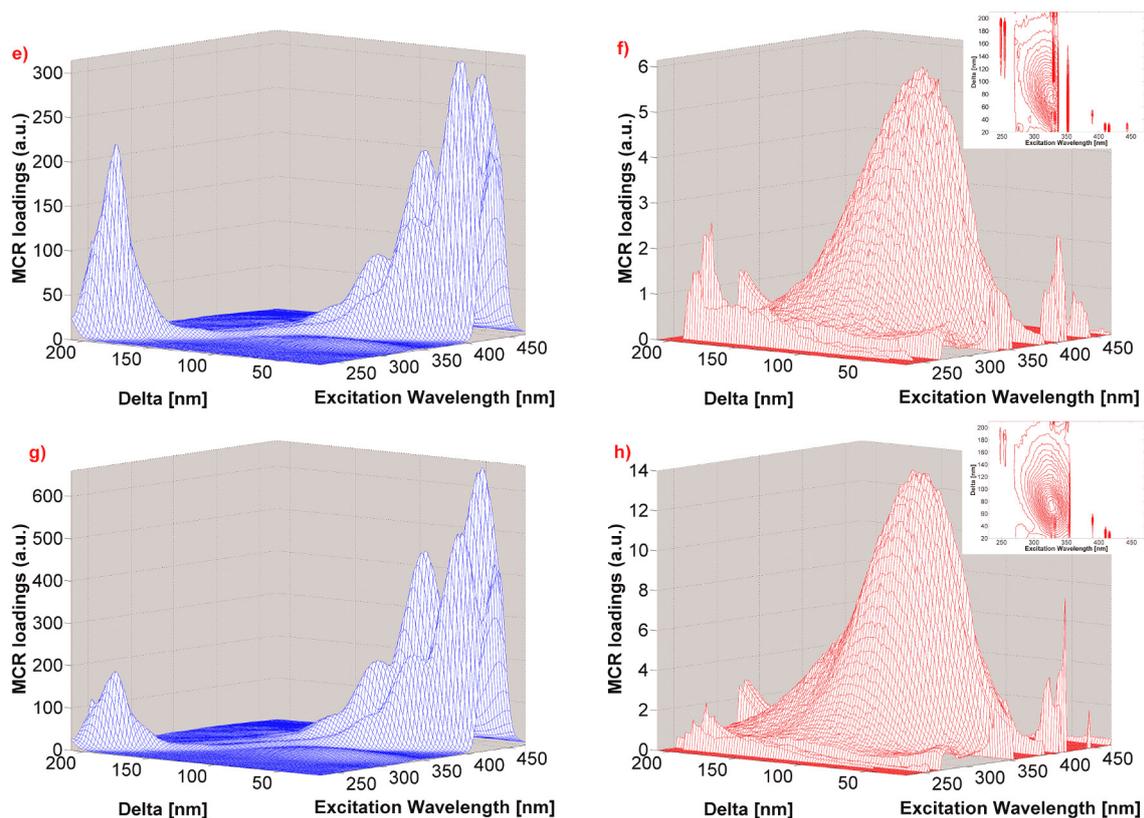
	Real anisotropy $\pm$ STD		Modelled anisotropy $\pm$ STD	
	$S_2$	$S_1$	$S_2$	$S_1$
0.2 $\mu$ M	-0.034 $\pm$ 0.010	0.204 $\pm$ 0.015	-0.037 $\pm$ 0.011	0.203 $\pm$ 0.015
0.5 $\mu$ M	-0.026 $\pm$ 0.004	0.202 $\pm$ 0.007	-0.026 $\pm$ 0.004	0.202 $\pm$ 0.007
0.8 $\mu$ M	-0.030 $\pm$ 0.004	0.192 $\pm$ 0.004	-0.029 $\pm$ 0.003	0.193 $\pm$ 0.004
1 $\mu$ M	-0.035 $\pm$ 0.004	0.204 $\pm$ 0.002	-0.033 $\pm$ 0.004	0.205 $\pm$ 0.002
Av/std	-0.031 $\pm$ 0.004	0.201 $\pm$ 0.005	-0.031 $\pm$ 0.004	0.201 $\pm$ 0.005

## 7.8 SWA-MCR modelled perylene data

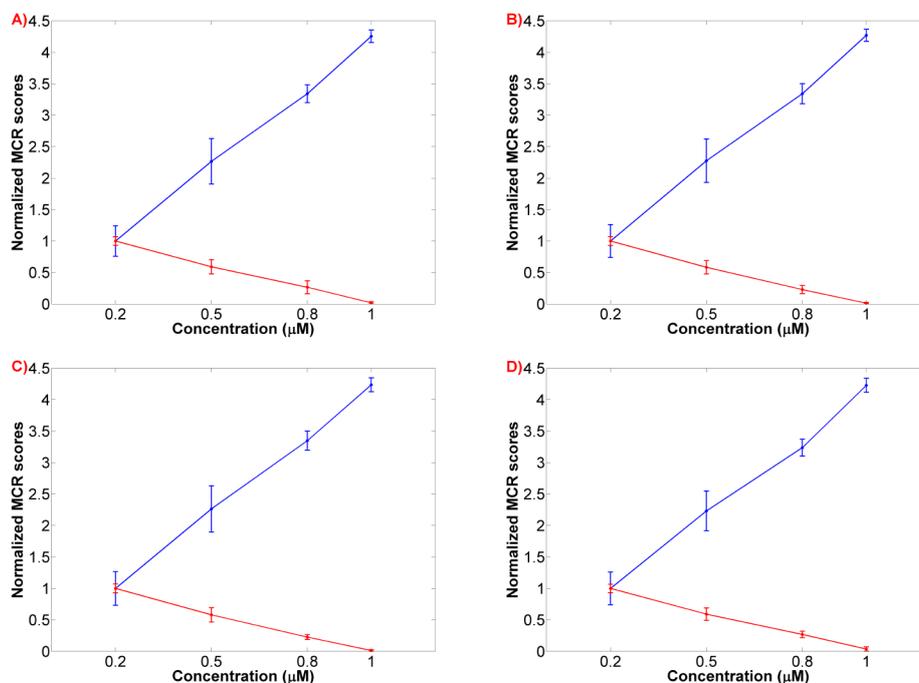


**Figure A- 15:** MCR loadings obtained for both components (Comp.1 blue, Comp.2 red) after modelling of the sample-wise augmented HH polarized-TSFS datasets (using 1. normalization and 2. non-negativity constraints).

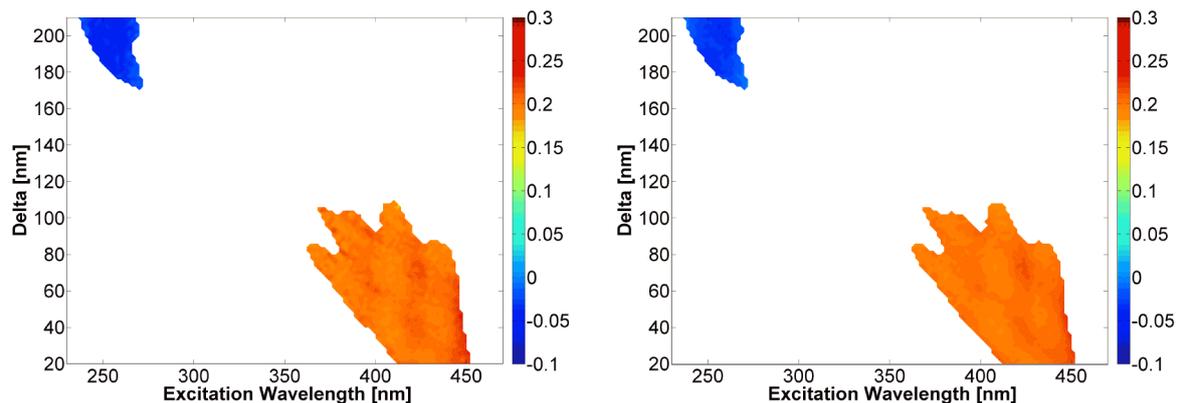




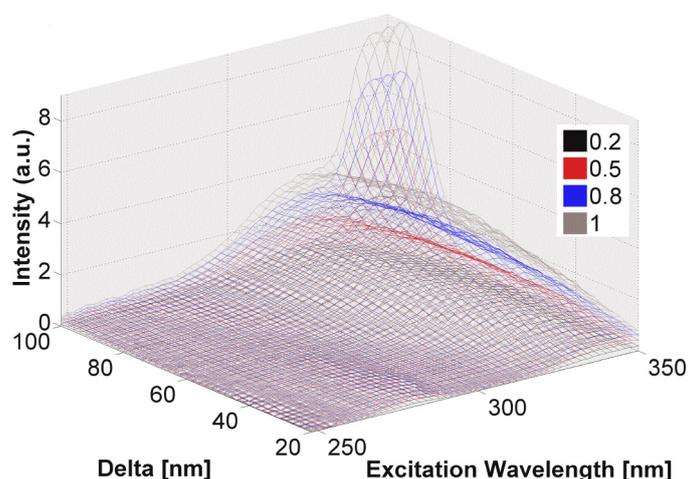
**Figure A- 16: MCR loadings obtained for both components (Comp.1 blue, Comp.2 red) after modelling of the sample-wise augmented HH (a/b), HV (c/d), VH (e/f), and VV (g/h) polarized-TSFS datasets (using 1. normalization 2. non-negativity 3. unimodality constraints).**



**Figure A- 17: MCR scores (scaled to 1<sup>st</sup> sample) obtained for each pure component after modelling of the SWA-TSFS perylene data: HH (A), HV (B), VH (C), and VV (D) polarized-TSFS datasets (using 1). Normalization, 2). non-negativity, and 3). unimodality constraints).**



**Figure A- 18: Real (left) and reconstructed (right) anisotropy patterns obtained for 1  $\mu\text{M}$  perylene sample in glycerol (average over 3 replicates). The reconstructed pattern was built using 1<sup>st</sup> component of SWA-TSFS models (using 1. normalization 2. non-negativity 3. unimodality constraints). The colour bar represents the anisotropy values.**



**Figure A- 19: TSFS plot in region where the impurity emitted, showing that the impurity was not entirely removed from 1<sup>st</sup> component extracted by SWA-TSFS model (using 1). Normalization, 2). non-negativity, and 3). unimodality constraints). The quantity of impurity remaining increased with higher perylene concentrations which explained the decreasing scores in Figure A- 17.**

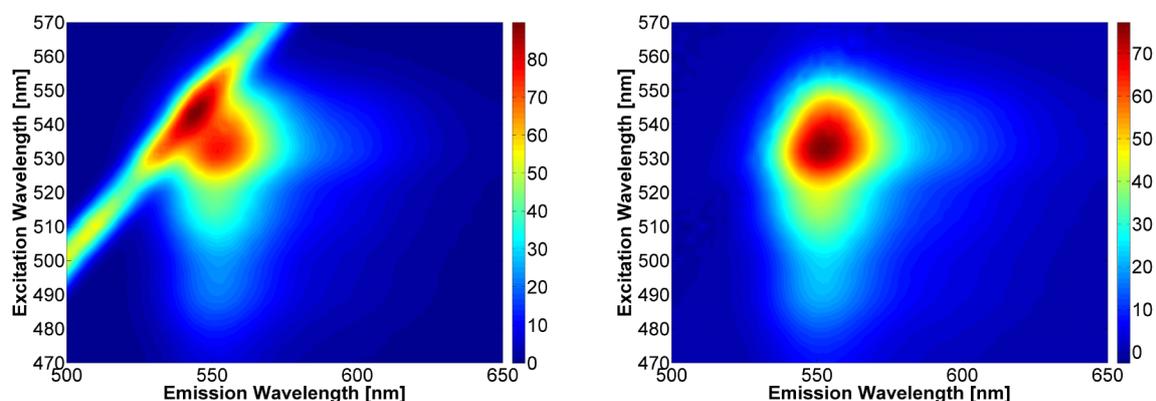
**Table A- 6: MCR model for SWA-TSFS data from Perylene sample set, using 1) normalization and 2) non-negativity constraints. Percentage of variance captured by each component (Fit %X). The model converged below 100 iterations.**

SWA-TSFS MCR model.				
Comp	HH Fit (%X)	HV (%X)	VH (%X)	VV (%X)
1	97.7%	97.9%	98.1%	97.8%
2	2.2%	2.0%	1.8%	2.1%
Sum	99.9%	99.9%	99.9%	99.9%
Lack of Fit	1.5%	1.4%	1.4%	1.1%

**Table A- 7: Real and modelled (from SWA-TSFS using 1. normalization 2. non-negativity 3. unimodality constraints) averaged  $r$  values calculated for each concentration, for  $\lambda_{ex}/\Delta\lambda$  ( $S_2$ ) = 240-270/20-210 nm and  $\lambda_{ex}/\Delta\lambda$  ( $S_1$ ) = 370-460/20-210 nm.**

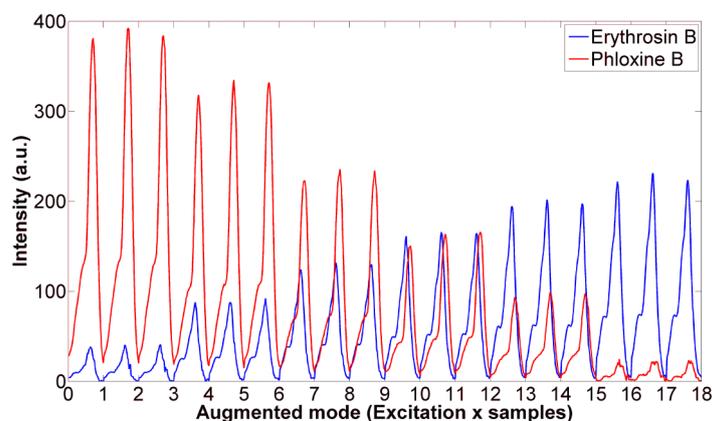
	Real anisotropy $\pm$ STD		Modelled anisotropy $\pm$ STD	
	$S_2$	$S_1$	$S_2$	$S_1$
<b>0.2 <math>\mu</math>M</b>	$-0.034 \pm 0.009$	$0.202 \pm 0.009$	$-0.024 \pm 0.005$	$0.201 \pm 0.005$
<b>0.5 <math>\mu</math>M</b>	$-0.029 \pm 0.004$	$0.195 \pm 0.004$	$-0.026 \pm 0.003$	$0.199 \pm 0.005$
<b>0.8 <math>\mu</math>M</b>	$-0.031 \pm 0.002$	$0.191 \pm 0.004$	$-0.030 \pm 0.002$	$0.193 \pm 0.003$
<b>1 <math>\mu</math>M</b>	$-0.027 \pm 0.003$	$0.201 \pm 0.001$	$-0.020 \pm 0.001$	$0.204 \pm 0.001$
<b>Av/std</b>	$-0.030 \pm 0.003$	$0.201 \pm 0.005$	$-0.025 \pm 0.004$	$0.199 \pm 0.005$

## 7.9 Fluorophore mixture modelling

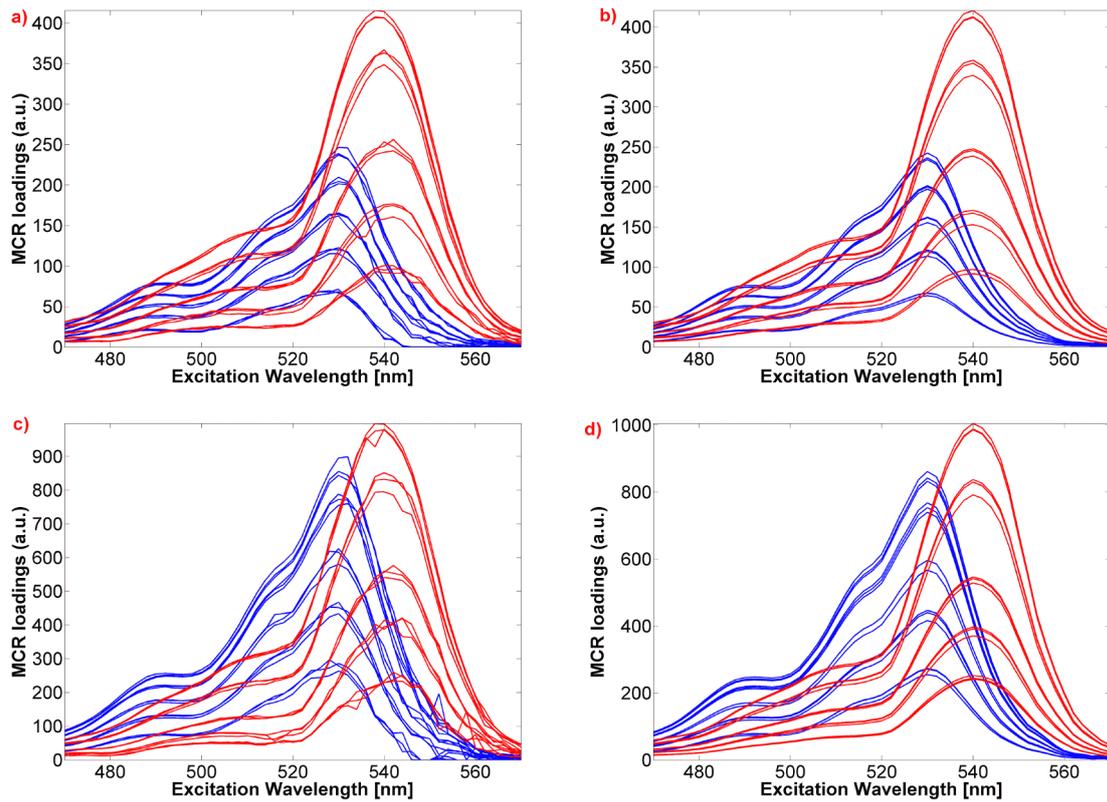


**Figure A- 20: HH-EEM spectra obtained by averaging samples 10 to 12 before (left) and after (right) Rayleigh scatter correction by modelling.**

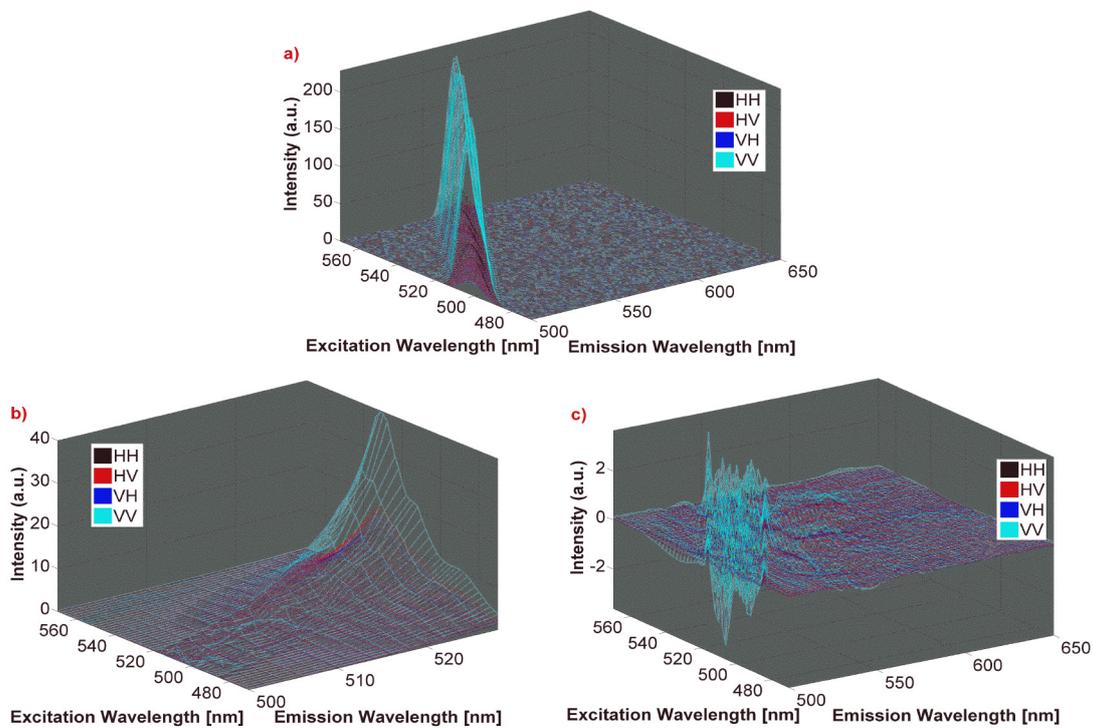
As shown in Figure A- 21, Eb and PB were resolved in all samples which was not correct since samples 1-3 contained only pure PB, and samples 16 to 18 were pure EB.



**Figure A- 21: Augmented mode obtained after resolution of HH dataset, using normalization and non-negativity constraints on both concentration and spectra matrices. For clarity the x axis is labelled with numbers designating each sample instead of excitation variables.**

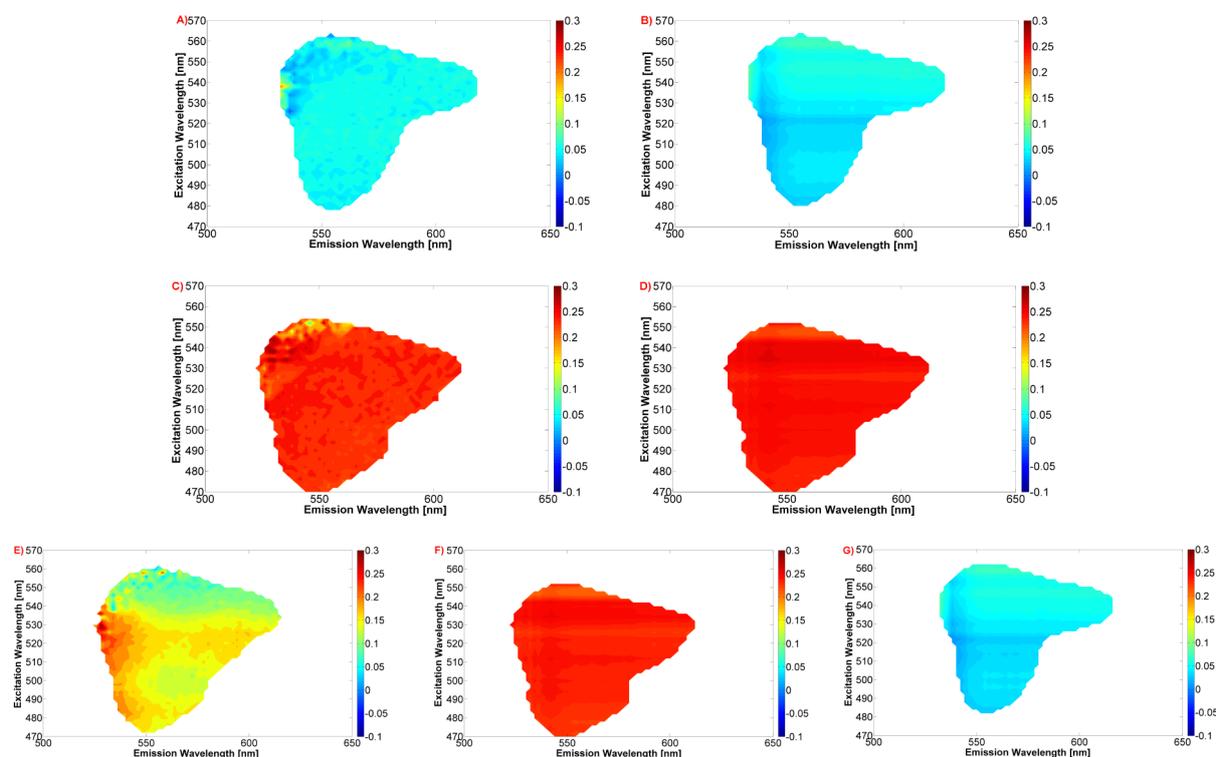


**Figure A- 22: Excitation loadings for each sample (EB/PB) obtained from augmented mode, after resolution of HH dataset (a/b) and VV dataset (c/d). The loadings are shown for the model run with the following constraints applied: (a/c) 1,2,6; (b/d) 1,2,5,6.**

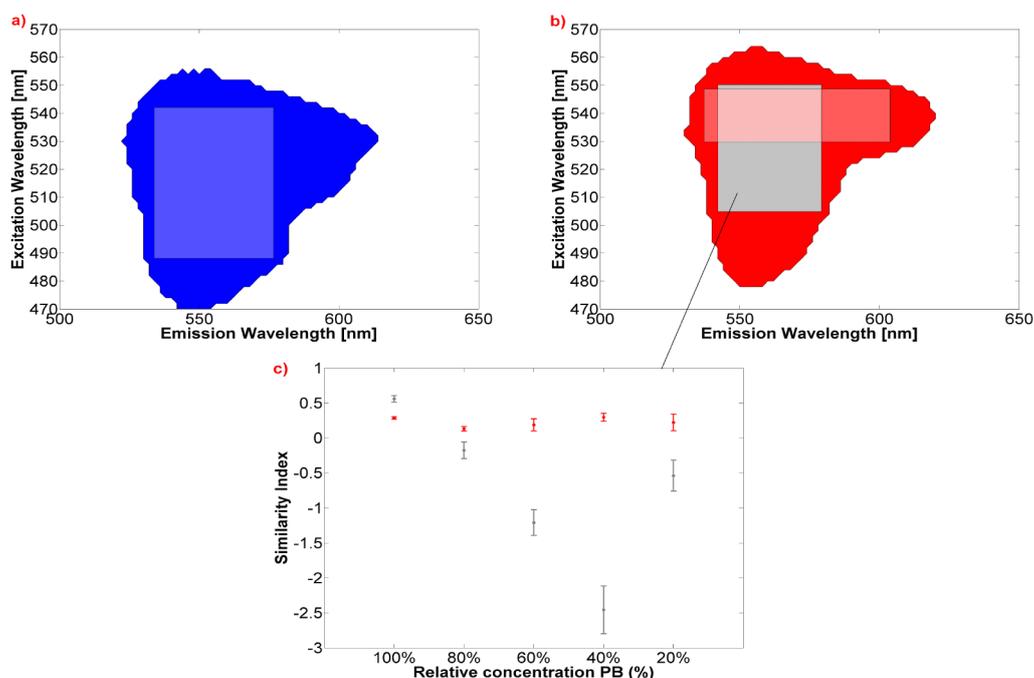


**Figure A- 23: This figure shows the average over all the EEM spectra of EB/PB of: (a) the pure Rayleigh scatter band extracted from each polarization setting; (b) the noise**

resulting from scatter correction in the area free from fluorescence; and (c) the residuals from the model 1,2,5,6 confirming the presence of higher residual scatter in VV-data post-correction.



**Figure A- 24: Comparison of the real and MCR recovered anisotropy plots obtained for: Phloxine B (A/B), and Erythrosin B (C/D), (bottom row) 60% EB/40%PB mixture anisotropy EEM pattern (E) and the recovered Component 1 (F) and 2 (G).**



**Figure A- 25: Areas used to calculate the similarity index between real and extracted anisotropy patterns for: (a) EB, and (b) PB. For PB the grey area led to the similarity index shown in (c).**

**Table A- 8: Table summarizing the composition of each sample, measured at 25°C in HEPES buffer (pH 7.2) using dWGP.**

Samples	Phloxine B (~ 0.6 $\mu$ M)	Erythrosin B (~ 6 $\mu$ M)
1, 2, 3	100%	0%
4, 5, 6	80%	20%
7, 8, 9	60%	40%
10, 11, 12	40%	60%
13, 14, 15	20%	80%
16, 17, 18	0%	100%

**Table A- 9: Estimation of rotational ambiguities for each component from the best EB-PB-EEM-MCR MCR model built using 1: normalization, 2: non-negativity, and 6: correspondence among species constraints/criteria.**

1,2,6	Component 1				Component 2			
	HH	HV	VH	VV	HH	HV	VH	VV
$f_n$ initial	0.463	0.460	0.443	0.613	0.762	0.760	0.780	0.632
$f_n$ max	0.463	0.460	0.443	0.613	0.762	0.760	0.780	0.632
$f_n$ min	0.463	0.460	0.443	0.613	0.762	0.760	0.780	0.632
$f_n$ max- $f_n$ min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

**Table A- 10: Real and modelled (calculated from MCR components) averaged anisotropy values (EB-PB-EEM-126 model), for each mixture, across the  $S_1$  ( $\lambda_{ex/em}$ ) = 520–620/470–570 nm bands.**

	Modelled anisotropy $\pm$ STD	
	Component 1 (EB)	Component 2 (PB)
Pure PB EB solutions, real anisotropy values.	<b>0.243 <math>\pm</math> 0.002</b>	<b>0.054 <math>\pm</math> 0.002</b>
0% / 100%	/	<b>0.048 <math>\pm</math> 0.001</b>
20% / 80%	0.245 $\pm$ 0.006	0.040 $\pm$ 0.001
40% / 60%	0.253 $\pm$ 0.004	0.036 $\pm$ 0.002
60% / 40%	0.240 $\pm$ 0.003	0.033 $\pm$ 0.002
80% / 20%	0.235 $\pm$ 0.002	0.064 $\pm$ 0.006
100% / 0%	<b>0.239 <math>\pm</math> 0.002</b>	/
Avg/std	<b>0.242 <math>\pm</math> 0.007</b>	<b>0.044 <math>\pm</math> 0.012</b>

## 7.10 Some historical and general information about insulin

### 7.10.1 The history behind the discovery of insulin

Insulin was isolated for the first time in 1921 by Frederick Banting, and a 4<sup>th</sup> year medical student called Charles Best, in the laboratory of Prof. McLeod in Toronto University<sup>536-537</sup>.

They were given by Prof MacLeod some lab space and dogs to investigate Banting's idea, who believed that his predecessors did not work correctly in trying to relate pancreas activity with the regulation of blood glucose level. The importance of the pancreas in sugar metabolism was known since 1890, when Minkowski and Von Mering realised that the removal of the entire pancreas in dogs led to severe diabetes and ultimately the dog's death. However, they had no idea that only a small part of the pancreas was responsible for sugar regulation, the Langerhans islets, called after Paul Langerhans who discovered them in 1869<sup>538</sup>. Hypotheses about the role of these islets on glucose regulation were suggested as early as 1894 by Sir Sharpey-Schafer, who also introduced the term "insulin" as being this mysterious substance potentially secreted in the islets, insulin coming from the latin *insula* (island).

The pancreas also contains acinar cells, responsible for the production and storage of digestive enzymes such as trypsin<sup>539</sup>. Upon removal of the entire pancreas, Banting suspected that his predecessors did not take into account that the digestive enzymes left post-extraction could destroy the molecule of interest. When Moses Barron published in 1920 that the acinar cells were dying by atrophy soon after the ligation of pancreatic ducts, but not the Langerhans islets<sup>540</sup>, this gave Banting an idea. In order to reach specifically the substance produced by the islets, his idea was to leave enough time post pancreas ducts ligation for acinar cells to die, so that the islets substance would remain intact. After a few years of experimentation, he managed to isolate the islet secretion, which proved to be efficient in reducing hyperglycemia of diabetic dogs. However, the process to get a degenerated pancreas was long (several weeks) and instead they managed to get the same result from fresh beef pancreases, using alcohol extraction<sup>536</sup>.

With the help of Collip, by careful re-crystallisation from ethanol they isolated insulin for the first time. In 1922, the first human clinical trial was performed on a 14-year-old diabetic boy in Toronto hospital<sup>541</sup>. The first trial failed due to impurities still present in the "formulation", even if a slight decrease in blood glucose was observed. Collip, during the same year, improved the isolation of insulin which became the first successful human diabetic treatment. The discovery of insulin finally brought Banting and MacLeod the Nobel prize for physiology/medicine in 1923.

It is interesting to note that a collaboration with the company Eli Lilly started in 1922, when the team in Toronto faced the problem of insulin mass production. In 1922, Toronto University also received a visit that still has consequences nowadays. The 1920 Nobel prize winner for physiology/medicine Prof Krogh, whose wife suffered from diabetes, had much

interest in the work developed by Banting and MacLeod, and came back to Copenhagen in 1923 with the permission to produce insulin in Scandinavia. This gave birth to the Nordisk Insulin Laboratorium, which became Novo Nordisk, a major Pharma-Chem company. More details and interesting facts about the discovery and controversy associated with insulin can be found in Louis Rosenfield's historical review<sup>538</sup>.

### 7.10.2 Evolution of insulin formulations

Insulin formulation has evolved a lot since the early Collip formulations, which were described as “a murky, light-brown liquid containing much sediment” when first used<sup>538</sup>. After the discovery of insulin, research moved towards trying to prolong the effect of insulin, in order to limit the number of daily injections which was one of the major early drawbacks<sup>517</sup>. This started to be obtained by Hagedorn in the 1930s by addition of protamine in formulation, a protein that helped to keep insulin longer in solution<sup>542-543</sup>. This formulation, characterised as “intermediate acting”, saw its active period improved in the 1940s by the addition of small amount of zinc. Issues of stability, over-dosing effects (risk of hypoglycaemia) and long offsets were solved when Hagedorn mixed stoichiometric ratios of insulin with protamine (also called *isophane*), and a small amount of zinc. This NPH (neutral protamine Hagedorn) crystalline formulation, is still popular nowadays as “intermediate acting” insulin: very flexible and efficient once combined with regular rapid acting insulin for treatment of severe diabetes<sup>544</sup>. Stable, long lasting insulins called *lente* or *ultralente* were also produced in the 1950s in Denmark by Hallas-Mø, by complexing small zinc amounts with neutral insulin in solution<sup>545</sup>.

Up until the 1960s and 70s, Human diabetes treatments (type I and II) used bovine and progressively more and more porcine pancreas extracts<sup>517,546</sup>. The commercial production of human insulin only started in the 80s<sup>547</sup>, when animal sources were not able to satisfy the demand that increased worldwide. Creating new ways to synthesize human insulin became the top priority, as this would solve the demand issue but also help to better control the purity of formulations which was one of the other major patient safety issues<sup>547</sup>. Scientists started to modify porcine insulin (which differs from human insulin by only one amino acid) using Trypsin or *Achromobacter lyticus*, that catalysed the conversion of pig B30 alanine to threonine either through a one-step transpeptidation, or through a two-step reaction involving reverse protease mechanism<sup>548-549</sup>. This type of enzyme catalysis gave birth to the first semi-synthetic human insulin and also the first mono-component insulin formulations from Novo Nordisk with a purity of 99%<sup>550</sup>. The same result can be obtained starting from the insulin precursor (proinsulin) instead of porcine insulin<sup>551-552</sup>.

Semi-synthesis of human insulin started to become interesting when recombinant DNA technology, or genetic engineering appeared. The concept of that method is to introduce a human gene in the genetic material (plasmide) of a simple bacteria, such as *Saccharomyces cerevisiae* (yeast) or *Escherichia coli* (*E. coli*), in order to let the bacteria produce a specific protein encoded by the gene<sup>553</sup>. Insulin was the first molecule destined for use in human treatment ever produced by this technology<sup>554-555</sup>, licensed and marketed by Eli Lilly in 1982<sup>556</sup>. It is generally the proinsulin that is expressed by the bacteria rather than insulin. Yeast for example secretes proinsulin in the supernatant<sup>557</sup>, which is transformed into fully functional human insulin through various procedures<sup>556</sup>. A mixture of proinsulin and insulin was also directly expressed by a transformed yeast cell<sup>558</sup>. *E. coli* and *Saccharomyces cerevisiae* are the main expression systems currently used for insulin production, and the latter is generally preferred for large scale production<sup>556</sup> (which corresponds also to the human insulin used in this thesis). This is different to most of the glycosylated therapeutic proteins, such as monoclonal antibodies that are expressed in mammalian cells like Chinese hamster ovary (CHO) or Baby hamster kidney (BHK)<sup>559-560</sup> because they require a complex machinery for glycosylation that is absent in *E. coli* or too different in yeast to give functional therapeutics<sup>561</sup>. Other alternatives for insulin production through expression of proinsulin in vegetal lines such as tobacco and lettuce chloroplasts<sup>562</sup> or from oilseeds<sup>563</sup> have also been explored, but they hardly compete with *E. coli* or yeast for large scale production in terms of cost, speed of growth and high yield. Advantages and disadvantages of *E. coli* vs yeast are still being discussed for therapeutics production<sup>564</sup>, but the important point is that the development of genetic engineering opened the road towards the production at unlimited quantity of therapeutic proteins<sup>565-566</sup>.

Synthesizing human insulin was unfortunately only one part of the puzzle. Another critical issue was the difference between the insulin activity profile in the patient post injection (intermediate acting as NPH, short acting or *lente*) and the normal biological secretion<sup>547</sup>. The body of a non-diabetic patient is a complex machine, capable of delivering, if needed, excess insulin to counter hyperglycemia episodes but capable also to maintain a constant glucose level in the blood by controlling insulin concentration, which is called the “basal” insulin<sup>565-566</sup>. Formulations on the other hand suffer from poor reproducibility patterns<sup>547</sup>, action peaks that are sometimes too sharp (eventually leading to hypoglycaemia<sup>567-568</sup>), lag times that are not always suited and incompatibilities between formulations making for example the compensation of a long lag time with rapid acting insulin a very difficult task<sup>517</sup>. There is

therefore a big difference in terms of pharmacokinetics, and a stable “basal” insulin representing up to 30% of daily need<sup>565</sup> is essentially what is lacking in most formulations to best mimic real body secretion patterns<sup>567</sup>. With the development of protein/genetic engineering, scientists were able to target specific modifications in insulin structure, in order to optimize pharmacokinetics and best fit patient needs<sup>547</sup>. This is how the first insulin analogues<sup>569</sup> appeared, some of them being very good basal insulins. Insulin analogues are not used in this thesis, therefore not many details are provided but they are important to mention as research is still ongoing to produce pharmacokinetic profiles as close as possible to human secretion patterns. Insulin glargine, the first long-lasting analogue introduced for clinical use in 2000<sup>547</sup>, was a major advancement in diabetes treatment due to the absence of an action peak and a long basal lasting effect<sup>570</sup>. This analogue is obtained through addition of two arginine residues in position B30 and by swapping the A21 arginine with glycine. Insulin detemir was available soon after<sup>571</sup>, and others such as ultra-long lasting Degludec were developed only a couple of years ago<sup>572</sup>. It is worthwhile mentioning that the development of insulin analogues went also in the opposite direction, as very fast acting insulins (active in minutes) are critical in some situations. Lispro or Aspart are the most famous examples<sup>573-574</sup>, and to close the loop, pegylated Lispro was recently published as a basal insulin<sup>575</sup>.

### **7.10.3 The current insulin market**

It was estimated that worldwide in 2015, 415 million adults were diabetic (type 1 or 2) and that this number is expected to grow to over 640 million by 2040<sup>576</sup>. In the US for example, the total cost associated with diabetes treatment in 2012 was over 240 billion US dollars (\$) <sup>577</sup>, representing 20% of total health expenses of the country. A recent study from 2018 estimated a cost of 327 billion dollars in 2017<sup>578</sup> and if the increase follows the trend observed in 2010, 1 adult in every 3 by 2050 will be diabetic in the US <sup>579</sup>. In terms of market, a recent study reported in 2014 that insulin sales worldwide should increase from 12 billion \$ to 32 billion in 2019<sup>580</sup>, and to 54 billion in the next 20 years<sup>556</sup>. In the meantime, it was estimated that the total sales associated with more than 300 biopharmaceutical products exceed 100 billion \$<sup>581</sup>. Insulin therefore represents therefore a significant part of the biopharma industry, but most importantly will be needed by more and more people as the WHO predicts that a diabetes epidemic is expected<sup>322</sup>. The production of insulin will thus continue to increase and the production of larger quantities of insulin in safer ways depends upon the development of better analytical methods and a better understanding of insulin biophysics. In terms of intrinsic

fluorescence, this means starting with the beginning: trying to understand the structure of insulin.

### 7.11 The primary structure elucidation of insulin

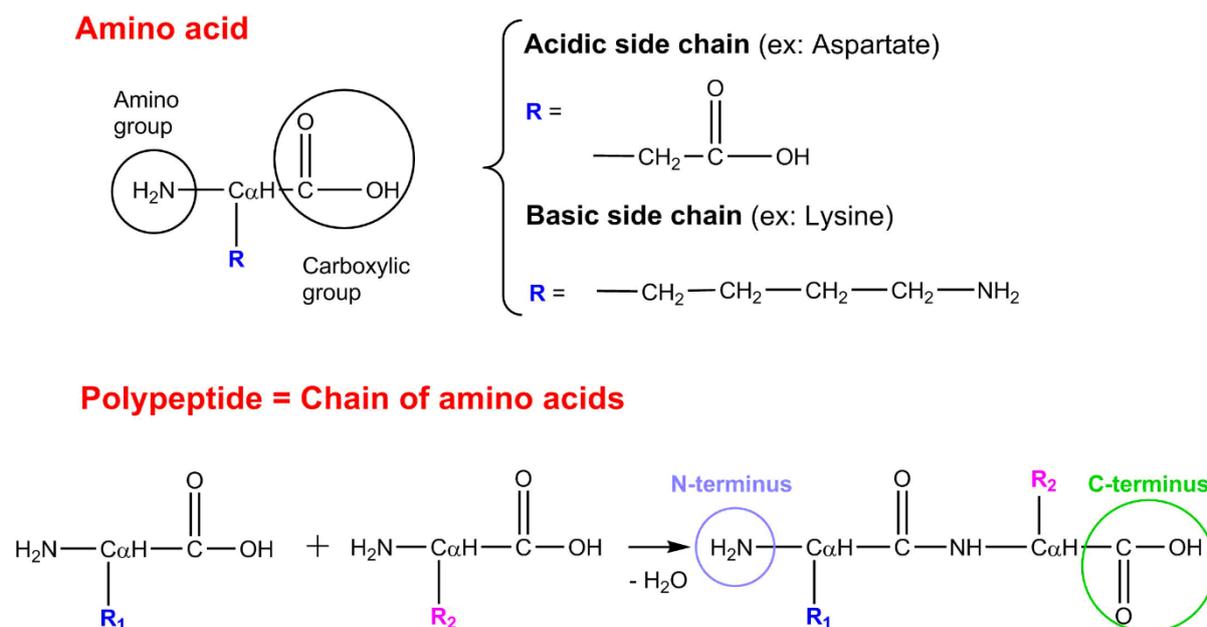
Insulin was first crystallised in 1926 by J. J. Abel *et al.*<sup>582</sup>, and identified as a protein in 1928 by Wintersteiner *et al.* who proved the presence of tyrosine residues and cysteine bridges in insulin<sup>583</sup>. Soon after (1931), insulin's molecular weight was determined by Sjögren and Svedberg through measurement of beef insulin sedimentation constants in ultracentrifugation studies<sup>584</sup>. The molecular weight they obtained of 35,100 Da<sup>aaa</sup> was considered by the authors quite accurate as it was very close to egg albumin value (34,500 Da). However, it is interesting to mention that without realising it, they determined with an impressive accuracy the value for the hexameric form of insulin, known years later to be the form adopted by insulin in crystals. Further studies in the 1940s refined that value to 12,000 Da (Dimer) after noticing that lower concentrations of insulin were affecting the sedimentation values, and this already started to explain the value previously obtained as a result of the association of monomer units in solution<sup>585</sup>. This value was finally determined to be 6500 Da (monomer) in the early 1950s<sup>586</sup>, following Neurath observations that insulin was possibly further dissociated in the correct environment.

The crystallisation of insulin also opened the way towards the determination of protein structure, and the full primary structure (amino-acid sequence) of bovine insulin by Sanger in the early 1950s<sup>587</sup>. His work was not assembled into one big paper, but rather spread in several papers in which he solved bit by bit insulin sequence. With that in mind and referring to what was mentioned above about molecular weight measurements, it is interesting to note that it is through the use of the “Sanger” reagent that Harfenist *et al.* were able to obtain a molecular weight of 6500 Da in 1952. This reagent, 1,2,4-fluorodinitrobenzene (FDNB), was created to bind specifically under mild conditions (therefore avoiding protein degradation) to the N-terminal amino-acids<sup>bbb</sup>.

---

<sup>aaa</sup> Da or Dalton is a molecular weight unit used in protein chemistry, where 1 Da = 1000 g.mol<sup>-1</sup>.

<sup>bbb</sup> Each amino-acid contains an amine (NH<sub>2</sub>) and carboxylic (COOH) group. Through condensation reaction, the amine group of one amino-acid reacts with carboxylic group of next amino acid, to form a polypeptide chain (= protein). Each chain has thus two ends: one where the amine group of a particular amino-acid is unbound (N-terminus or N-terminal amino-acid) and one where the carboxylic group of another amino-acid is unbound (C-terminus). These are terms often used in protein chemistry.



**Figure A- 26: Illustration (drawn with ChemDraw) of the general chemical formula of amino acids and proteins.**

Following complete hydrolysis in 20% HCl, the isolation and identification of the bright yellow DNP-amino acid complexes by partition chromatography allowed him to prove in particular that a phenylalanine and glycine residue were the N-terminal residues of each insulin chains<sup>588</sup>. This method could also be extended to the identification of residues present just after the N-terminal ones, using partial hydrolysis this time<sup>589</sup>. This technique of N-terminal labelling along with complete or partial hydrolysis, and identification of fractionised peptides by partition chromatography, is one example illustrating how Sanger's work brought about many advances, and explains why his work revolutionised the area of protein chemistry and protein sequencing. After splitting up insulin into two fractions (oxidized chain A and B) with performic acid<sup>590</sup>, Sanger using the same sequencing method but through use of different acids or enzymes for partial hydrolysis, determined first in 1951 the complete sequence of the basic<sup>ccc</sup> B chain (or phenylalanyl chain) of 30 residues<sup>591-592</sup>, before solving the acidic A chain (glycyl chain) of 21 residues<sup>593-594</sup>. To finish the study, he also proved the existence of 3 disulfide bridges in position A7<sup>ddd</sup>-B7, A20-B19, and A6-A11<sup>595</sup>. The method employed was slightly different as cysteine bridges were this time preserved during hydrolysis, fractionised one another using paper ionophoresis and oxidised as cysteic acid peptides. By comparing their sequences to the ones of A and B already known, the position of cysteine bridges could be

<sup>ccc</sup> Any polypeptide chain is considered basic or acidic, depending if most amino-acids that compose it have an acidic or a basic side-chain R (see Figure 1 for illustration).

<sup>ddd</sup> A7 refers here to residue n<sup>o</sup>7 in chain A, knowing that numbering always start from N-terminus residue.

deduced. This work ended up with the first complete protein sequence characterization, awarding him with his first Chemistry Nobel Prize in 1958<sup>eee</sup>. More importantly, Sanger brought this revolutionary view of proteins being well defined chemical structures<sup>587</sup>.

## 7.12 Historical details behind the first elucidation of insulin's crystal structure

Hodgkin (nee Crowfoot) was one of the pioneers in x-ray crystallography and contributed significantly to the elucidation of important biological molecule structures<sup>596</sup> (Nobel Prize in 1964). She started to work on insulin in the 1930s, characterising single insulin crystals through x-ray photographs<sup>597</sup>. This work directly followed Scott's discovery, who realised while trying to improve Abel's crystallisation method in 1934, that the presence of zinc (Zn) or other divalent cations was crucial for the appearance of crystals<sup>598</sup>. This allowed Hodgkin to grow single zinc-insulin crystals and to characterize their shape as rhombohedral. The 1969 study confirmed with a 3D structure something that was already proven<sup>376,599-600</sup> *i.e.* that pig insulin naturally crystallises as a hexamer formed from three dimers coordinated around two Zn atoms. A quaternary structure (arrangement of multiple folded proteins such as hexamer) was thus resolved before even looking at the tertiary or secondary structure of monomeric insulin. The electron density map generated from x-ray measurements enabled a resolution at 2.8 Å with some atomic positions (residues A12 to A18) refined in 1971<sup>601</sup>. Very similar results (2.5 Å) were obtained independently in China by the *Beijing insulin group* only two years after Hodgkin<sup>602</sup>. Fruitful collaborations between UK and China emerged<sup>603-605</sup> and the insulin structure continued to be refined over the years, starting with Blundell *et al.* in 1972 who improved the resolution to 1.9 Å<sup>326</sup>, structure also refined by the Chinese group in 1974<sup>606</sup> (1.8 Å) for example.

## 7.13 Insulin: a molecule with similar crystal and solution structures

Insulin is transported and is physiologically active as monomer, and the determination of insulin structure in this specific state, ideally in solution but also as a crystal, is critical. The big issue with monomeric insulin is that crystal structures cannot be obtained due to the automatic self-assembly that occurs at the concentrations needed for crystallisation<sup>329,350</sup>. The

---

<sup>eee</sup> After studying insulin, Sanger started to think that the genetic material behind the creation of such ordered chemical structures as proteins must also have some kind of code. Using electrophoresis, he developed the first method to sequence nucleic acids in RNA and DNA awarding him with a second Chemistry Nobel Prize in 1980.

only way to get a crystal form of an active form of monomeric insulin was obtained by the enzymatic removal of the last five residues of the B chain (B26 to B30)<sup>350,607-608</sup>, which are normally involved in the formation of the hydrophobic  $\beta$ -sheet structure responsible for dimerization. The structure of that des-(B26-B30)-insulin or despentapeptide insulin (DPI) obtained by X-ray crystallography proved however to be remarkably similar to the molecule 2<sup>fff</sup> (UK convention) conformation from hexamer crystals, confirming that crystal packing forces did not alter the monomer folding in any important way<sup>350</sup>. This was also proven with molecular dynamics simulations<sup>609</sup>.

This work on DPI helped to get closer to the real goal, *i.e.* solving the structure of monomeric insulin in solution, as insulin is secreted, transported and interacts with receptor in solution. Indeed, aggregation caused not only issues in obtaining monomeric crystals, but also complicated high resolution NMR studies of monomer structure in solution<sup>357,610</sup>. The propensity of the molecule to self-associate in water caused broadening of <sup>1</sup>H-NMR signals<sup>611</sup>, thus early NMR studies were not well resolved<sup>612-613</sup>. Working at acidic pH<sup>402,611,614</sup> (pH 2-3) helped to partially overcome that issue as it largely reduces the presence of high-order aggregates (*i.e.* hexamers) promoting smaller aggregates like dimers. Using selective mutation, insulin could be induced to remain monomeric at concentrations suitable for solution structure determination by NMR. DPI (1 mM) at pH 1.9 for example was found to have a conformation in solution highly similar to molecule 2 in 2-Zn porcine crystal<sup>336,615</sup>, even though high flexibility seems to be observed at N-terminal ends of A and B chains and C-terminal end of B chain. Other site-directed mutagenesis such as (B16-Tyr--> His)<sup>423</sup> also give biologically active monomer at mM concentration in aqueous solution (pH 2.5). NMR and circular dichroism (CD) spectroscopies confirmed once again the high similarity with molecule 2 conformation, even if flexibility was only observed at the very end of N-terminal B chain.

The situation gets a bit more complex when acetic acid is added to solution. Weiss *et al.*<sup>611</sup> were first to prove with CD and UV difference spectra that at pH 2 in 20% acetic acid, non-mutated insulin existed as a monomer, even in the ~mM concentration range. Through <sup>1</sup>H-NMR data and site-specific mutations, they could prove that a native-like structure highly similar to the crystal structure was also preserved for non-modified insulin<sup>611</sup>. Further proof of the non-denaturation and presence of a stable folded monomeric state in acetic acid was shown

---

<sup>fff</sup> Many papers use the Chinese convention where similarity to molecule 1 is reported. We decided to use UK convention for the thesis.

in several studies<sup>358,616</sup>. However, the big difference was that the molecule 1 conformation seemed to be preferred as suggested by the absence of NOEs<sup>§§§</sup> between PheB25 and TyrA19. Even if authors admitted some reservations due to the partial disorder observed on the side chain of PheB25 revealing high flexibility in solution, the adoption of molecule 1 conformation was confirmed by further study<sup>617</sup> and also suggested by analysis of DPI at pH 3 in 20% acetic acid<sup>618</sup>. It was also suggested to be the form adopted by insulin prior to interaction with the receptor<sup>619-620</sup>. The PheB25 ring by pointing outwards, can more easily bind to the hydrophobic pocket of the receptor, establishing one of the primary and most important contacts in insulin recognition. Following this, joint protein/receptor arrangements involving in particular the separation of C-terminal B chain to N-terminal A chain<sup>359,621</sup>, create secondary binding events by exposure of other hydrophobic residues that make insulin fully biologically active. Nevertheless, this illustrates the debate present over the monomer conformation in solution. Chang *et al.*<sup>622</sup> highlighted that issue and proposed the hypothesis that in 20% acetic acid, hydrophobic interactions between the phenyl ring of PheB25 and the organic solvent stabilize the residue in the outward position.

A similar controversy is also present for dimers. Jorgensen *et al.*<sup>623</sup> published interesting results on B9 (Asp) mutant of human insulin. The B9 mutation enabled a serious reduction in self-association so that only dimers existed in solution at pH 1.9. They determined the solution structure with <sup>1</sup>H NMR and found that the monomer units adopted molecule 2 configuration, with only minor deviations from the crystal structure<sup>623</sup>. Conversely, however, molecular dynamics in water showed the insulin dimers to be asymmetric<sup>624</sup> with neither of the conformations being favoured in solution, an idea also supported by Weiss *et al.*<sup>341,625</sup>. In that regard, it seems to be identical to the dimers encountered in hexamers crystal structures (Figure 45). The arrangement of dimers does not seem to change either if hexamers are in solution<sup>622</sup>, illustrating the very high similarity observed between Zn complex in solid or liquid phase<sup>349</sup>.

## **7.14 The T/R transition: impacts on insulin hexamer structure and biological relevance**

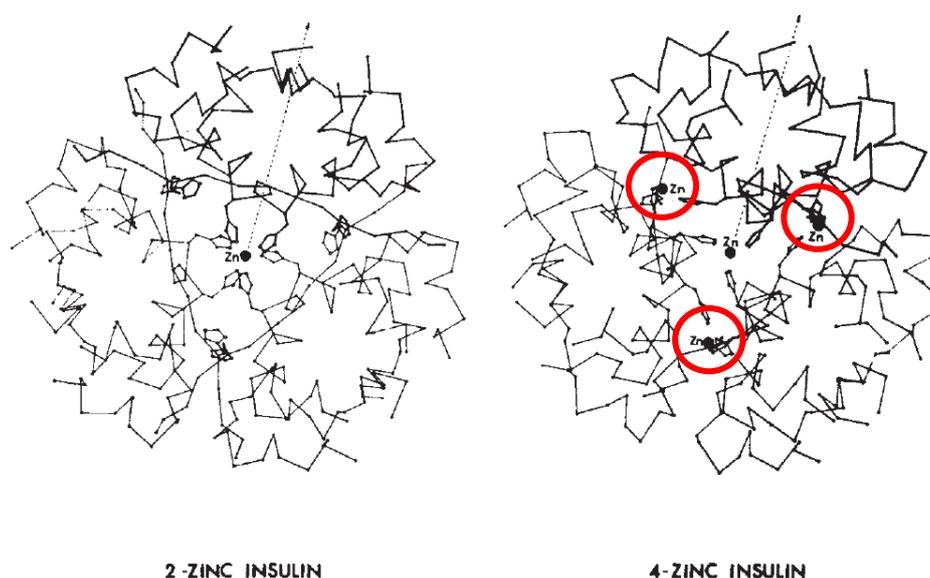
Insulin monomer can undergo a transition from the T to the R state, and this transition is linked to the existence of various forms of insulin hexamers. We saw that the 2-Zn form, also

---

<sup>§§§</sup> NOEs : Nuclear Overhauser Effect, referring to the relaxation that occurs between a nucleus excited magnetically and its neighbour. This effect is used to estimate distances between nuclei as it increases with shorter distances.

called  $T_6$ , was the first one discovered. But a 4-Zn structure was soon after discovered. This structure was named  $T_3R_3$  and exist in reversible equilibrium with the  $T_6$  form<sup>387</sup>.

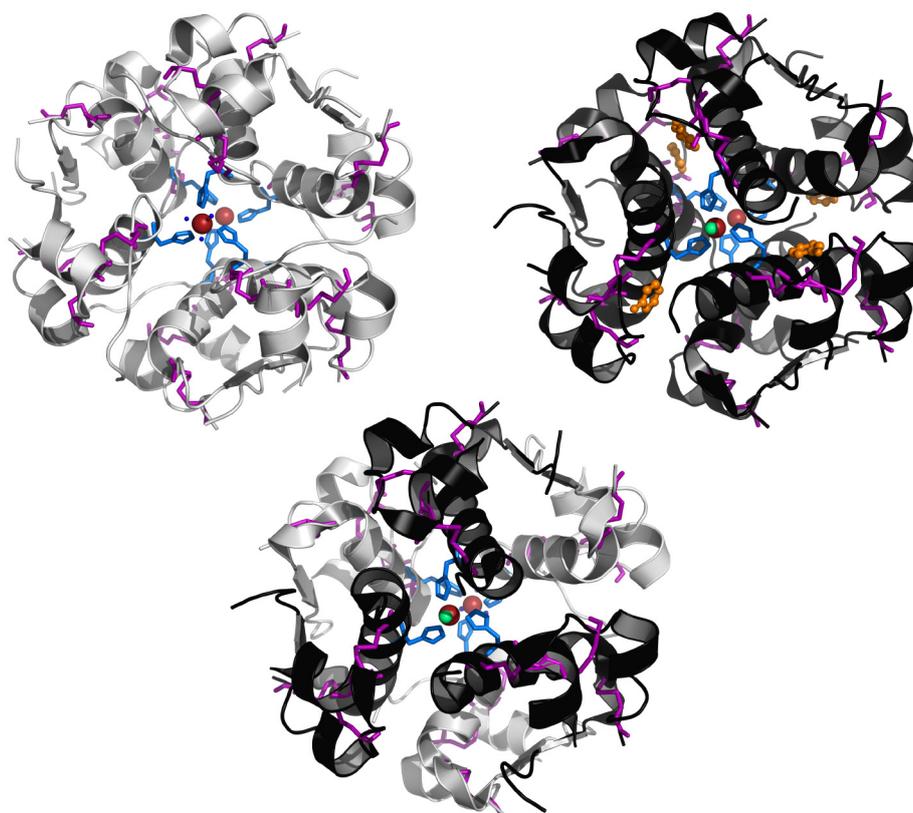
4-Zn means that a maximum of 4 metal ions can be present in the hexamer. In addition to the position along the threefold axis located between the three monomers sharing T configuration, three additional off-axial binding sites can appear in the cavities created by the  $T \rightarrow R$  transition<sup>378,384</sup>. In one dimer, the HisB10 imidazole ring by rotation of its side chain turns towards the HisB5 of another dimer, creating these solvent shielded binding sites (Figure A- 27). Due to the restricted space in the cavities, a tetrahedral rather than octahedral coordination is adopted by  $Zn^{2+}$  and completed by two molecules of water<sup>378</sup>.



**Figure A- 27: Comparison of the 2-zinc insulin and 4-zinc insulin hexamers viewed down the threefold axis. Only  $\alpha$ -carbon atoms, histidine side chains and zinc ions are shown. For clarity, the solvent shielded sites were circled in red. Reproduced from Bentley<sup>378</sup> *et al.*, with permission from *Springer Nature*.**

The presence of zinc in off-axial positions was originally thought to increase structural stability however, the off-axial binding sites seemed to not be always fully occupied<sup>384</sup>. At lower zinc concentrations for example, the  $T_3R_3$  structure remained but arranged around only two Zn ions<sup>383</sup> (shown in Figure A- 28). The cavities were filled with clusters of water molecules and both Zn ions occupied positions along the threefold axis. The main difference with the 2-Zn  $T_6$  structure is that one Zn ion is buried at the bottom of a narrow tunnel engendered by the R arrangement. This  $Zn^{2+}$  becomes completely isolated from the solvent<sup>383,626</sup>, thus less prone to dissociation. Re-examination of the 4-Zn structure (suggested in reality to be 3 Zn)<sup>384</sup> also show this site being preferentially filled compared to the off-axial

positions. This suggests that the buried site might be the real cause for that molecule's extra-stability, explaining as well why industry used it as a long-acting insulin. At the bottom of that tunnel, the space around that ion is highly restricted<sup>627</sup> imposing a change in the coordination geometry from octahedral to tetrahedral<sup>626</sup>. The orientation of the HisB10 side chain is preserved and a single exogenous anion (in this case Cl<sup>-</sup>) replaces the three water molecules. In any case, it seems that whether the 2 or 4 Zn T<sub>3</sub>R<sub>3</sub> is considered, both hexamers can be distinguished by their faces (trimers): one face where an octahedral coordination is conserved and arranged around three T monomers, and one face where the geometry is tetrahedral and arranged around three R monomers. In the 1990s, Derewanda *et al.*<sup>628</sup> realised that a third hexamer conformation existed when phenol was added to the solution. The hydrophobic pockets exposed by T → R transition of the dimer-dimer interface<sup>626</sup> are large enough to accommodate a wide variety of small organic molecules like phenol, substituted phenol, or naphthalenes<sup>626,629-631</sup>. With different effectiveness<sup>630</sup>, they all stabilise the six monomers in the R conformation through H-bonds with CysA6 and CysA11<sup>628</sup> and van der Waals interactions, forming the R<sub>6</sub> hexamer (Figure A- 28).



**Figure A- 28:** Representation of T<sub>6</sub> (top left), T<sub>3</sub>R<sub>3</sub> (bottom), R<sub>6</sub> (top right) 2-Zn hexamer insulin, using respectively 4INS, 1TRZ and 1ZNJ from the PDB databank. The monomers in T configurations are plotted in light grey, the R states in black. Zn, Cl and H<sub>2</sub>O are shown respectively as red-brick, green, and tiny blue spheres. The six phenol rings are shown in the R<sub>6</sub> structure as orange sticks.

Because the 2-Zn hexamer conformational changes are ligand mediated, the molecule was qualified of allosteric<sup>hhh</sup> protein<sup>385</sup>. The positive homotropic effect<sup>626</sup> engendered by these neutral organic molecules, act in tandem with the binding of monovalent anions to tetrahedral Zn<sup>2+</sup>, strongly promoting the T<sub>6</sub> → R<sub>6</sub> transition through stabilisation of the R<sub>6</sub> state<sup>632-633</sup>. Similar to the organic ligands, a variety of monovalent anions<sup>627,634</sup> can also play that role, and some (*e.g.* SCN<sup>-</sup>) are much more effective to than others (*e.g.* Cl<sup>-</sup>)<sup>631</sup>. In the end the T<sub>3</sub>R<sub>3</sub> state is only an intermediate in the T<sub>6</sub> → R<sub>6</sub> transition<sup>627</sup>, and can be isolated if phenol concentration is low enough to avoid complete transition<sup>635</sup>.

The biologic relevance of this T/R transition is still a matter of speculation, especially when related to insulin-receptor binding. It is known that a conformational switch occurs upon binding<sup>359,362</sup>; with the T state being inactive<sup>341,366</sup>, the R state was naturally suspected to be potentially the one binding. Research has since proven that it is unlikely to be the case<sup>388,636</sup>, and while the mechanism of binding remains unresolved, there is an agreement that: (*i*) the T/R transition is probably uncoupled from receptor binding<sup>365,386</sup>; (*ii*) the R-state is less stable but more active than T state and may have some features that favour binding to the receptor<sup>362,386,637</sup>; (*iii*) the T/R transition at least underlines the flexible sites probably crucial for binding (*e.g.* critical hinge point at GlyB8<sup>386,388</sup>). The inherent flexibility of the insulin molecule is thus a key factor for receptor recognition, and a detachment model of B-chain C-terminal region (B21-B30) caused by rearrangement and resulting in exposure of hydrophobic residues is believed to form the surface engaged in receptor binding<sup>362,533-534</sup>.

Even if at first glance information about insulin-receptor complexation can appear out of context in this chapter, its relevance is high. As explained in more detail in section 5.2.2, many of insulin fluorophores that participate in self-assembly are also suspected to be engaged in receptor binding. Therefore, being able to successfully resolve insulin intrinsic emission into its constituent emitters could become an interesting tool to monitor this kind of unresolved and complex association.

Also from a biopharmaceutical point of view, insulin is very often formulated as the hexamer<sup>448,517,636</sup>; various additives such as phenol/metacresol or chloride/acetate ions are often employed for their respective antimicrobial and isotonic properties, were also found to increase

---

<sup>hhh</sup> An allosteric effect is observed when a ligand (other than the primary ligand) binds to a protein in sites other than the main active site and causes a modification in the conformation/activity as well as the binding properties of the protein. If this initial binding increases the affinity for ligand binding in other sites, then the allosteric effect is said positive. Also, if all the ligands involved are the same compound, the allosteric effect is homotropic.

the formulation shelf-life, stabilising in fact the Zn-buried R<sub>6</sub> state<sup>74,340,410,638</sup>. Monitoring the allosteric transition from T<sub>6</sub> to R<sub>6</sub> with ARMES could thus be an interesting study to do in the future, especially in response to Maltesen's<sup>639</sup> comment that clearly states the difficulty of characterising spectroscopically this transition in solution due to the "lack of intrinsic chromophoric properties of the insulin hexamer".

### 7.15 Details about Tryptophan steady-state fluorescence

Trp fluorescence is due to an aromatic indole group and is strongly influenced by its microenvironment, with emission maxima that can shift by ~50 nm going from non-polar to polar environments. These complex emission properties are believed by many to be due to the presence of nearly isoenergetic excited states <sup>1</sup>L<sub>a</sub> and <sup>1</sup>L<sub>b</sub>. One argument in favour of this theory was the resolution under cryogenic conditions of the excitation anisotropy spectrum of Trp, that proved that both electronic transitions were overlapping in the 250-300 nm absorption band<sup>640,641,457</sup>. Associated to the near orthogonality of their dipoles (Figure 52), it also enabled to explain the complex/non-constant anisotropy observed for that band.

Depending on the solvent, either excited state can be favoured to become the main contributor to emission. In completely nonpolar environments *e.g.* 100% cyclohexane<sup>642</sup> or in proteins like azurin with very unusual environments for Trp (completely buried and yet not hydrogen bonded)<sup>643</sup>, emission is structured, blue shifted ( $\lambda_{em} = 308$  nm) and is believed to arise mainly from <sup>1</sup>L<sub>b</sub> (Figure 52). In polar solvents however *e.g.* 100% ethanol<sup>642</sup> or for glucagon where Trp is fully exposed<sup>20</sup>, <sup>1</sup>L<sub>a</sub> is believed to dominate the fluorescence with an emission band that becomes broad, unstructured and that can be red shifted up to  $\lambda_{em} = 354$  nm.

There are different reasons explaining the inversion of emission dipoles with solvent polarity. One is the direct involvement of the nitrogen atom of pyrrole ring in the <sup>1</sup>L<sub>a</sub> transition, increasing the sensitivity of that dipole to hydrogen bonding which stabilises it at energy levels lower than <sup>1</sup>L<sub>b</sub><sup>28</sup>. Another reason is that <sup>1</sup>L<sub>a</sub> has a larger excited state dipole moment than <sup>1</sup>L<sub>b</sub><sup>644</sup>, resulting in shifts to longer wavelength in polar solvents. Also upon excitation of <sup>1</sup>L<sub>a</sub>, the electron density that migrates from the pyrrole to the benzene ring is highly sensitive to the charged entities (*i.e.* electrical field of protein/solvent) surrounding the dipole, and can cause the emission to shift in the red<sup>645</sup>.

There is consensus that <sup>1</sup>L<sub>a</sub> is the Trp fluorescing state in (almost) all proteins<sup>645</sup>, which makes sense considering the high influence that even trace quantities of hydrogen bonding solvent on Trp fluorescence can have<sup>642</sup>, and the fact that environments favouring <sup>1</sup>L<sub>b</sub> are very

rare. What is particularly interesting is that the degree of red shift can be roughly correlated with the degree of Trp solvent exposure in proteins (Figure 52), generating discrete classes of Trp<sup>134,646</sup> that have specific  $\lambda_{\text{ex}} / \lambda_{\text{em}}$  values but also quantum yields and lifetime<sup>647,34</sup> (Table A-11). Trp intrinsic fluorescence displays thus a unique sensitivity to solvent polarity and/or microenvironment, explaining all the attention it received. Applications are numerous with the study of protein folding and conformational states<sup>648-649</sup>, protein dynamics<sup>650-651</sup>, intermolecular interactions<sup>652-653</sup>, interaction protein-membrane<sup>654</sup> etc.

**Table A- 11: Fluorescent characteristics of the most frequent Trp classes, as defined by Burnstein *et al.*<sup>134,646</sup>. Class A and S are missing and correspond to Trp with structured emission, emitting at 308 and 316 nm but encountered in very rare cases. Class I: buried low-polar region; Class II: limited solvent exposure; Class III: totally exposed.**

Trp classes	Lifetime (ns)	Quantum yield	$\lambda_{\text{em}}$ (nm)	Stokes shift (nm)
Class I	2.1	0.11	330-332	48-50
Class II	4.4	0.3	340-342	53-55
Class III	5.4	0.2	350-353	59-61

## 7.16 Certificate analysis of Human insulin lot used for this thesis

**SIGMA-ALDRICH®**

[sigma-aldrich.com](http://sigma-aldrich.com)

3050 Spruce Street, Saint Louis, MO 63103, USA

Website: [www.sigmaaldrich.com](http://www.sigmaaldrich.com)

Email USA: [techserv@sial.com](mailto:techserv@sial.com)

Outside USA: [eurtechserv@sial.com](mailto:eurtechserv@sial.com)

### Certificate of Analysis

Product Name:  
Insulin human - meets USP testing specifications

Product Number: 10908  
 Batch Number: SLBP5090V  
 Brand: SIGMA  
 CAS Number: 11061-68-0  
 MDL Number: MFCD00131380  
 Formula: C257H383N65O77S6  
 Formula Weight: 5,807.57 g/mol  
 Storage Temperature: Store at -20 °C  
 Quality Release Date: 02 NOV 2015  
 Recommended Retest Date: AUG 2018

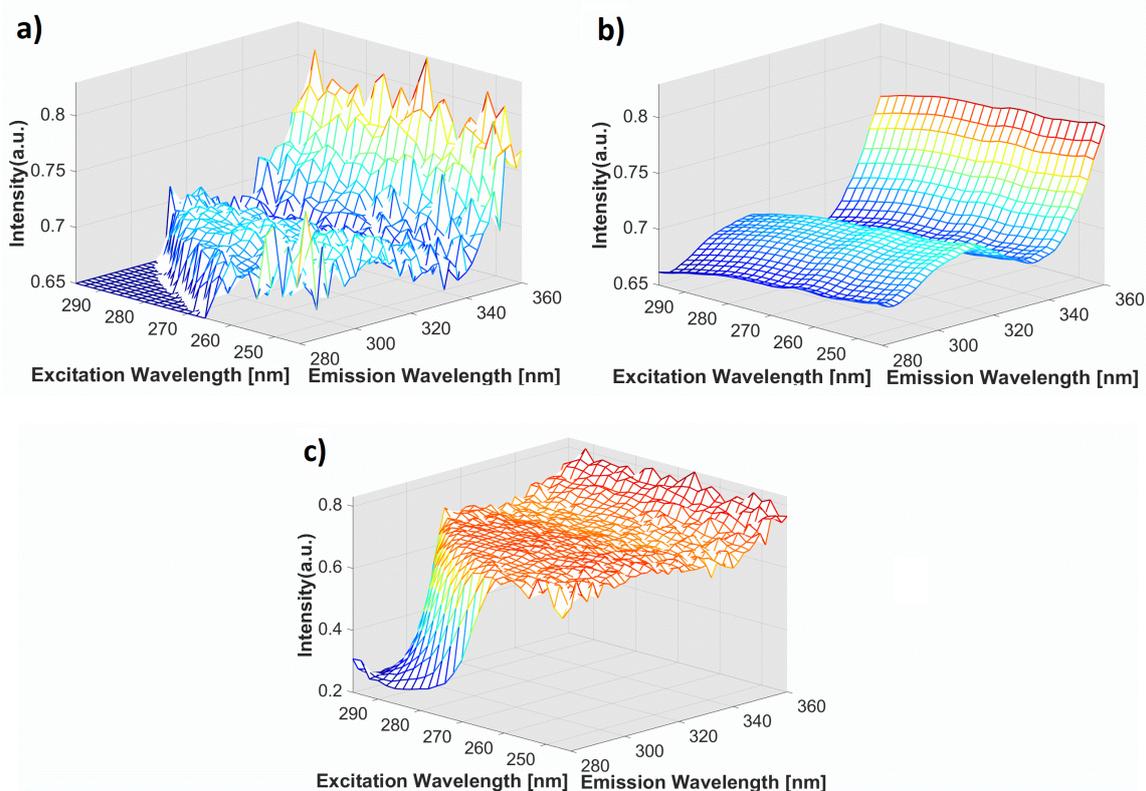
Test	Specification	Result
Identity	Pass	Pass
Total Microbial Count	≤ 300	< 10
Microbial Enumeration tests and Absence of Specified Microorganism; Expressed in CFU/g		

Test	Specification	Result
Endotoxin Level	$\leq 10$ EU/mg	$< 0$ EU/mg
USP Endotoxins units/mg		
Loss on Drying	$\leq 10.0$ %	5.0 %
Zinc (Zn)	$\leq 1.0$ %	0.4 %
Dry basis		
Limit Test	$\leq 1.0$ %	0.2 %
High molecular weight proteins		
Residual Solvents USP 467	Meets Requirements	Meets Requirements
Two class 2 solvents (meets option 1 limit) and Three class 3 solvents.		
Related Compound	$\leq 2.0$ %	0.9 %
Host Cell Derived Proteins	Pass	Pass
$< \text{ or } = 10\text{ppm}$		
Assay	$\geq 27.5$ UN/mg	28.7 UN/mg
dry basis		
USP Insulin Human Units/mg		

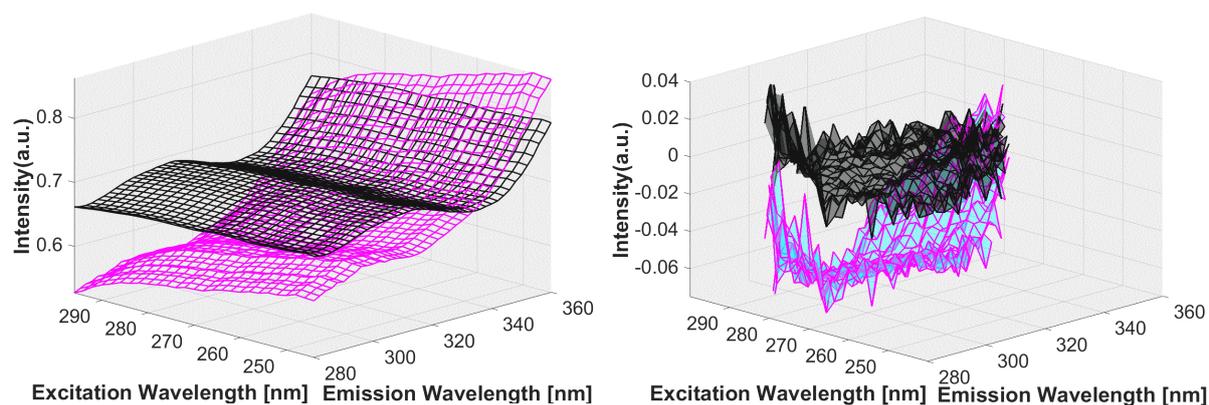
Product Number: I0908  
Batch Number: SLBP5090V

*Robyn Buelach*

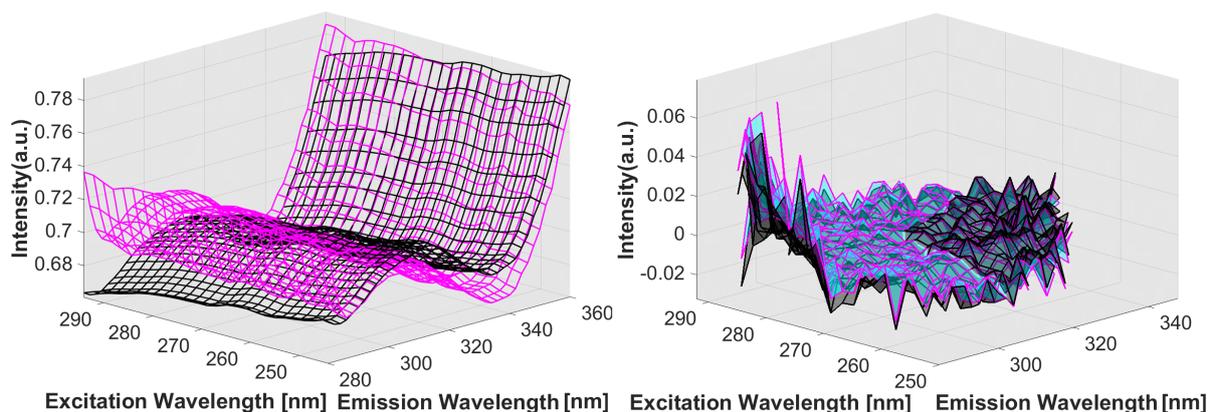
### 7.17 Details for the G factor denoising



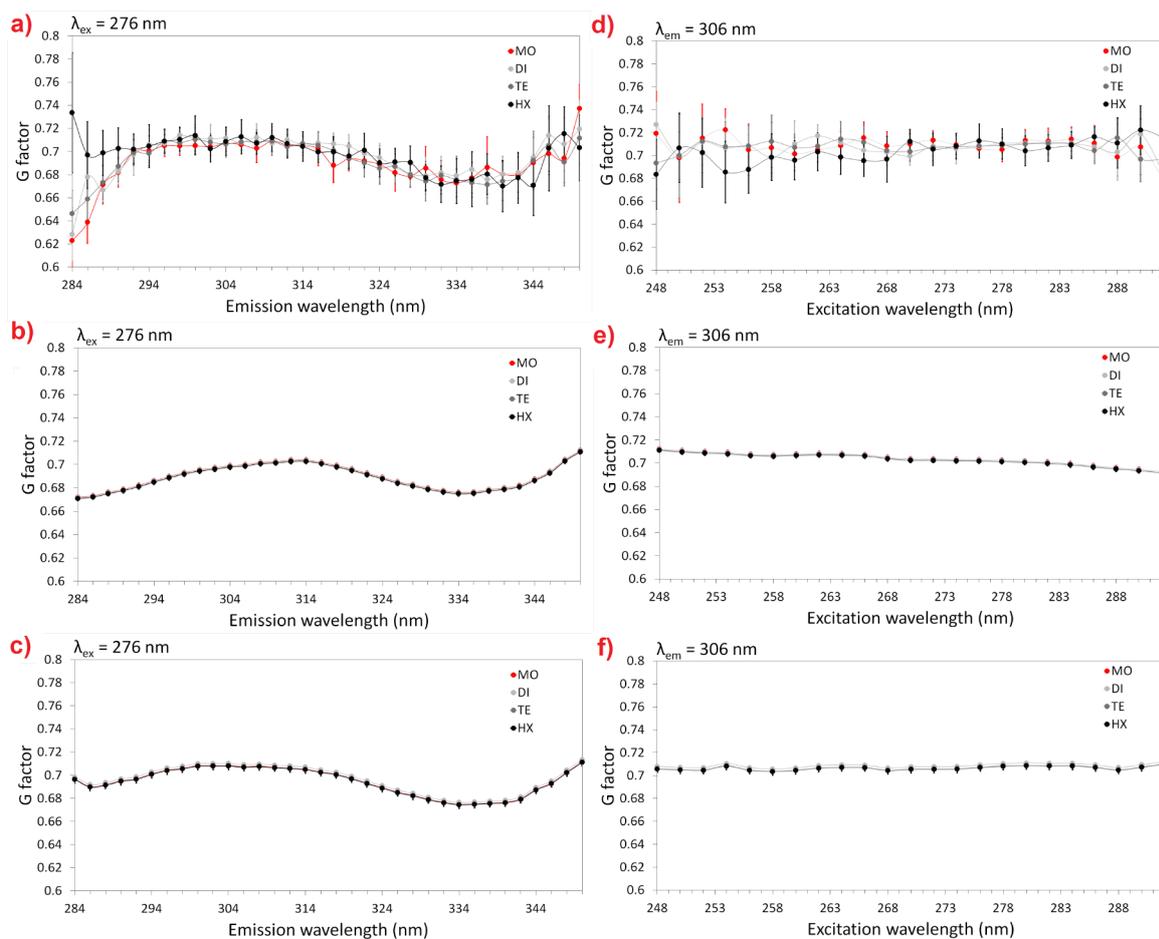
**Figure A- 29: G factor shown prior (a) and post (b) denoising using a one component PARAFAC modelling approach. The values below 0.65 found along the Rayleigh band (c), were replaced by 0.65 prior modelling to reduce their negative impact on G shape. All plots were obtained by averaging the data over 12 monomer replicates.**



**Figure A- 30:** (left) Overlap between the G factor modelled by PARAFAC using either in pink non-optimised parameters (entire  $\lambda_{\text{ex/em}}$  region; G values unchanged along Rayleigh band) or in black optimised parameters (selected fluorescence region; G values replaced by 0.65 along Rayleigh band). Both were plotted on same region to clearly show the change in the denoised G shape. Their corresponding impact on *aniso*-MDF (right) is shown by plotting the difference spectrum between the raw *aniso*-MDF and the *aniso* obtained from either of the denoised G. All plots were obtained by averaging the data over 12 monomer replicates.



**Figure A- 31:** (left) Overlap between the G factor calculated from the PARAFAC-denoised  $\text{EEM}_{\text{HH}}$  and  $\text{EEM}_{\text{HV}}$  spectra (pink) and the G factor directly modelled by PARAFAC (black) using optimised parameters. Their corresponding impact on *aniso*-MDF (right) is shown by plotting the difference spectrum between the raw *aniso*-MDF and the *aniso* obtained from either of the denoised G. All plots were obtained by averaging the data over 12 monomer replicates.



**Figure A- 32: G factor calculated for each insulin form, showing the average/STD obtained over the 12 replicates along the emission ( $\lambda_{\text{ex}} = 276$  nm, a/b/c) and excitation ( $\lambda_{\text{em}} = 306$  nm, d/e/f). G was calculated either from raw  $\text{EEM}_{\text{HH}}$  and  $\text{EEM}_{\text{HV}}$  (a/d), was modelled by PARAFAC as one non-negative component (b/e) or was calculated from the PARAFAC-denoised  $\text{EEM}_{\text{HH}}$  and  $\text{EEM}_{\text{HV}}$  (c/f). This graph has the merit to show the stability of the G factor along the entire data collection once noise is removed (c/f versus a/d). It also shows the superiority of c/f versus b/e to best resemble the a/d spectra and thus best denoise the G factor.**

## 7.18 PARAFAC models on the L-Tyr data

In order to investigate if the spectra obtained from measuring free L-Tyr in various buffers showed any spectral shifts, PARAFAC models were built using the same conditions as the insulin models with (i) blank subtraction ensuring the removal of the non-trilinear Raman scattering; (ii) correction of the 1<sup>st</sup> order Rayleigh scatter via modelling as used in previous chapter; and (iii) minimal smoothing.

The  $\lambda_{\text{ex/em}}$  range was reduced to  $\lambda_{\text{ex}} = 244\text{-}296$  nm and  $\lambda_{\text{em}} = 280\text{-}360$  nm in order to remove regions with no signal or where residual scatter not overlapping with the fluorescence is present. This then produces a three-way array of size 12 samples  $\times$  41  $\lambda_{\text{em}}$   $\times$  27  $\lambda_{\text{ex}}$ . All models

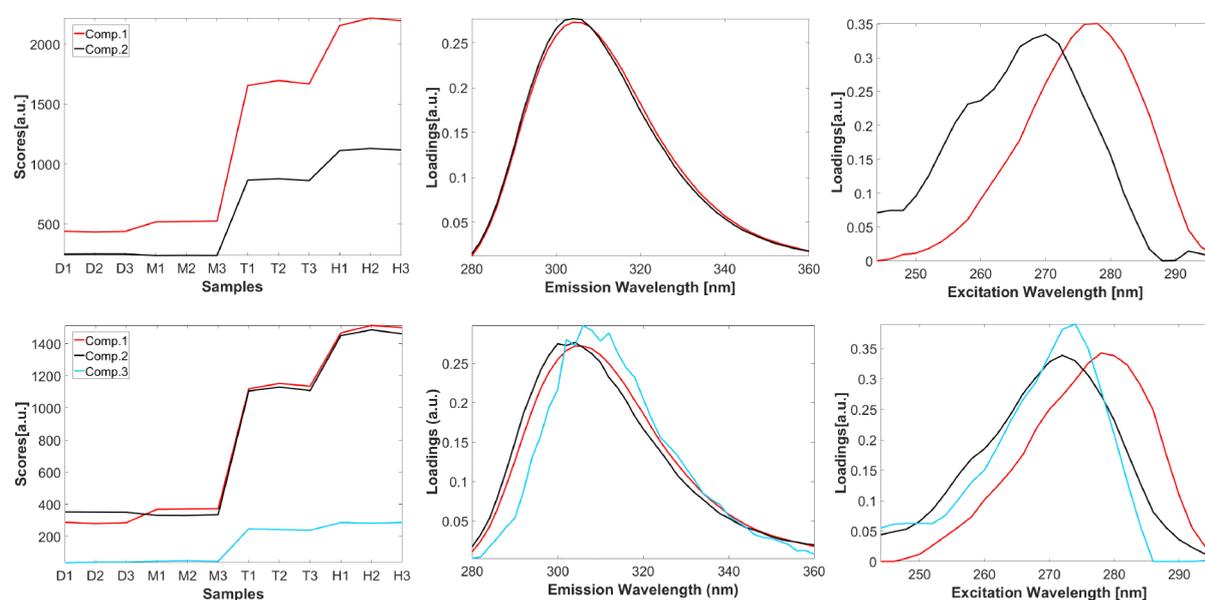
were run with non-negativity in all modes in order to obtain meaningful solutions. The convergence criterion was set so that the relative difference in fit between two successive iterations was below  $10^{-10}$ . This might be a bit excessive for this dataset but might not be for insulin data, as a lower threshold value than  $10^{-6}$  is recommended for difficult data<sup>655</sup>. Initial estimates were obtained from the best fit model obtained out of 10 small runs (a few iterations each) as proposed by Tomasi and Bro<sup>656</sup>, which is also the recommended initialization method for difficult models where *i.e.* many factors are present, or where loadings are highly collinear, or for models presenting local minima in their solutions. In order to check for the latter but also to look at the reproducibility of the modelling, each model was repeated 10 times and the best one (best fit, higher core consistency) was selected for discussion.

One of the critical aspects in chemometrics and especially in factor-based methods like PARAFAC or MCR, is to find the right quantity of meaningful components to best fit and represent the dataset under study. One of the key aspects to meaningful factors is the use of appropriate constraints, but because there is no limit to the quantity of components that can be fitted to a dataset, there is always the risk to over or under fit the data. It can sometimes be easy to detect overfitting if *e.g.* the components extracted do not make physico-chemical sense, or under-fitting by looking at the residuals. One major rule is of course to know the chemical system being studied and use that information to have an idea of the quantity of components to extract. However even with the knowledge of the system, the choice of the appropriate quantity of components can be hard. The core consistency or CORCONDIA<sup>657</sup> (CC) is a clear-cut diagnostic method developed by Bro *et al.*, that enables to find out, at least for well-behaved trilinear data, if the right quantity of components were selected to best model the data. CC generally stays close to 100% and decreases smoothly with an increasing number of real components, until a sudden drop is observed at  $n$  components. This drop indicates that components start to span non-trilinear variation in the data, making the trilinear PARAFAC model inappropriate. The CC method says that with  $n-1$  factors, the PARAFAC model will be appropriate with factors that cover only trilinear variation in the data, but also describe the data with the optimised number of components as fewer factors might fail in describing all the low-rank trilinear variation<sup>657</sup>.

The normalization is not a pre-processing that is often used for the analysis of EEM data with PARAFAC. A first series of models with non-normalized data were built using the EEM<sub>||</sub> of L-Tyr collected across the different buffers (Table A- 12, Figure A- 33).

**Table A- 12: Summary of the fit (or variance<sup>iii</sup>) captured by each component for the non-normalized EEM<sub>||</sub> PARAFAC models of L-Tyr in various buffers. The CC (%) and the total variance captured by the model (%) are also indicated.**

	EEM <sub>  </sub> 1 comp.	EEM <sub>  </sub> 2 comp.	EEM <sub>  </sub> 3 comp.
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	274/304	278/304	278/304
<b>Fit model (%)</b>	100	79.08	49.80
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	270/304	272/304
<b>Fit model (%)</b>	/	20.92	48.29
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	/	274/306
<b>Fit model (%)</b>	/	/	1.91
<b>Variance explained (%)</b>	99.98	99.99	99.99
<b>CORCONDIA (%)</b>	100	-2817.9	-4059.6

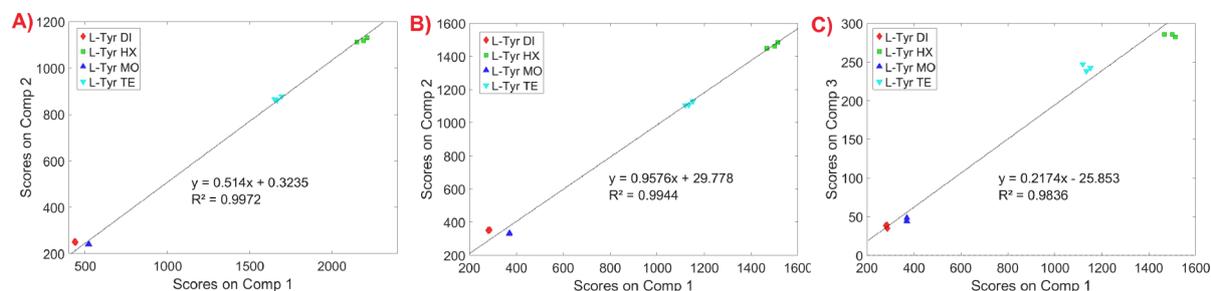


**Figure A- 33: PARAFAC models built using the non-normalized pre-treated EEM<sub>||</sub> data consisting of L-Tyr collected in the various insulin buffers (dimeric D, monomeric M, tetrameric T, and hexameric H). The scores and spectral loadings are shown for the two (top row) and three (bottom row) component models.**

Looking at the odd excitation loadings shape of the two and three component models, the very negative CC obtained and the simplicity of the dataset, it is tempting to suggest that both models overfit the data. However due to the large variation in intensity between samples in this dataset (Figure 55), we might not be able to trust the scores and loadings obtained for either of these models as representative of the true chemical variance. Samples with high intensity have indeed a much larger influence on the model, and can ultimately play against the need of

<sup>iii</sup> Fit model (%) represents the sum-squared signal captured by each component relatively to the total amount of signal captured by the model. To obtain the variance (%) captured by each component, one simply has to multiply the fit model by the total variance captured by the model. Because the latter is almost 100% in all the models we built, fit model and variance are in fact virtually identical for the components.

variability in the data, which is one of the most important pre-requisites for a good PARAFAC model<sup>289</sup>. One aspect which reveals this type of issue is the obtention of strong collinearity between factors scores<sup>289</sup>, which is exactly what was found *e.g.* for the three component model (Figure A- 34).

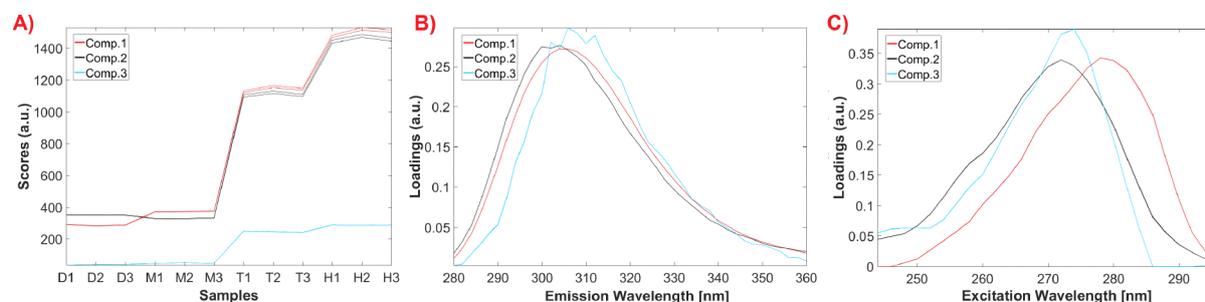


**Figure A- 34: Scatter plots of the scores obtained for the two component PARAFAC model (A, Comp.1 vs Comp.2), and for the three component model (B, Comp.1 vs Comp.2) and (C, Comp.1 vs Comp.3).**

This large influence also leads to models that are highly reproducible (Table A- 13 and Figure A- 35). The reproducibility is thus not a guarantee of the chemical meaning of a model.

**Table A- 13: Summary of the variance, CC, time and iterations obtained for each of the 10 repetitions of the three component PARAFAC model, built using the non-normalized pre-treated EEM<sub>||</sub> data consisting of L-Tyr in the various buffers.**

	Variance (%)	CC (%)	Time (s)	Iterations
<b>Model 1</b>	99.992	- 4059.646	68.645	6303
<b>Model 2</b>	99.992	- 4139.659	81.834	8175
<b>Model 3</b>	99.992	- 4059.631	73.677	6999
<b>Model 4</b>	99.992	- 4059.634	52.354	5059
<b>Model 5</b>	99.992	- 4139.677	88.272	9347
<b>Model 6</b>	99.992	- 4059.628	46.214	4147
<b>Model 7</b>	99.992	- 4059.611	51.436	4841
<b>Model 8</b>	99.992	- 4059.649	71.293	6923
<b>Model 9</b>	99.992	- 4059.614	70.305	6901
<b>Model 10</b>	99.992	- 4139.711	69.363	6915

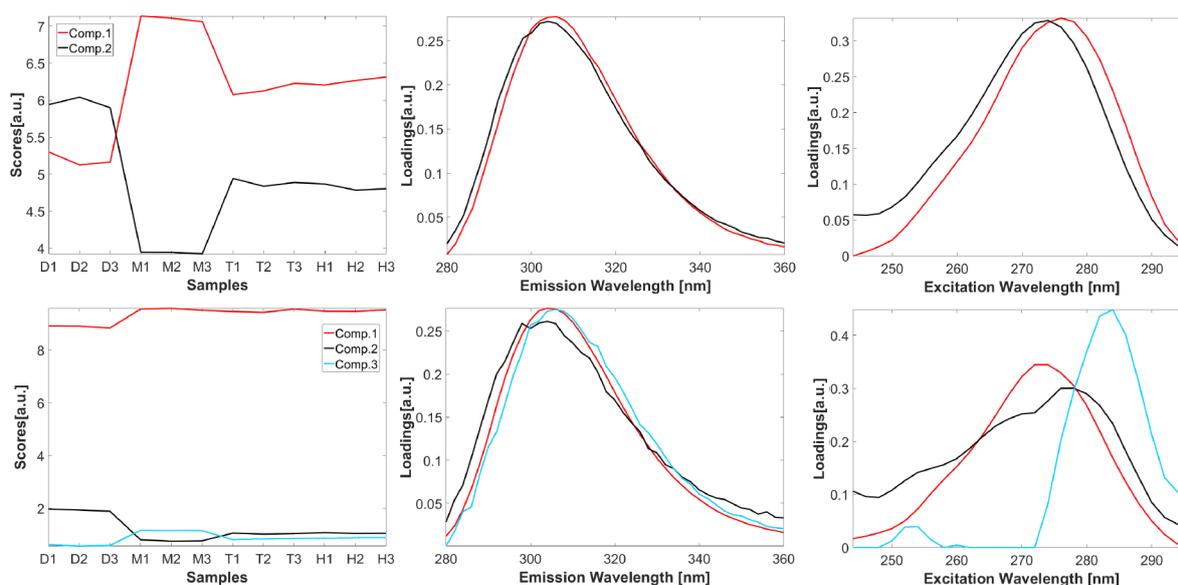


**Figure A- 35: Overlap of the scores (A) and loadings (B/C) obtained for the 10 repetitions of the three component PARAFAC model built using the non-normalized pre-treated EEM<sub>||</sub> data, consisting of L-Tyr collected in various buffers. The solutions ~perfectly overlap.**

To accurately investigate at the possible influence of buffers on the spectral shape of L-Tyr and enable the PARAFAC model to focus on real physico-chemical variations, each EEM has thus to be normalized to one in order to give the same weight to all samples. The new models build using the normalised data are presented in Table A- 14 and Figure A- 36.

**Table A- 14: Summary of the fit captured by each component for the normalized EEM<sub>||</sub> PARAFAC models of L-Tyr in various insulin buffers. The CC and the total variance captured by the model are also indicated.**

	EEM <sub>  </sub> 1 comp.	EEM <sub>  </sub> 2 comp.	EEM <sub>  </sub> 3 comp.
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	274/304	276/306	274/304
<b>Fit model (%)</b>	100	63.13	97.31
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	274/304	278/304
<b>Fit model (%)</b>	/	38.87	1.81
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	/	/	284/306
<b>Fit model (%)</b>	/	/	0.88
<b>Variance explained (%)</b>	99.96	99.98	99.98
<b>CORCONDIA (%)</b>	100	-66.1	-16182.1



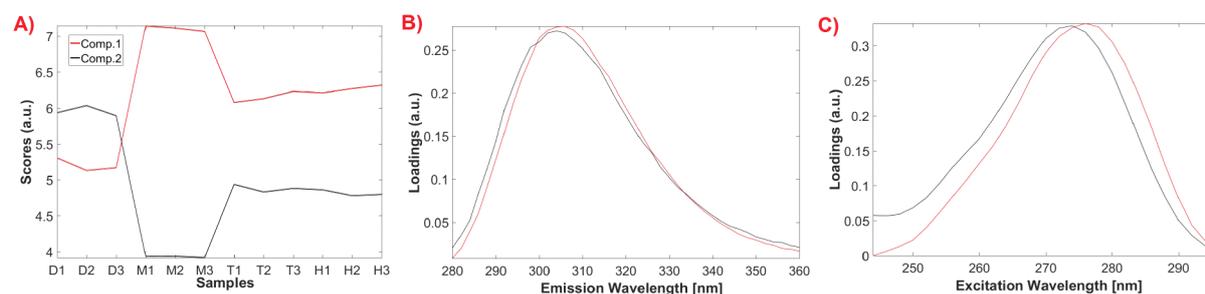
**Figure A- 36: PARAFAC models built using the normalized pre-treated EEM<sub>||</sub> data consisting of L-Tyr collected in the various insulin buffers. The scores and spectral loadings are shown for the two (top row) and three (bottom row) component models.**

As one can see, the results are very different from the non-normalized case. Due to the strange shape of the excitation loadings in the three component model (Figure A- 36, bottom right) as well as the very low % of variance represented by component 2 and 3 and the very negative CC obtained (Table A- 14), it is clear that we are over-fitting the data. The repetition of 10 models is also very unstable (data not shown). On the other hand, the two component

model has more believable features with loadings shapes that look like Tyr spectra, and two components that both explain a significant amount of variance. The repetition of 10 models is also very stable (Table A- 15 and Figure A- 37).

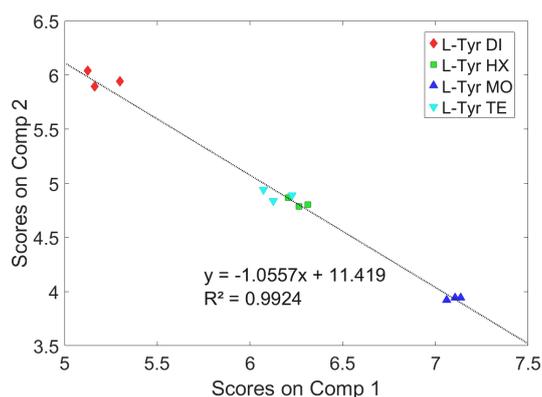
**Table A- 15: Summary of the variance, CC, time and iterations obtained for each of the 10 repetitions of the two component PARAFAC model, built using the normalized pre-treated EEM<sub>||</sub> data consisting of L-Tyr in the various buffers.**

	Variance (%)	CC (%)	Time (s)	Iterations
<b>Model 1</b>	99.977	- 66.052	20.467	1431
<b>Model 2</b>	99.977	- 66.389	21.950	1747
<b>Model 3</b>	99.977	- 66.390	21.668	1735
<b>Model 4</b>	99.977	- 66.052	15.428	879
<b>Model 5</b>	99.977	- 66.389	22.012	1755
<b>Model 6</b>	99.977	- 66.053	20.624	1385
<b>Model 7</b>	99.977	- 66.053	21.015	1443
<b>Model 8</b>	99.977	- 66.053	20.967	1393
<b>Model 9</b>	99.977	- 66.053	21.107	1437
<b>Model 10</b>	99.977	- 66.389	23.463	1755



**Figure A- 37: Overlap of the scores (A) and loadings (B/C) obtained for the 10 repetitions of the two component PARAFAC model, built using the normalized pre-treated EEM<sub>||</sub> data consisting of L-Tyr collected in various buffers. The solutions ~perfectly overlap.**

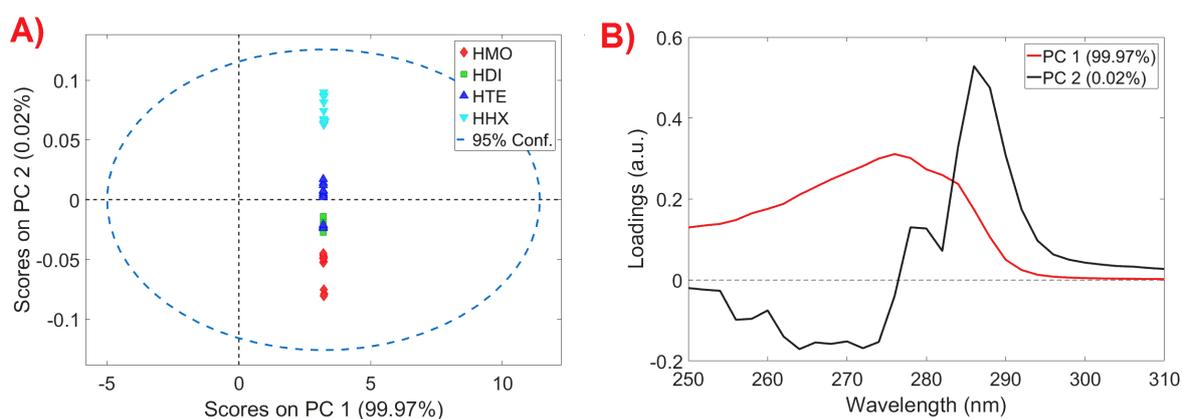
The scores for component 1 vs 2 show however a perfect collinearity (Figure A- 38). A collinearity in the scores might sometimes be related to a chemical property of the system or sometimes due to improper sampling<sup>309,493</sup>, however, here we are working with a very simple system *i.e.* diluted L-Tyr in buffers measured at the same concentration. Associated to the decrease in CC going from one component (100%) to two components model (-66.1%), this perfect collinearity is symptomatic of data being overfitted, *i.e.* one component is split into two that both represent partly the same underlying phenomenon.



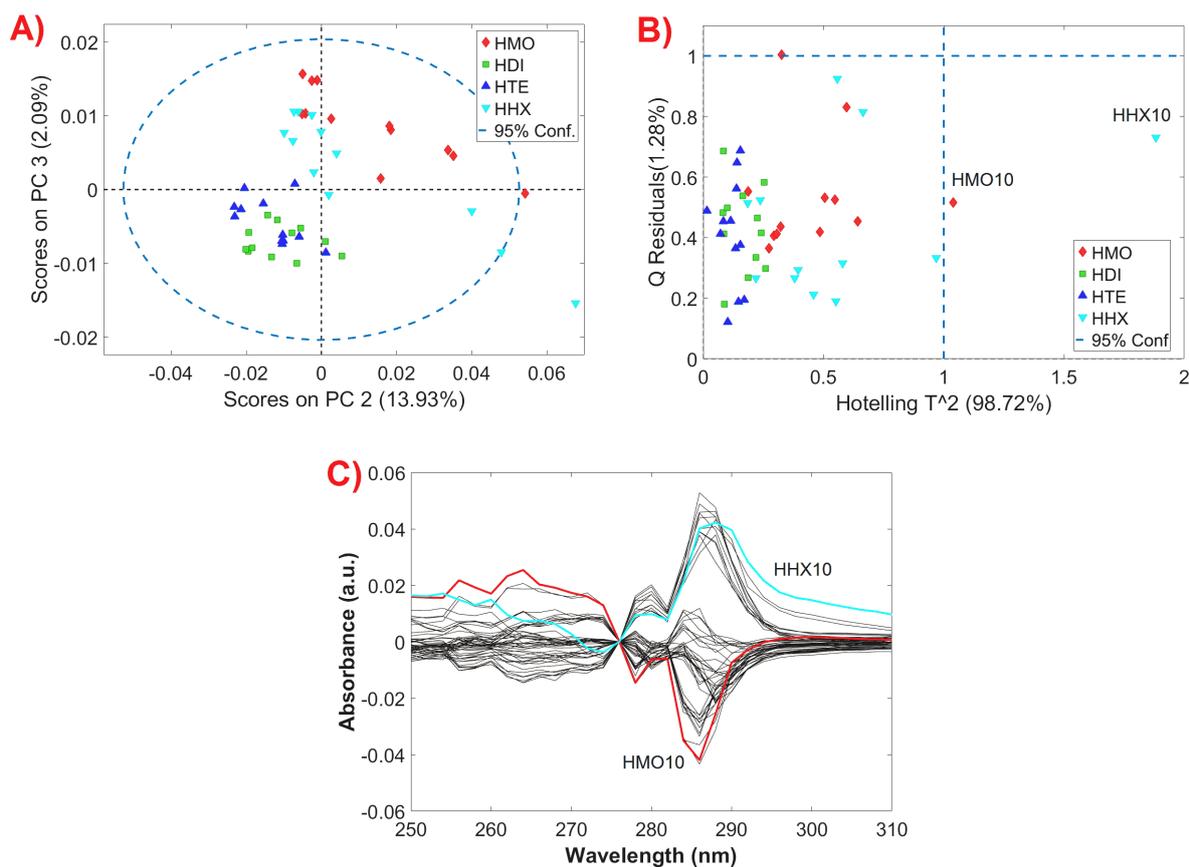
**Figure A- 38: Scatter plot of the scores (Comp.1 vs Comp.2) obtained for the two component PARAFAC model built using the normalized L-Tyr EEM<sub>||</sub> data.**

This simple analysis therefore serves as validation and control experiment, to prove that the use of different buffers only influences the emission of L-Tyr in solution and not its spectral shape. It is also interesting to note that on data where the samples weight is removed (normalized), a model that is reproducible still does not necessarily mean that the model is good. Overfitting as seen here might be present, leading the model to converge towards a stable local minima solution, that does not relate to real chemical information.

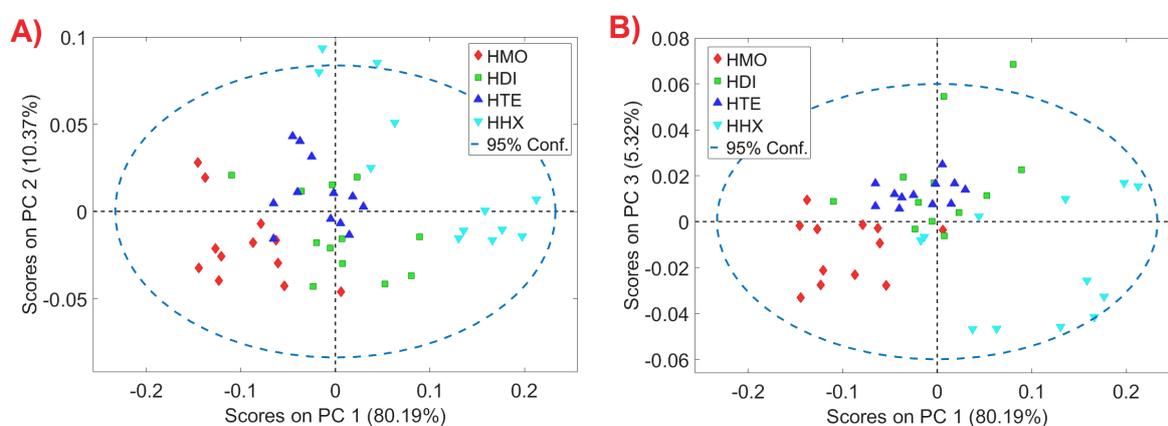
### 7.19 Details on the PCA/MCR models built on the UV-Vis data

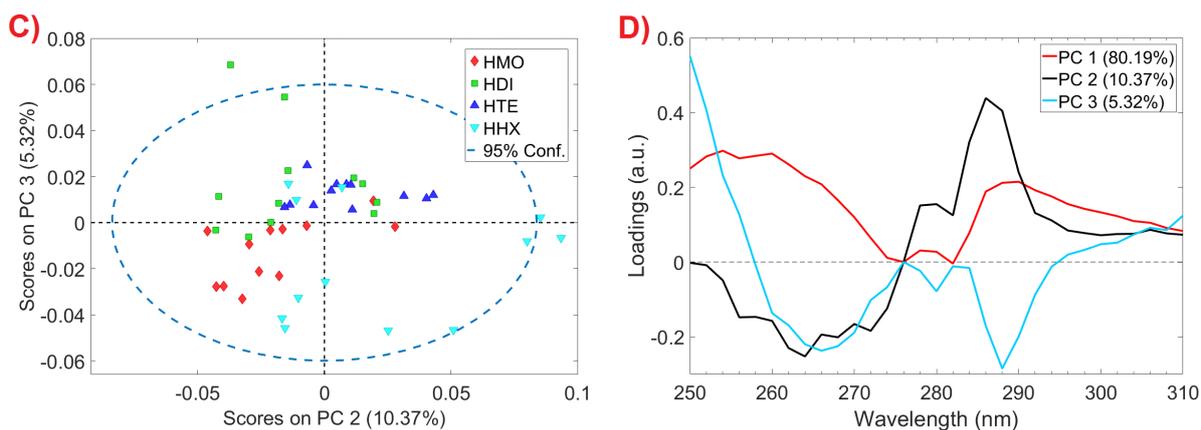


**Figure A- 39: (A/B) Scores scatter plots and loadings of the two component PCA model built using normalized UV-Vis spectra of 2 mg/mL insulin samples.**

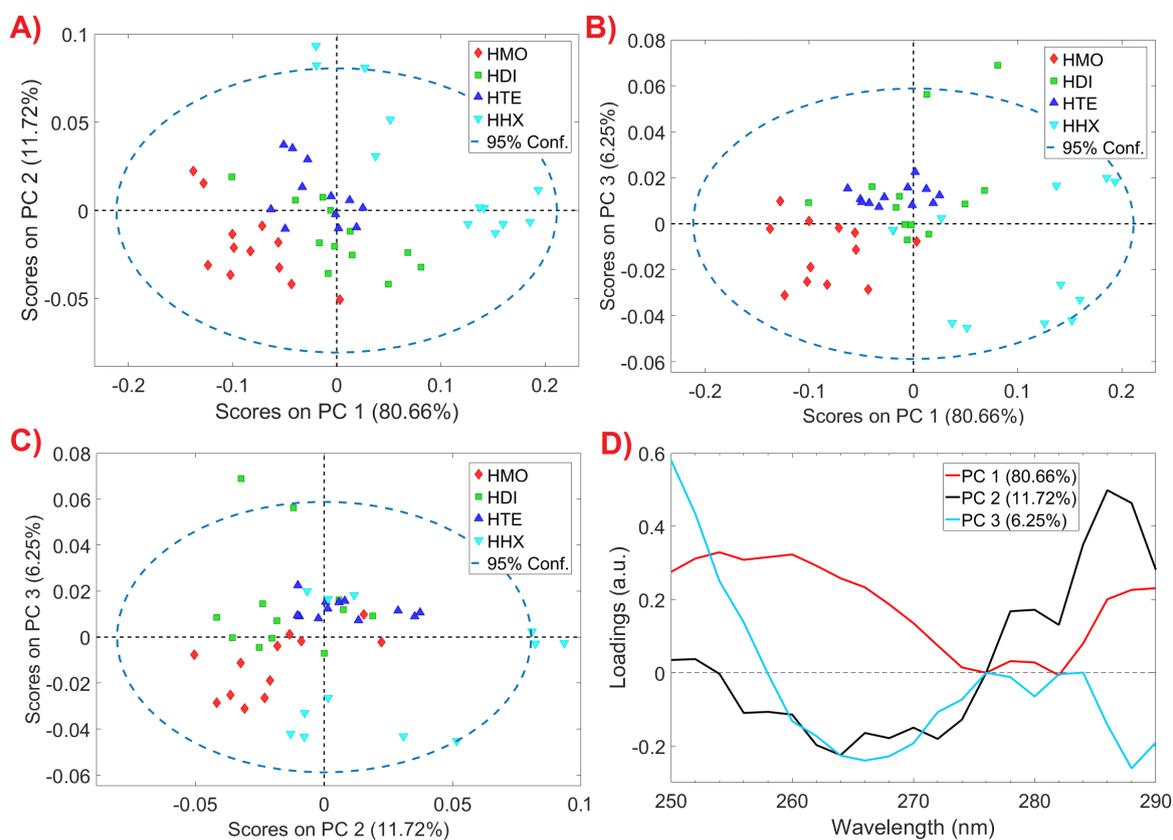


**Figure A- 40:** (A) Scatter plot of PC2vsPC3 obtained for the PCA model built using mean-centered UV-Vis data of insulin samples at 2 mg/mL. (B) The hotelling's T<sup>2</sup> versus Q statistics show that two samples could be outliers. This diagnostic can be a bit severe considering that mean-centring brings out differences that are otherwise very small. The mean-centre spectra of these two samples were plotted against the entire dataset (C). Their similarity to other samples proves that these two samples are not outliers. It is also interesting to see that the strongest outlier HHX10 was outlier due to a stronger scatter seen between 290-310 nm.



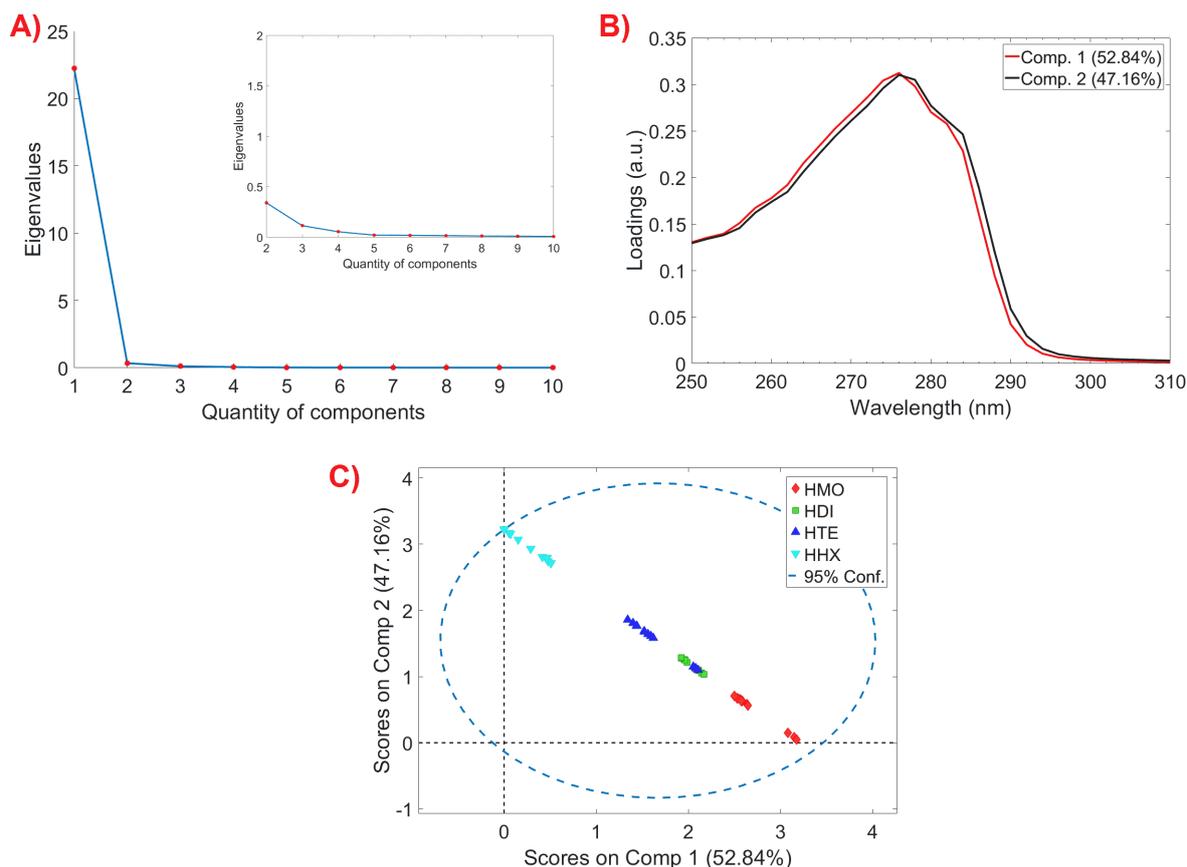


**Figure A- 41: Scatter plot (A/B/C) and loadings (D) of the PCA model built using mean-centred UV-Vis data of 0.2 mg/mL insulin samples. The model was built with 3 PCs and explained 95.87% of the variance.**



**Figure A- 42: Scores scatter plots (A/B/C) and loadings (D) for the three component PCA model built using mean-centred and cut 0.2 mg/mL insulin data. The Hotelling's  $T^2$  vs  $Q$  statistics is not shown, but neither of the samples were real outliers.**

The first step in the MCR model was to try to estimate the quantity of components to extract. The rank analysis was done with SVD, as in the previous chapter. It was difficult to choose between two and three components, but the option with two components gave perfectly collinear scores as shown Figure A- 43.

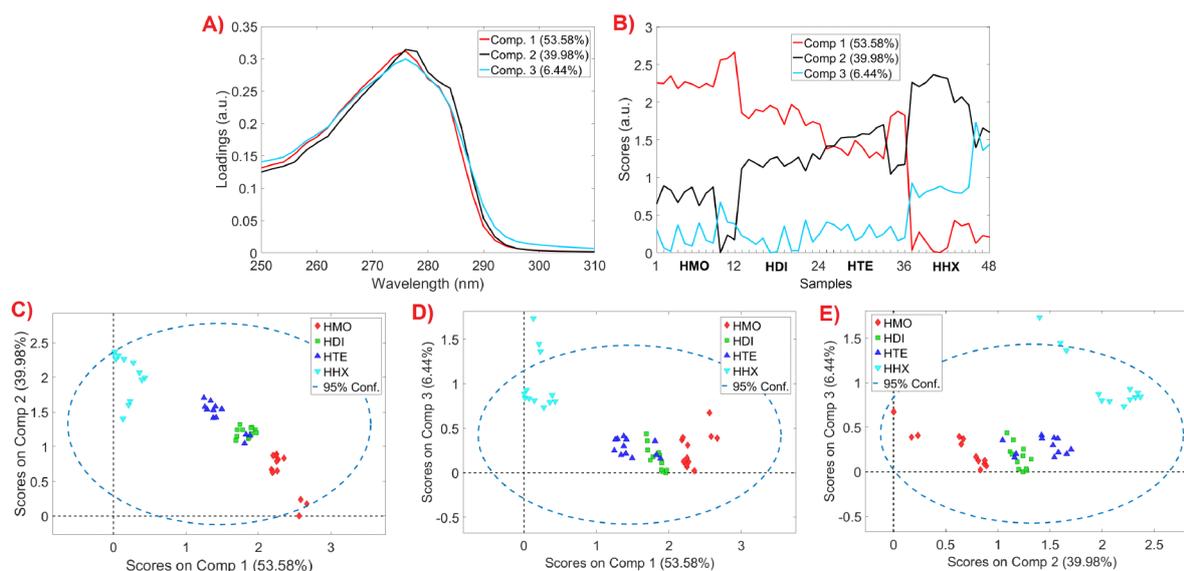


**Figure A- 43: (A) Singular values (or Eigenvalues) obtained via SVD on the normalized insulin UV-Vis data at 2 mg/mL. An attempt of two component MCR-ALS model was built using this data for which the loadings (B) and the scatter plot PC1vsPC2 (C) are shown.**

The model was therefore built with three components, a quantity that was also suggested by the independent PCA models built above. Non-negativity was applied on both concentration and pure spectra profiles, as well as normalization of the spectral profiles to equal length (Euclidean) to avoid intensity ambiguities between components.

The convergence criterion was set so that the relative difference in fit between two successive iterations was below  $10^{-8}$ . Exterior normalization was used as initial estimates which is one of the two options available in PLS toolbox for MCR<sup>iii</sup>.

<sup>iii</sup> The two options available to initialize the MCR-ALS algorithm in the PLS Toolbox, are based on the same principle, which is to use as pure variable or samples the ones that are the most outside of the data cloud, therefore the purest. Their selection is based either with their Euclidean distance to the cloud, or with their distance after normalization which is the method selected here.



**Figure A- 44: Loadings (A) and scores (B) shown for the three component MCR model built using the normalized insulin data at 2 mg/mL. The scores scatter plots are shown in C/D/E.**

The model explained 99.99% of the variance, the LOF was 0.31% and the rotational ambiguities were estimated in Table A- 16.

**Table A- 16: Estimation of rotational ambiguities per component, for three component MCR model built using the insulin absorbance spectra at 2 mg/mL. The model was built using normalization and non-negativity.**

	Comp.1 (53.58%)	Comp.2 (39.98%)	Comp.3 (6.44%)
$f_n \text{ max}$	0.413	0.390	0.259
$f_n \text{ min}$	0.381	0.371	0.234
$f_n \text{ max-f}_n \text{ min}$	0.032	0.019	0.025

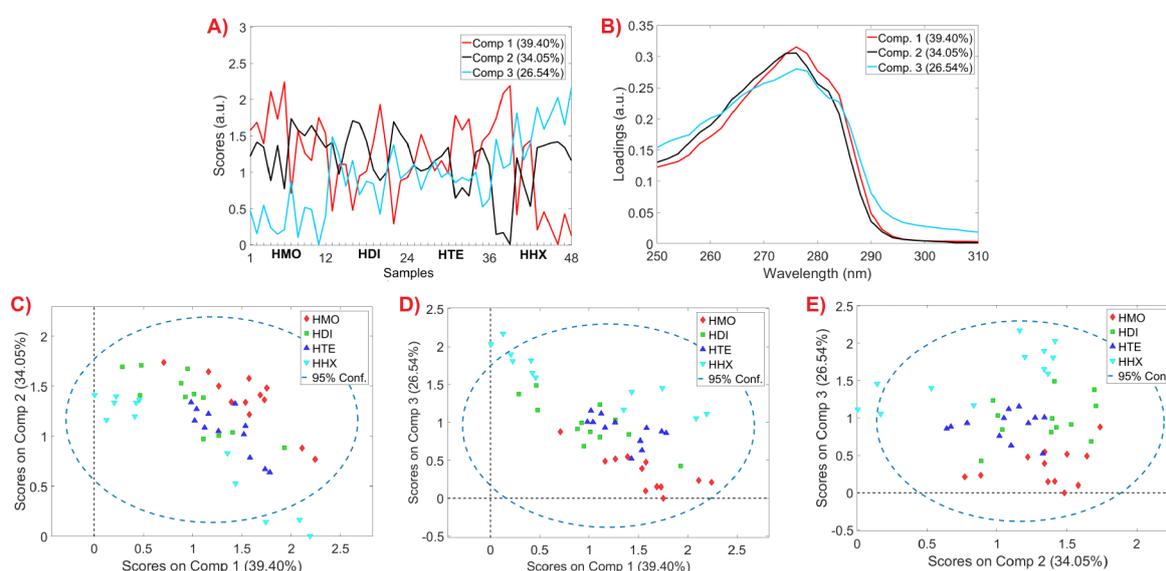
Looking at the scatter score plots in Figure A- 44, it is clear that the MCR model can separate the insulin forms and valuable information is present in the spectral loadings:

- Starting from the end, Comp.3 is carrying the scatter information and shows shoulders at 258 and 264 nm that are related to Phe absorption. As seen in D/E, this component enables to separate HHX from the rest of the samples.
- Comp.2 is clearly related to the red-shift that occurs during insulin aggregation and enables to separate the species based on their extent of red-shift (C/E).
- Comp.1 is the opposite of Comp.2 and represents a blue shifted signal which enables to separate the species based on their lack of aggregation (C/D).
- Also if we consider the spectral loadings of Comp.2 and 3 together (Figure A- 44, A), we can clearly confirm that the decreasing slope seen above 285 nm and that permitted

species separation in PCA (Figure 57), is shifted upon aggregation because of a shift in the absorbance peak and the presence of scattered light in the peak tail. This confirms the findings of an older publication on insulin where the author measured the pH dependency of insulin aggregation with UV-Vis difference spectra<sup>658</sup>. The author warned that even if the method provides information on internal hydrogen bonding, a potential contribution of light scattering may also need to be considered, which is what we just proved.

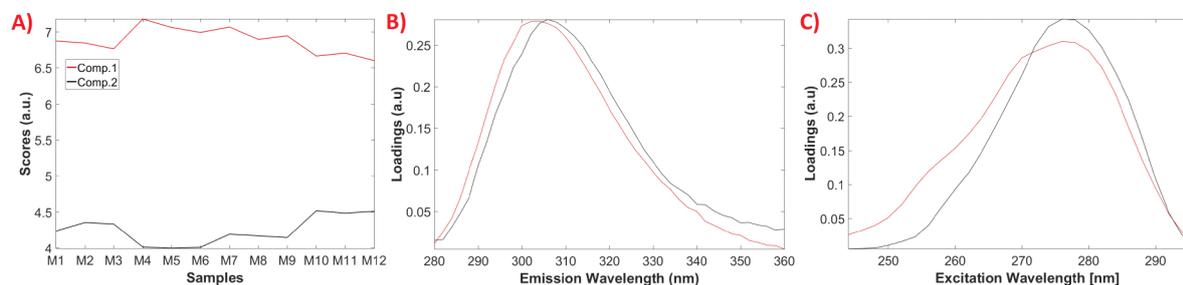
The components extracted are thus interesting, but they do not correspond to pure Tyr or Phe contributions. The model (Figure A- 44) has however the advantage to confirm what was difficult to see in PCA but also to clearly separate insulin species. It is also interesting to note that once again HTE are found to lie between HDI and HHX. The higher proximity to HDI is clear once again, and the same samples HTE R10-12 are found to overlap with HDI. A similar MCR model was built for the 0.2 mg/mL dataset.

The model, shown in Figure A- 45, explains also over 99.99% of the data with components that have similar shapes. However similar to what was seen for the PCA model, the model fails to separate insulin species probably due to the low signal to noise ratio (SNR) and a scatter component that was more noise and measurement related than being due to protein size changes.



**Figure A- 45: Loadings (A) and scores (B) shown for the three component MCR model built using the normalized insulin data (0.2 mg/mL). The scores scatter plots are shown in C/D/E.**

## 7.20 Repeated PARAFAC models obtained on the EEM<sub>||</sub> monomeric insulin data at 0.2 mg/mL

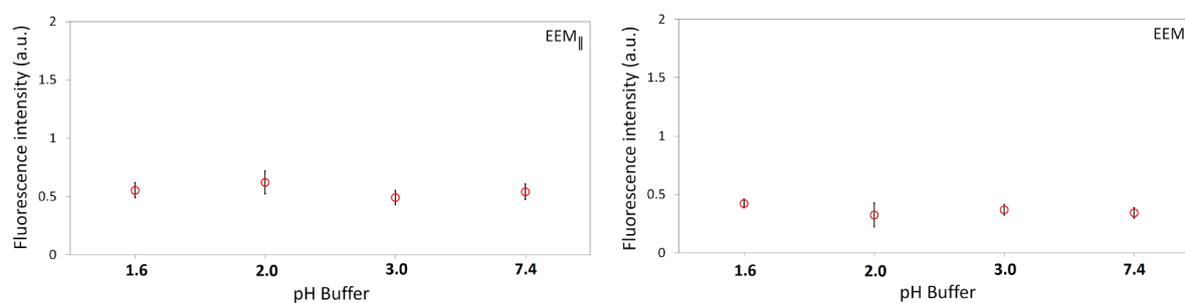


**Figure A- 46:** Overlap of the scores (A) and loadings (B/C) obtained for the 10 repetitions of the two component PARAFAC model, built using the normalized pre-treated EEM<sub>||</sub> data consisting of the 12 insulin monomer replicates at 0.2 mg/mL. Most of the solutions perfectly overlap.

**Table A- 17:** Summary of the variance, CC, time and iterations obtained for each of the 10 repetitions of the two component PARAFAC model, built using the normalized pre-treated EEM<sub>||</sub> data consisting of the 12 insulin monomer replicates at 0.2 mg/mL.

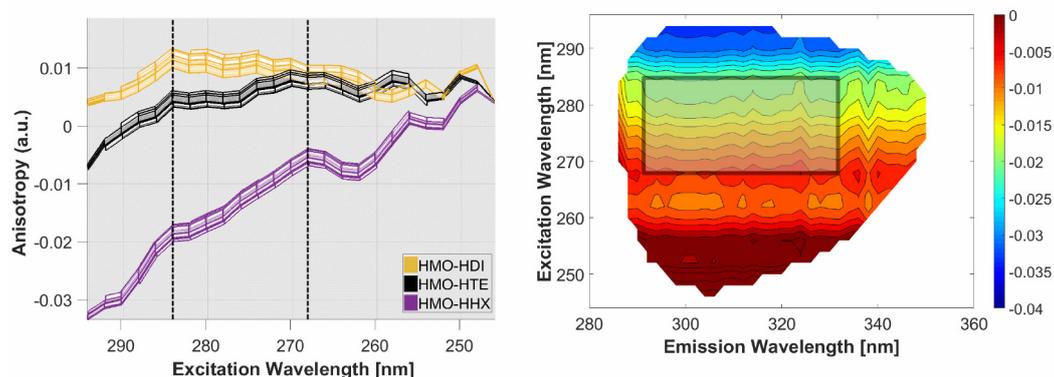
	Variance (%)	CC (%)	Time (s)	Iterations
<b>Model 1</b>	99.980	99.153	74.662	10000
<b>Model 2</b>	99.980	99.158	75.833	10000
<b>Model 3</b>	99.980	99.158	75.629	10000
<b>Model 4</b>	99.980	99.158	76.597	10000
<b>Model 5</b>	99.980	99.164	76.253	10000
<b>Model 6</b>	99.980	99.154	75.099	10000
<b>Model 7</b>	99.980	99.160	74.927	10000
<b>Model 8</b>	99.980	99.153	74.942	10000
<b>Model 9</b>	99.980	99.157	74.770	10000
<b>Model 10</b>	99.980	99.157	74.708	10000

## 7.21 Fluorescence of Tyrosinate in the various buffers used

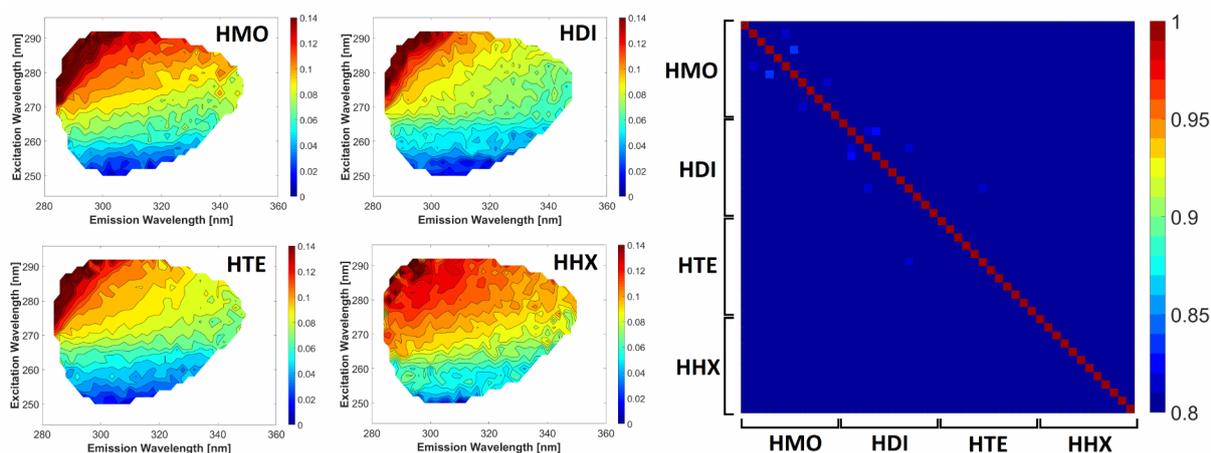


**Figure A- 47:** Average fluorescence intensity ( $\pm$  STD) obtained at  $\lambda_{ex/em} = 295/345$  nm across all the samples (0.2 mg/mL) and all the buffers used. The intensity was plotted on EEM<sub>||</sub> and EEM<sub>⊥</sub> in case the tyrosinate emission would be polarized/depolarized.

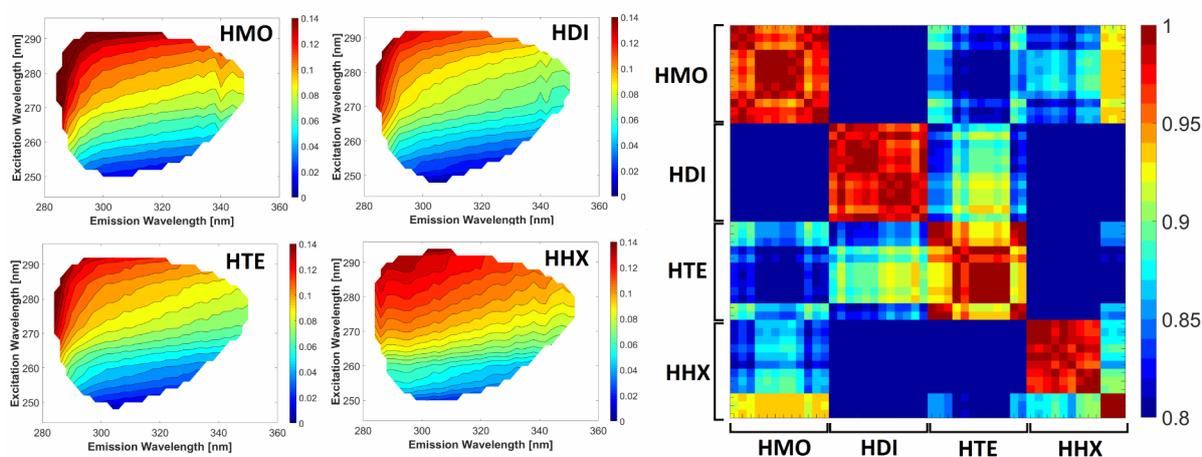
## 7.22 Details about the analysis of insulin *aniso*-MDF spectra



**Figure A- 48:** On the left is shown along the excitation, the difference *aniso*-MDF spectra calculated between the denoised HMO data and the rest of the insulin forms. On the right is shown as a contour plot the specific result obtained between HMO and HHX. The region selected to calculate the SimI is shown in dash (left) and as rectangle (right).

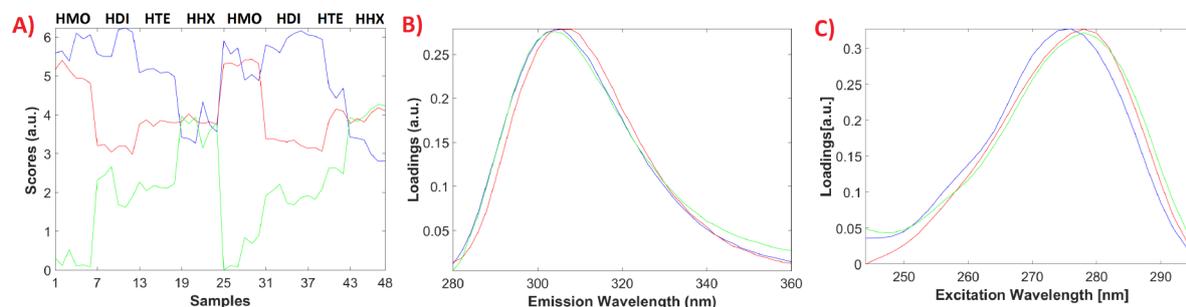


**Figure A- 49:** (Left) Average *aniso*-MDF spectra obtained for each insulin form at 0.2 mg/mL. (Right) Corresponding SimI calculated on the selected region.



**Figure A- 50:** (Left) Average denoised *aniso*-MDF spectra obtained for each insulin form at 0.2 mg/mL. (Right) Corresponding SimI calculated on the selected region.

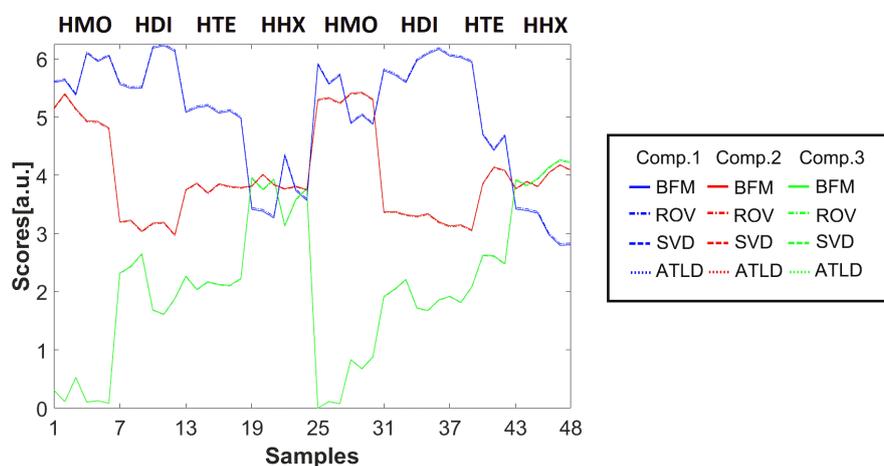
## 7.23 Robustness of the PARAFAC model built using the EEM<sub>||</sub> insulin oligomers data, at 0.2 mg/mL

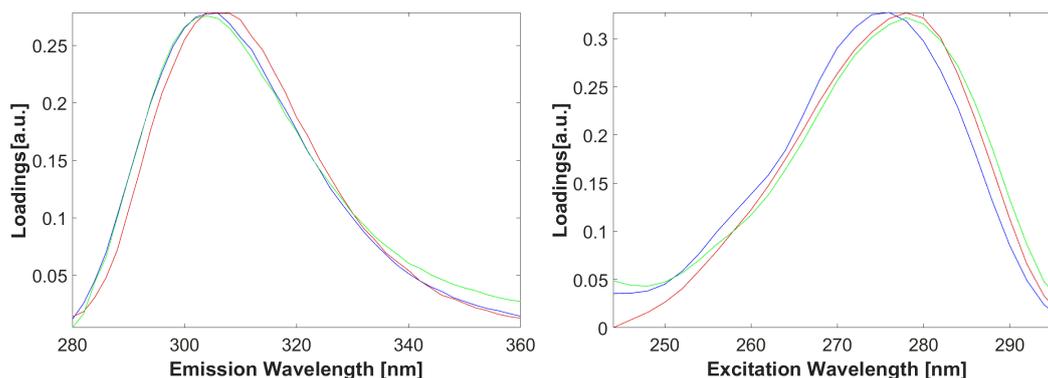


**Figure A- 51: Overlap of the scores (A) and loadings (B/C) obtained for the 10 repetitions of the three components PARAFAC model, built using the normalized pre-treated 0.2 mg/mL EEM<sub>||</sub> insulin data using the best fit model (BFM) as initialization. Most of the solutions perfectly overlap.**

**Table A- 18: Summary of the variance, CC, time and iterations obtained for each of the 10 repetitions of the three component PARAFAC model, built using the normalized pre-treated 0.2 mg/mL EEM<sub>||</sub> insulin data using the best fit model (BFM) as initialization.**

	Variance (%)	CC (%)	Time (s)	Iterations
<b>Model 1</b>	99.977	- 3391.842	37.252	2883
<b>Model 2</b>	99.977	- 3391.853	31.091	2215
<b>Model 3</b>	99.977	- 3391.881	82.337	7981
<b>Model 4</b>	99.977	- 3391.857	32.245	2403
<b>Model 5</b>	99.977	- 3391.891	83.429	7899
<b>Model 6</b>	99.977	- 3391.861	81.432	7635
<b>Model 7</b>	99.977	- 3391.849	55.068	4809
<b>Model 8</b>	99.977	- 3391.867	82.992	7943
<b>Model 9</b>	99.977	- 3391.733	27.300	1755
<b>Model 10</b>	99.977	- 3391.899	80.419	7845



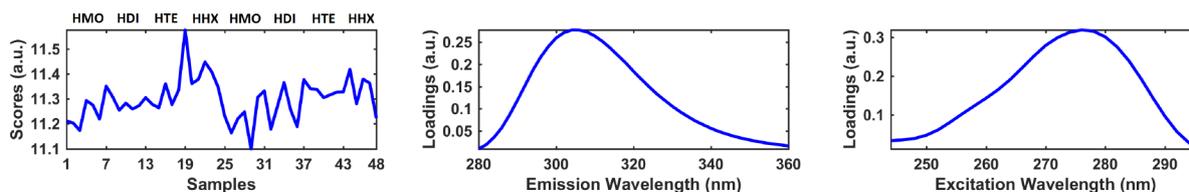


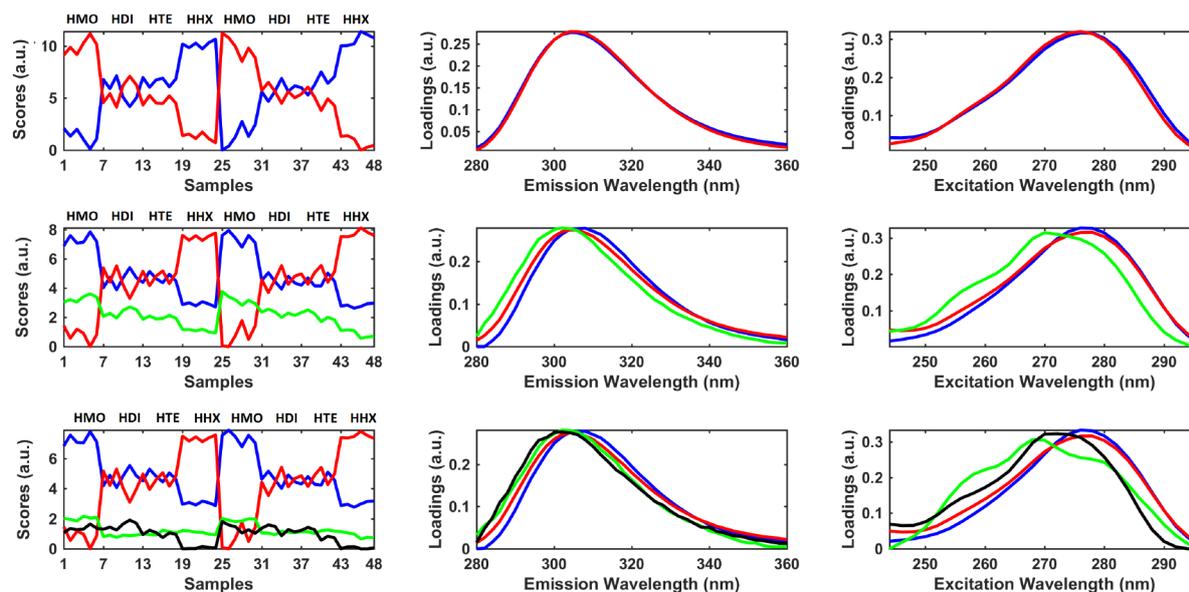
**Figure A- 52:** Almost perfect overlap of the scores (top) and loadings (bottom) obtained for the three component PARAFAC models built using the normalized pre-treated 0.2 mg/mL EEM<sub>||</sub> insulin data, using either BFM, ROV, SVD or ATLD as initialization methods.

**Table A- 19:** Summary of the fit captured by each component for the three components PARAFAC models built using the normalized insulin EEM<sub>||</sub> data, using either the best fit model of 10 small runs (BFM), random orthogonalized values (ROV), singular value decomposition (SVD) or alternating trilinear decomposition (ATLD) as initialization methods.

	BFM	ROV	SVD	ATLD
<b>C1 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	276/306	276/306	276/306	276/306
<b>Fit model (%)</b>	53.47	53.93	53.47	53.93
<b>C2 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	278/304	278/304	278/304	278/304
<b>Fit model (%)</b>	34.05	33.73	34.05	33.73
<b>C3 <math>\lambda_{ex}/\lambda_{em}</math> (nm)</b>	278/304	278/304	278/304	278/304
<b>Fit model (%)</b>	12.48	12.34	12.48	12.34
<b>Variance explained (%)</b>	99.98	99.98	99.98	99.98
<b>CORCONDIA (%)</b>	-3391.73	-3364.38	-3391.88	-3364.37
<b>Split-half half (%)</b>	93.3	89.5	90.5	97.3
<b>Split-half random (%)</b>	90.8	93.7	99.1	94.1

## 7.24 Robustness of the PARAFAC model built using the EEM<sub>⊥</sub> insulin oligomers data, at 0.2 mg/mL

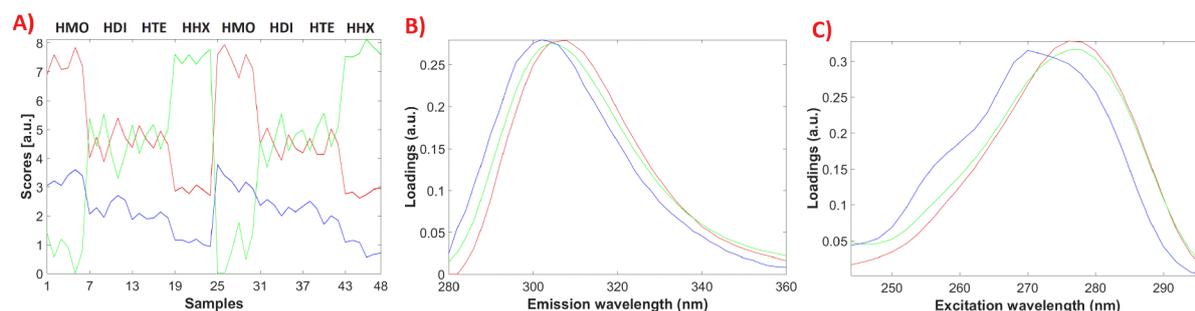




**Figure A- 53:** Scores (left) and loadings (middle + right) of the PARAFAC models built using the pre-treated and normalised 0.2 mg/mL EEM<sub>1</sub> insulin data. The models were fitted with an increasing number of components, starting from one (top) and going up to four components (bottom). Comp.1/2/3/4 are respectively in blue/red/green/black.

**Table A- 20:** This table sums up the CC, total variance captured, and results from split-half analysis for the 1 to 4 components PARAFAC models, built using the normalized insulin EEM<sub>1</sub> data at 0.2 mg/mL.

	1 Comp.	2 Comp.	3 Comp.	4 Comp.
Variance explained (%)	99.94	99.97	99.98	99.98
CORCONDIA (%)	100	-4966.33	-12958.74	-9289.07
Split-half half (%)	100	100	99	70.2
Split-half random (%)	100	100	99	82.3

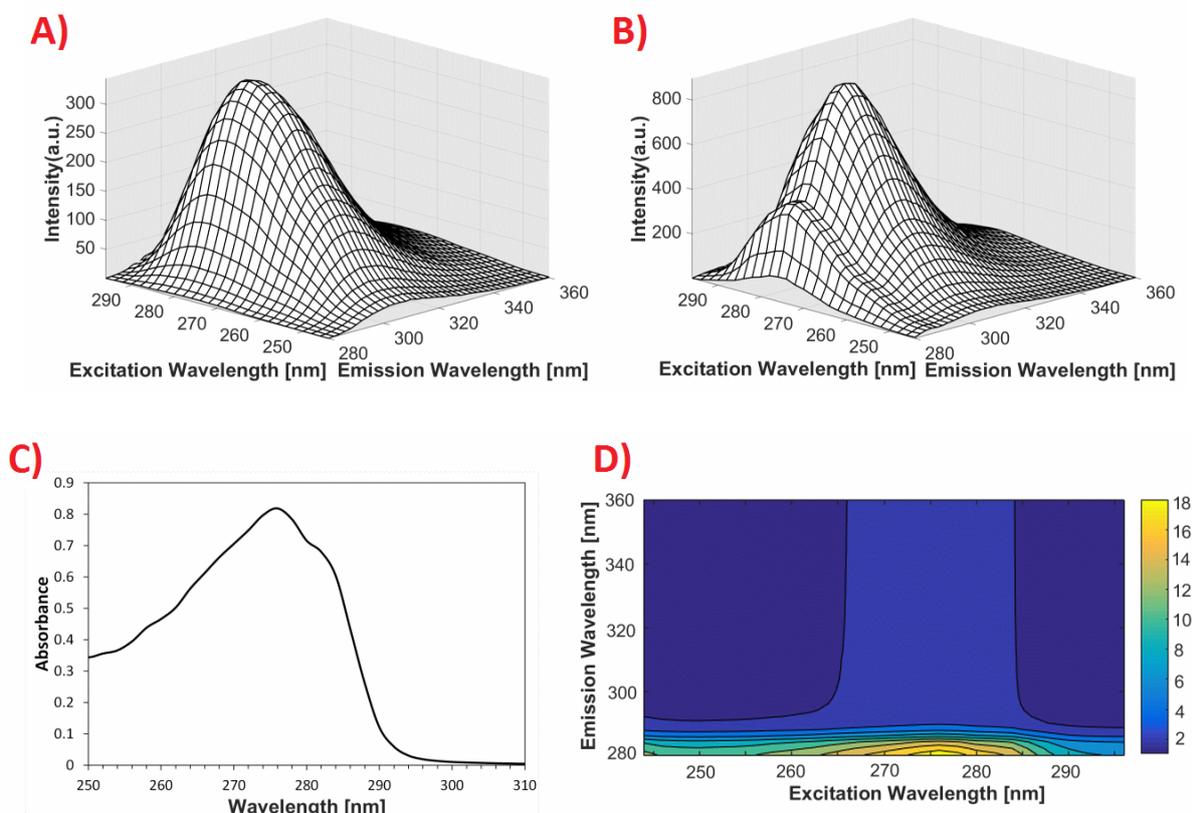


**Figure A- 54:** Overlap of the scores (A) and loadings (B/C) obtained for the 10 repetitions of the three components PARAFAC model, built using the normalized pre-treated 0.2 mg/mL EEM<sub>1</sub> insulin data using the best fit model (BFM) as initialization. Most of the solutions perfectly overlap.

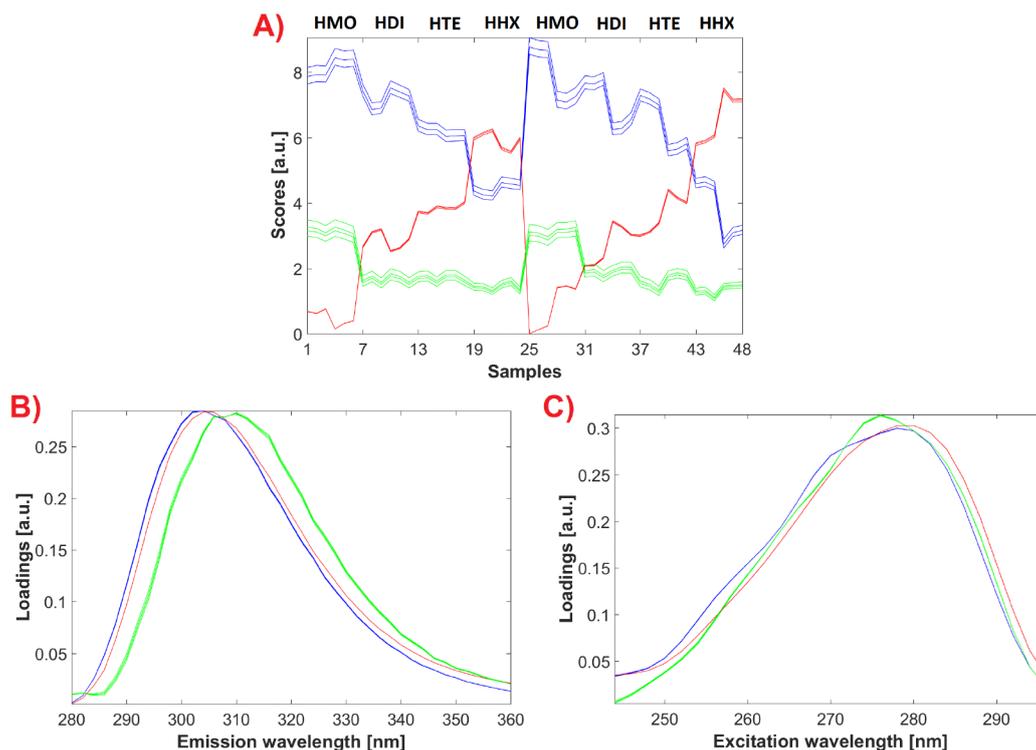
**Table A- 21: Summary of the variance, CC, time and iterations obtained for each of the 10 repetitions of the three component PARAFAC model, built using the normalized pre-treated 0.2 mg/mL EEM<sub>⊥</sub> insulin data using the best fit model (BFM) as initialization.**

	Variance (%)	CC (%)	Time (s)	Iterations
<b>Model 1</b>	99.981	-12958.761	24.876	1523
<b>Model 2</b>	99.981	-12958.828	23.575	1487
<b>Model 3</b>	99.981	-12995.919	22.190	979
<b>Model 4</b>	99.981	-12996.007	23.554	1385
<b>Model 5</b>	99.981	-12995.852	27.133	1709
<b>Model 6</b>	99.981	-12995.985	21.613	1109
<b>Model 7</b>	99.981	-12958.733	25.794	1525
<b>Model 8</b>	99.981	-12958.712	27.271	1741
<b>Model 9</b>	99.981	-12995.895	20.538	1103
<b>Model 10</b>	99.981	-12958.773	26.407	1749

### 7.25 Further information about the PARAFAC models built using the 2 mg/mL insulin samples



**Figure A- 55: The EEM<sub>||</sub> spectra of the first HMO sample are shown respectively in A and B before and after the IFE correction by ABA. In C is shown the absorbance spectrum of this sample and in D the correction matrix that was calculated from it.**



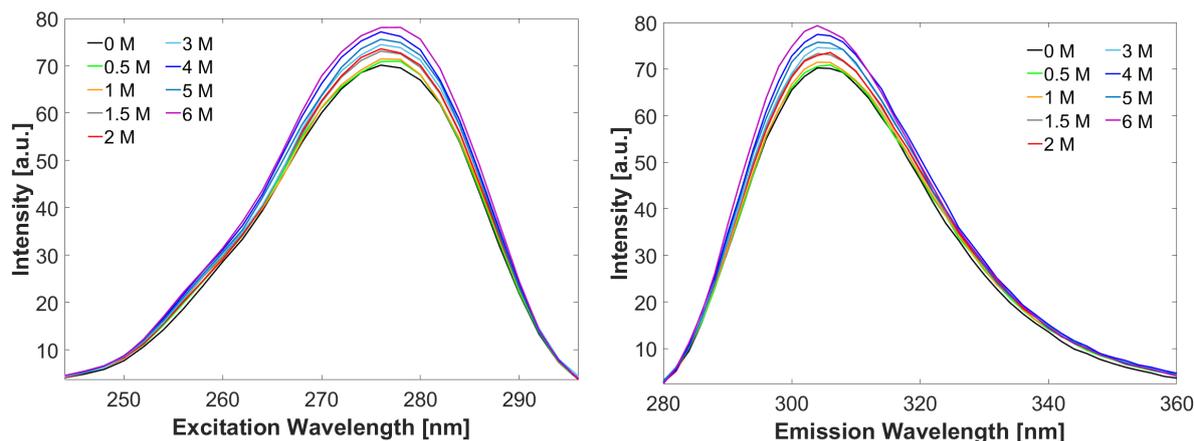
**Figure A- 56: Overlap of the scores (A) and loadings (B/C) obtained for the 10 repetitions of the three components PARAFAC model, built using the normalized (not IFE corrected) 2 mg/mL EEM<sub>||</sub> insulin data using the best fit model (BFM) as initialization.**

**Table A- 22: Summary of the variance, CC, time and iterations obtained for each of the 10 repetitions of the three component PARAFAC model, built using the normalized (not IFE corrected) 2 mg/mL EEM<sub>||</sub> insulin data, using the best fit model (BFM) as initialization.**

	Variance (%)	CC (%)	Time (s)	Iterations
<b>Model 1</b>	99.989	-1856.196	112.096	10000
<b>Model 2</b>	99.989	-1795.663	89.257	8109
<b>Model 3</b>	99.989	-1795.659	82.417	7373
<b>Model 4</b>	99.989	-1795.654	89.019	8013
<b>Model 5</b>	99.989	-1896.509	109.060	10000
<b>Model 6</b>	99.989	-1795.659	88.017	8057
<b>Model 7</b>	99.989	-1795.659	84.949	8043
<b>Model 8</b>	99.989	-1795.654	86.693	7987
<b>Model 9</b>	99.989	-1795.661	85.242	7655
<b>Model 10</b>	99.989	-1974.618	108.717	10000

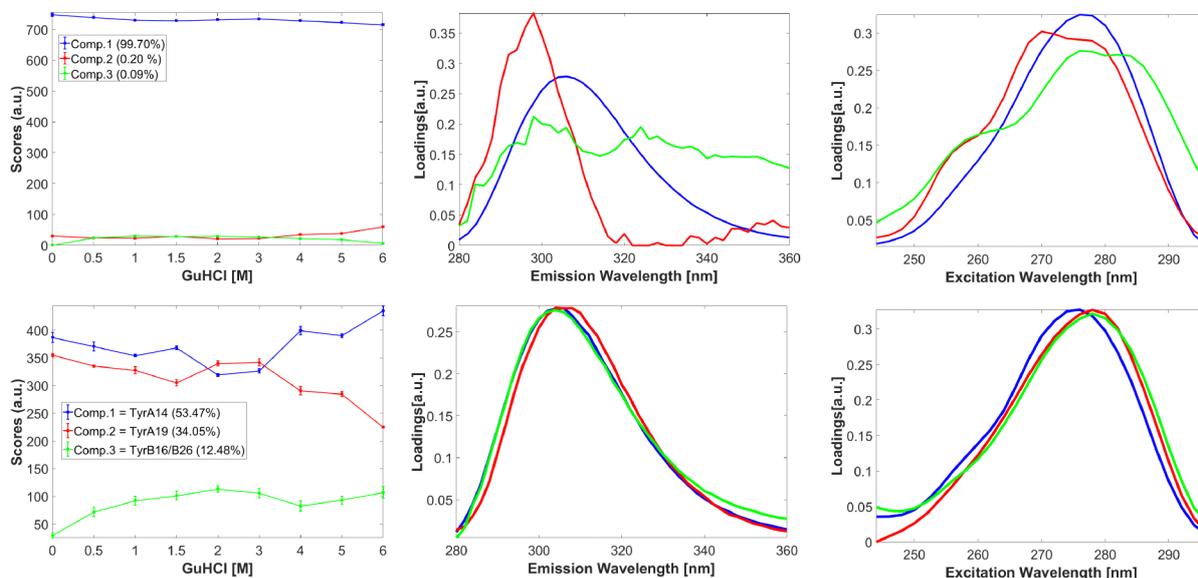
## 7.26 Denaturation of insulin monomers by GuHCl

ARMES data were collected by exposing monomeric insulin samples (in 20% acetic acid at 0.2 mg/mL) to increasing concentration of GuHCl. The polarised EEM<sub>||</sub> insulin data did not show much variation during the denaturation process, as shown in Figure A- 57.



**Figure A- 57: Excitation (left) and emission (right) spectra of a monomeric insulin solution at 0.2 mg/mL, exposed to an increasing concentration of GuHCl (0 to 6M). The spectra are averaged over three replicate measurements and are observed respectively at  $\lambda_{em} = 306$  nm and  $\lambda_{ex} = 276$  nm.**

At first glance, the intrinsic emission of insulin seems thus not appropriate to monitor the appearance of partially unfolded monomers. It is also not possible to build successful PARAFAC models on this data, due to the presence of multi-mode dependencies (Figure A-58, top row).



**Figure A- 58: Scores (left) and loadings (middle + right) of the three component PARAFAC models built using the 0.2 mg/mL EEM<sub>||</sub> insulin data exposed to increasing GuHCl concentration. On the top row is presented the model directly built using the data, and on the bottom the prediction obtained by the model developed in section 5.10.1. Error bars (in the scores) are obtained over three replicate measurements.**

However when the data (EEM<sub>||</sub>) is fitted with the model developed in section 5.10.1, clear changes can be observed in the scores that are coherent with a denaturation process

(Figure A- 58, bottom row). For example, up to 2M GuHCl, the intensity of the FRET species (Comp.3) increases, meaning that the transfer of energy towards TyrA19 or TyrA14 is not total anymore. The distance between fluorophores must thus increase which is expected during a denaturation process. This is also characterised by a simultaneous decrease in the intensity of the other components, due to their energy loss (especially for TyrA19).

Another very interesting aspect of the scores plot is the clear appearance of three regions upon addition of denaturant: [0 to 1.5-2]; [1.5-2 to 3-4] and [3-4 to 6] M GuHCl, that match the studies from Ahmad *et al.*<sup>400,474</sup>. These regions correspond to varying unfolded states of insulin monomer, and it is believed that the partially unfolded species responsible for insulin fibrillation exist between 1.5 - 4 M GuHCl (above 4 M, the monomer totally unfold)<sup>400,474</sup>. It seems thus that the model presented in Figure A- 58 managed to isolate and characterise the partially unfolded monomer, which is important and also very promising.

## References

1. Chennamsetty, N.; Voynov, V.; Kayser, V.; Helk, B.; Trout, B. L., Design of therapeutic proteins with enhanced stability. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106* (29), 11937-11942.
2. Jenkins, N.; Meleady, P.; Tyther, R.; Murphy, L., Strategies for analysing and improving the expression and quality of recombinant proteins made in mammalian cells. *Biotechnology and Applied Biochemistry* **2009**, *53*, 73-83.
3. Hermeling, S.; Crommelin, D. J. A.; Schellekens, H.; Jiskoot, W., Structure-Immunogenicity Relationships of Therapeutic Proteins. *Pharmaceutical Research* **2004**, *21* (6), 897-903.
4. Rosenberg, A. S., Effects of protein aggregates: an immunologic perspective. *The AAPS journal* **2006**, *8* (3), E501-7.
5. Mahler, H. C.; Friess, W.; Grauschopf, U.; Kiese, S., Protein aggregation: pathways, induction factors and analysis. *Journal of pharmaceutical sciences* **2009**, *98* (9), 2909-2934.
6. Den Engelsman, J.; Garidel, P.; Smulders, R.; Koll, H.; Smith, B.; Bassarab, S.; Seidl, A.; Hainzl, O.; Jiskoot, W., Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. *Pharma Res* **2011**, *28* (4), 920-933.
7. Harper, J. D.; Lansbury, P. T. J., Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annual review of biochemistry* **1997**, *66*, 385-407.
8. Fink, A. L., Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design* **1998**, *3* (1), R9-R23.
9. den Engelsman, J.; Garidel, P.; Smulders, R.; Koll, H.; Smith, B.; Bassarab, S.; Seidl, A.; Hainzl, O.; Jiskoot, W., Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. *Pharm Res* **2011**, *28* (4), 920-33.
10. Eftink, M. R., Fluorescence techniques for studying protein structure. *Methods of Biochemical Analysis* **1991**, *35*, 127-205.
11. Warner, I. M.; Christian, G. D.; Davidson, E. R.; Callis, J. B., Analysis of Multicomponent Fluorescence Data. *Analytical Chemistry* **1977**, *49* (4), 564-573.
12. Patra, D.; Mishra, A. K., Recent developments in multi-component synchronous fluorescence scan analysis. *Trac-Trends In Analytical Chemistry* **2002**, *21* (12), 787-798.
13. 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.
14. 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.
15. 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), 037001.
16. Casamayou-Boucau, Y.; Ryder, A. G., Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR). *Analytica Chimica Acta* **2018**, *1000*, 132-143.
17. Uversky, V. N.; Garriques, L. N.; Millett, I. S.; Frokjaer, S.; Brange, J.; Doniach, S.; Fink, A. L., Prediction of the association state of insulin using spectral parameters. *J Pharm Sci* **2003**, *92* (4), 847-58.

18. Tantipolphan, R.; Romeijn, S.; Engelsman, J.; Torosantucci, R.; Rasmussen, T.; Jiskoot, W., Elution behavior of insulin on high-performance size exclusion chromatography at neutral pH. *Journal of pharmaceutical and biomedical analysis* **2010**, *52* (2), 195-202.
19. Lakowicz, J. R., Chapter 1. Introduction to fluorescence. In *Principles of Fluorescence Spectroscopy, 3rd Edition*, Springer, Ed. 2006.
20. Ghisaidoobe, A. B.; 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.
21. Valeur, B.; Berberan-Santos, M. N., A Brief History of Fluorescence and Phosphorescence before the Emergence of Quantum Theory. *Journal of Chemical Education* **2011**, *88* (6), 731-738.
22. Stokes, G. G., On the change of refrangibility of light. *Philosophical Transactions of the Royal Society of London* **1852**, *142*, 463-562.
23. Valeur, B., Chapter 1. Basic principles of fluorescence spectroscopy. In *Molecular Fluorescence: Principles and Applications*, Sons, J. W. a., Ed. Wiley-VCH Verlag GmbH: 2002.
24. Roberts, J. D.; Caserio, M., Chapter 6. Bonding in organic molecules. Atomic-orbital models. In *Basic principles of organic chemistry, second edition*, Roberts, J. D.; Caserio, M., Eds. W. A. Benjamin, Inc.: 1977.
25. Valeur, B., Chapter 3. Characteristics of fluorescence emission. In *Molecular Fluorescence: Principles and Applications*, Sons, J. W. a., Ed. Wiley-VCH Verlag GmbH: 2002.
26. Liu, J.; Liu, C.; He, W., Fluorophores and Their Applications as Molecular Probes in Living Cells. *Current Organic Chemistry* **2013**, *17* (6), 564-579.
27. Hawe, A.; Sutter, M.; Jiskoot, W., Extrinsic fluorescent dyes as tools for protein characterization. *Pharmaceutical Research* **2008**, *25* (7), 1487-1499.
28. Lakowicz, J. R., Chapter 16. Protein fluorescence. In *Principles of Fluorescence Spectroscopy, 3rd Edition*, Springer, Ed. 2006.
29. Lakowicz, J. R., Chapter 2. Instrumentation for fluorescence spectroscopy. In *Principles of Fluorescence Spectroscopy, 3rd Edition*, Springer, Ed. 2006.
30. Redmond, R. W., Introduction to fluorescence and photophysics. In *Handbook of Biomedical Fluorescence*, Mycek, M.-A.; Pogue, B. W., Eds. Marcel Dekker, Inc.: 2003.
31. Eaton, D. F., Reference materials for fluorescence measurement. *Pure and Applied Chemistry* **1988**, *60* (1107-1114).
32. Fasman, G. D.; Sober, H. A., *Handbook of biochemistry and molecular biology*. CRC Press Cleveland: 1977; Vol. 4.
33. Noronha, M.; Lima, J. C.; Lamosa, P.; Santos, H.; Maycock, C.; Ventura, R.; Maçanita, A. L., Intramolecular Fluorescence Quenching of Tyrosine by the Peptide  $\alpha$ -Carbonyl Group Revisited. *The Journal of Physical Chemistry A* **2004**, *108* (12), 2155-2166.
34. Callis, P. R.; Liu, T., Quantitative Prediction of Fluorescence Quantum Yields for Tryptophan in Proteins. *The Journal of Physical Chemistry B* **2004**, *108* (14), 4248-4259.
35. Cowgill, R. W., Fluorescence and protein structure. XVII. On the mechanism of peptide quenching. *Biochimica et biophysica acta* **1970**, *200* (1), 18-25.
36. Cowgill, R. W., Fluorescence and protein structure: XI. Fluorescence quenching by disulphide and sulfhydryl groups. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1967**, *140* (1), 37-44.
37. Weisenborn, P. C. M.; Meder, H.; Egmond, M. R.; Visser, T. J. W. G.; van Hoek, A., Photophysics of the single tryptophan residue in *Fusarium solani* Cutinase: Evidence for the occurrence of conformational substates with unusual fluorescence behaviour. *Biophysical chemistry* **1996**, *58* (3), 281-288.

38. Alcalá, J. R.; Gratton, E.; Prendergast, F. G., Fluorescence lifetime distributions in proteins. *Biophysical journal* **1987**, *51* (4), 597-604.
39. Swaminathan, R.; Krishnamoorthy, G.; Periasamy, N., Similarity of fluorescence lifetime distributions for single tryptophan proteins in the random coil state. *Biophysical journal* **1994**, *67* (5), 2013-2023.
40. Rolinski, O. J.; Amaro, M.; Birch, D. J., Early detection of amyloid aggregation using intrinsic fluorescence. *Biosens Bioelectron* **2010**, *25* (10), 2249-52.
41. Rolinski, O. J.; Wellbrock, T.; Birch, D. J.; Vyshemirsky, V., Tyrosine Photophysics During the Early Stages of beta-Amyloid Aggregation Leading to Alzheimer's. *J Phys Chem Lett* **2015**, *6* (15), 3116-20.
42. Gally, J. A.; Edelman, G. M., The effect of temperature on the fluorescence of some aromatic amino acids and proteins. *Biochimica et biophysica acta* **1962**, *60* (3), 499-509.
43. Valeur, B., Chapter 4. Effects of intermolecular photophysical processes on fluorescence emission. In *Molecular Fluorescence: Principles and Applications*, Sons, J. W. a., Ed. Wiley-VCH Verlag GmbH: 2002.
44. Lakowicz, J. R., Chapter 8. Quenching of fluorescence. In *Principles of Fluorescence Spectroscopy, 3rd Edition*, Springer, Ed. 2006.
45. Möller, M.; Denicola, A., Protein tryptophan accessibility studied by fluorescence quenching. *Biochemistry and Molecular Biology Education* **2002**, *30* (3), 175-178.
46. Eftink, M. R.; Ghiron, C. A., Fluorescence quenching studies with proteins. *Analytical Biochemistry* **1981**, *114* (2), 199-227.
47. Kayser, V.; Turton, D. A.; Aggeli, A.; Beevers, A.; Reid, G. D.; Beddard, G. S., Energy migration in novel pH-triggered self-assembled beta-sheet ribbons. *Journal of the American Chemical Society* **2004**, *126* (1), 336-343.
48. Dusa, A.; Kaylor, J.; Edridge, S.; Bodner, N.; Hong, D. P.; Fink, A. L., Characterization of oligomers during alpha-synuclein aggregation using intrinsic tryptophan fluorescence. *Biochemistry* **2006**, *45* (8), 2752-2760.
49. Wu, P. G.; Brand, L., Resonance Energy Transfer: Methods and Applications. *Analytical Biochemistry* **1994**, *218* (1), 1-13.
50. Piston, D. W.; Kremers, G. J., Fluorescent protein FRET: the good, the bad and the ugly. *Trends in biochemical sciences* **2007**, *32* (9), 407-414.
51. Braslavsky, S. E., Glossary of terms used in photochemistry, 3rd edition (IUPAC Recommendations 2006). *Pure and Applied Chemistry* **2007**, *79* (3), 293-465.
52. Teale, F. W. J., The ultraviolet fluorescence of proteins in neutral solution. *Biochemical Journal* **1960**, *76* (2), 381-388.
53. Weber, G., Fluorescence-polarization spectrum and electronic-energy transfer in proteins. *Biochemical Journal* **1960**, *75* (2), 345-352.
54. Valeur, B., Chapter 9. Resonance energy transfer and its applications. In *Molecular Fluorescence: Principles and Applications*, Sons, J. W. a., Ed. Wiley-VCH Verlag GmbH: 2002.
55. Teoh, C. L.; Bekard, I. B.; Asimakis, P.; Griffin, M. D.; Ryan, T. M.; Dunstan, D. E.; Howlett, G. J., Shear flow induced changes in apolipoprotein C-II conformation and amyloid fibril formation. *Biochemistry* **2011**, *50* (19), 4046-4057.
56. Nath, S.; Meuis, J.; Hendrix, J.; Carl, S. A.; Engelborghs, Y., Early aggregation steps in alpha-synuclein as measured by FCS and FRET: evidence for a contagious conformational change. *Biophysical journal* **2010**, *98* (7), 1302-1311.
57. Boteva, R.; Zlateva, T.; Dorovska-Taran, V.; Visser, A. J.; Tsanev, R.; Salvato, B., Dissociation equilibrium of human recombinant interferon gamma. *Biochemistry* **1996**, *35* (47), 14825-14830.

58. Valeur, B., Chapter 6. Principles of steady-state and time-resolved fluorometric techniques. In *Molecular Fluorescence: Principles and Applications*, Sons, J. W. a., Ed. Wiley-VCH Verlag GmbH: 2002.
59. Parker, C. A.; Rees, W. T., Fluorescence spectrometry. A review. *Analyst* **1962**, 83-111.
60. Luciani, X.; Mounier, S.; Redon, R.; Bois, A., A simple correction method of inner filter effects affecting FEEM and its application to the PARAFAC decomposition. *Chemometrics And Intelligent Laboratory Systems* **2009**, 96 (2), 227-238.
61. 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.
62. Gu, Q.; Kenny, J. E., Improvement of inner filter effect correction based on determination of effective geometric parameters using a conventional fluorimeter. *Analytical Chemistry* **2009**, 81 (1), 420-426.
63. Wang, T.; Zeng, L.-H.; Li, D.-L., A review on the methods for correcting the fluorescence inner-filter effect of fluorescence spectrum. *Applied Spectroscopy Reviews* **2017**, 52 (10), 883-908.
64. 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: 2008.
65. Lakhali, L.; Acha, V.; Aussenac, T., PARAFAC analysis of front-face fluorescence data: Absorption and scattering effects assessed by means of Monte Carlo simulations. *Chemometrics and Intelligent Laboratory Systems* **2012**, 116, 112-122.
66. Gil, D. B.; de la Pena, A. M.; Arancibia, J. A.; Escandar, G. M.; Olivieri, A. C., Second-order advantage achieved by unfolded-partial least-squares/residual bilinearization modeling of excitation-emission fluorescence data presenting inner filter effects. *Analytical Chemistry* **2006**, 78 (23), 8051-8058.
67. Calvet, C.; Li, B.; Ryder, A. G., Rapid quantification of tryptophan and tyrosine in chemically defined cell culture media using fluorescence spectroscopy. *Journal of pharmaceutical and biomedical analysis* **2012**, 71, 89-98.
68. Kalman, B.; Clarke, N.; Johansson, L. B. A., Dynamics of a new fluorescent probe, 2,5,8,11-tetra-tert-butylperylene in solution, cubic lyotropic liquid crystals, and model membranes. *The Journal of Physical Chemistry* **1989**, 93 (11), 4608-4615.
69. Attri, A. K.; Fernandez, C.; Minton, A. P., pH-dependent self-association of zinc-free insulin characterized by concentration-gradient static light scattering. *Biophysical chemistry* **2010**, 148 (1-3), 28-33.
70. Attri, A. K.; Fernandez, C.; Minton, A. P., Self-association of Zn-insulin at neutral pH: investigation by concentration gradient--static and dynamic light scattering. *Biophys Chem* **2010**, 148 (1-3), 23-7.
71. Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V. N.; Fink, A. L., Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. *Biochemistry* **2001**, 40 (20), 6036-46.
72. Nielsen, L.; Frokjaer, S.; Brange, J.; Uversky, V. N.; Fink, A. L., Probing the mechanism of insulin fibril formation with insulin mutants. *Biochemistry* **2001**, 40 (28), 8397-409.
73. Ahmad, A.; Uversky, V. N.; Hong, D.; Fink, A. L., Early events in the fibrillation of monomeric insulin. *J Biol Chem* **2005**, 280 (52), 42669-75.
74. Brange, J.; Skelbæk-Pedersen, B.; Langkjær, L.; Damgaard, U.; Ege, H.; Havelund, S.; Heding, L. G.; Jørgensen, K. H.; Lykkeberg, J.; Markussen, J.; Pingel, M.; Rasmussen, E., C.

- Insulin Preparations. In *Galenics of insulin: the physico-chemical and pharmaceutical aspects of insulin and insulin preparations.*, Brange, J., Ed. Springer-Verlag Berlin Heidelberg: 1987.
75. Manoharan, C.; Singh, J., Addition of zinc improves the physical stability of insulin in the primary emulsification step of the poly(lactide-co-glycolide) microsphere preparation process. *Polymers* **2015**, *7* (5), 836-850.
76. Porter, R. R., Partition chromatography of insulin and other proteins. *Biochemical Journal* **1953**, *53* (2), 320-328.
77. Monographs. In *The Merck Index: An Encyclopedia of Chemicals, Drugs and Biologicals, Fourteenth edition*, O'Neil, M. J.; Heckelman, P. E.; Koch, C. B.; Roman, K. J., Eds. Merck & Co., Inc.: 2006.
78. Lakowicz, J. R., Chapter 10. Fluorescence anisotropy. In *Principles of Fluorescence Spectroscopy*, 3rd edition ed.; Springer: 2006.
79. Jameson, D. M., Chapter 5. Polarization and Anisotropy. In *Introduction to Fluorescence*, Taylor & Francis: 2014.
80. Perrin, F., Sur le mouvement brownien de rotation. *Comptes rendus* **1925**, *181*, 514-516.
81. Perrin, F., Theorie de la fluorescence polarizee (influence de la viscosite). *Comptes rendus* **1925**, *180*, 581-583.
82. Azumi, T.; McGlynn, S. P., Polarization of the luminescence of phenanthrene. *Journal of Chemical Physics* **1962**, *37*, 2413.
83. Jameson, D. M.; Croney, J. C.; Moens, P. D. J., Fluorescence: Basic Concepts, Practical Aspects, and Some Anecdotes. *Methods in Enzymology* **2003**, *360*, 1-43.
84. Kessel, D., Artifacts in Fluorescence Emission Spectroscopy Related to Wood's Anomaly. *Photochemistry and Photobiology* **1991**, *54* (3), 481-483.
85. Jablonski, A., On the notion of emission anisotropy. *Bulletin de l'Académie polonaise des sciences. Série des sciences mathématiques, astronomiques, et physiques.* **1960**, *8*, 259-264.
86. Jameson, D. M.; Ross, J. A., Fluorescence polarization/anisotropy in diagnostics and imaging. *Chemical Review* **2010**, *110* (5), 2685-2708.
87. Perrin, F., Polarisation de la lumiere de fluorescence. Vie moyenne des molecules dans l'etat excite. *Journal de Physique et Le Radium* **1926**, *7*, 390-401.
88. Perrin, F., La fluorescence des solutions: induction moleculaire. Polarisation et duree d'emission. Photochimie. *Annals of Physics* **1929**, *12*, 169-275.
89. Debye, P., Zur Theorie der anomalen Dispersion im Gebiete der langwelligen elektrischen Strahlung. . *Verhandlungen der Deutschen Physikalischen Gesellschaft* **1913**, *15*, 738-777.
90. Alexiev, U.; Farrens, D. L., Fluorescence spectroscopy of rhodopsins: insights and approaches. *Biochimica et biophysica acta* **2014**, *1837* (5), 694-709.
91. Holowka, D.; Wensel, T.; Baird, B., A nanosecond fluorescence depolarization study on the segmental flexibility of receptor-bound immunoglobulin E. *Biochemistry* **1990**, *29*, 4607-4612.
92. Hanson, D. C.; Yguerabide, J.; Schumaker, V. N., Segmental flexibility of Immunoglobulin G antibody Molecules in Solution: a New Interpretation. *Biochemistry* **1981**, *20*, 6842-6852.
93. Kawato, S.; Kinoshita, K.; Ikegami, A., Dynamic structure of lipid bilayers studied by nanosecond fluorescence techniques. *Biochemistry* **1977**, *16*, 2319-2324.
94. Kawato, S.; Kinoshita, K.; Ikegami, A., Effect of cholesterol on the molecular motion in the hydrocarbon region of lecithin bilayers studied by nanosecond fluorescence techniques. *Biochemistry* **1978**, *17*, 5026-5031.

95. Whal, P.; Paoletti, J.; Le Pecq, J.-B., Decay of fluorescence emission anisotropy of the ethidium bromide-DNA complex evidence for an internal motion in DNA. *Proceedings of the National Academy of Sciences of the United States of America* **1970**, *65* (2), 417-421.
96. Duhamel, J.; Kanyo, J.; Dinter-Gottlieb, G.; Lu, P., Fluorescence emission of ethidium bromide intercalated in defined DNA duplexes: evaluation of hydrodynamics components. *Biochemistry* **1996**, *35* (51), 16687-16697.
97. Deprez, E.; Tauc, P.; Leh, H.; Mouscadet, J.; Auclair, C.; Hawkins, M.; Brochon, J., DNA binding induces dissociation of the multimeric form of HIV-1 integrase: a time-resolved fluorescence anisotropy study. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98* (18), 10090-10095.
98. Yengo, C. M.; Berger, C. L., Fluorescence anisotropy and resonance energy transfer: powerful tools for measuring real time protein dynamics in a physiological environment. *Current Opinion in Pharmacology* **2010**, *10*, 731-737.
99. Weber, G., Polarization of the fluorescence of macromolecules. I. Theory and experimental method. *Biochemical Journal* **1952**, *51*, 145-155.
100. Weber, G., Polarization of the fluorescence of macromolecules. II. Fluorescent conjugates of ovalbumin and bovine serum albumin. *Biochemical Journal* **1952**, *51*, 155-167.
101. Pappas, D.; Burrows, S. M.; Randall, D. R., Exploring biomolecular interactions by single-molecule fluorescence. *TrAC Trends in Analytical Chemistry* **2007**, *26* (9), 884-894.
102. Pallicer, J. M.; Kramer, S. D., Evaluation of fluorescence anisotropy to assess drug-lipid membrane partitioning. *Journal of pharmaceutical and biomedical analysis* **2012**, *71*, 219-227.
103. Dandliker, W. B.; Kelly, R. J.; Dandliker, J., Fluorescence polarization immunoassay. Theory and experimental method. *Immunochemistry* **1973**, *10* (4), 219-227.
104. Guo, X.-G.; Castellano, F. N.; Li, L.; Lakowiz, J. R., Use of Long-Lifetime Re(I) Complex in fluorescence polarization immunoassays of High-Molecular-Weight Analytes. *Analytical Chemistry* **1998**, *70*, 632-637.
105. Owicki, J. C., Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. *Journal of Biomolecular Screening* **2000**, *5* (5), 297-306.
106. Spencer, R. D.; Toledo, F. B.; Williams, B. T.; Yoss, N. L., Fluorescence polarization assay of protease. *Clinical Chemistry* **1973**, *19*, 838-844.
107. Maeda, H., Assay of proteolytic enzymes by fluorescence polarization technique. *Analytical Biochemistry* **1979**, *92*, 222-227.
108. Cleemann, F.; Karuso, P., Fluorescence anisotropy assay for the traceless kinetic analysis of protein digestion. *Analytical Chemistry* **2008**, *80*, 4170-4174.
109. Bolger, R.; Checovich, W., A new protease activity assay using fluorescence polarization. *Biotechniques* **1994**, *17* (3), 585-589.
110. Levine, L. M.; Michener, M. L.; Toth, M. V.; Holwerda, B. C., Measurement of specific rotease activity utilizing fluorescence polarization. *Analytical Biochemistry* **1997**, *247* (1), 83-88.
111. Coffin, J.; Latev, M.; Bi, X.; Nikirov, T. T., Detection of phosphopeptides by fluorescence polarization in the presence of cationic polyamino acids: application to kinase assays. *Analytical Biochemistry* **2000**, *278*, 206-212.
112. Fowler, A.; Swift, D.; Longman, E.; Acornley, A.; Hemsley, P.; Murray, D.; Unitt, J.; Dale, I.; Sullivan, E.; Coldwell, M., An evaluation of fluorescence polarization and lifetime discriminated polarization for high throughput screening of serine/threonine kinases. *Analytical Biochemistry* **2002**, *308* (2), 223-231.
113. Seethala, R.; Menzel, R., A fluorescence polarization competition immunoassay for tyrosine kinases. *Analytical Biochemistry* **1998**, *255*, 257-262.

114. Zaman, G. J.; Garritsen, A.; De Boer, T.; Van Boeckel, C. A., Fluorescence assays for high-throughput screening of protein kinases. *Combinatorial Chemistry & High Throughput Screening* **2003**, *6* (4), 313-320.
115. Lea, W. A.; A., S., Fluorescence polarization assays in small molecule screening. *Expert Opinion on Drug Discovery* **2011**, *6* (1), 17-32.
116. Millar, R. P.; Newton, C. L., The year in G protein-coupled receptor research. *Molecular Endocrinology* **2010**, *24* (1), 261-274.
117. Thompson, R. B.; Maliwal, B. P.; Fierke, C. A., Expanded dynamic range of free zinc ion determination by fluorescence anisotropy. *Analytical Chemistry* **1998**, *70* (9), 1749-1754.
118. Lakowicz, J. R.; Gryczynski, I.; Gryczynski, Z.; Tolosa, L.; Randers-Eichhorn, L.; Rao, G., Polarization-Based sensing of glucose using an oriented reference film. *Journal of Biomedical Optics* **1999**, *4* (4), 443-449.
119. Axelrod, D., Fluorescence polarization microscopy. *Methods in Cell Biology* **1989**, *30*, 333-352.
120. Axelrod, D., Carbocyanine dye orientation in red cell membrane studied by microscopic fluorescence polarization. *Biophysical journal* **1979**, *26* (3), 557-573.
121. Gough, A. H.; Taylor, D. L., Fluorescence anisotropy imaging microscopy maps calmodulin binding during cellular contraction and locomotion. *Journal of Cell Biology* **1993**, *121* (5), 1095-1107.
122. Dix, J. A.; Verkman, A. S., Mapping of fluorescence anisotropy in living cells by ratio imaging. Application to cytoplasmic viscosity. *Biophysical journal* **1990**, *57* (2), 231-240.
123. Fushimi, K.; Dix, J. A.; Verkman, A. S., Cell membrane fluidity in the intact kidney proximal tubule measured by orientation-independent fluorescence anisotropy imaging. *Biophysical journal* **1990**, *57* (2), 241-254.
124. Yaroslavsky, A. N.; Neel, V.; Anderson, R. R., Fluorescence polarization imaging for delineating nonmelanoma skin cancers. *Optics Letter* **2004**, *29* (17), 2010-2012.
125. Pu, Y.; Wang, W. B.; Das, B. B.; Alfano, R. R., Time-resolved spectral wing emission kinetics and optical imaging of human cancerous and normal prostate tissues. *Optics Communications* **2009**, *282* (21), 4308-4314.
126. Zhang, G.; Demos, S. G.; Alfano, R. R., Far-red and NIR spectral wing emission from tissues under 532 and 632 nm photo-excitation. *Lasers in the life science* **1999**, *9* (1), 1-16.
127. Mc Ginty, J.; Dunsby, C.; Auksoorius, E.; P., B. R. K.; De Beule, P.; Elson, D. S.; Galletly, N.; Grant, D.; Hofmann, O.; Kennedy, G.; Kumar, S.; Lanigan, P. M. P., Chapter 4 Multidimensional fluorescence imaging. In *FRET and FLIM techniques*, Elsevier B.V.: 2009; Vol. 33, pp 133-169.
128. Yan, L.; Rueden, C. T.; White, J. G.; Eliceiri, K. W., Applications of combined spectral lifetime microscopy for biology. *Biotechniques* **2006**, *41* (3), 249-253.
129. Welch, C. M.; Elliott, H.; Danuser, G.; Hahn, K. M., Imaging the coordination of multiple signalling activities in living cells. *Nature Reviews Molecular Cell Biology* **2011**, *12* (749-756).
130. Roberts, M. S.; Dancik, Y.; Prow, T. W.; Thorling, C. A.; Lin, L. L.; Grice, J. E.; Robertson, T. A.; Konig, K.; Becker, W., Non-invasive imaging of skin physiology and percutaneous penetration using fluorescence spectral and lifetime imaging with multiphoton and confocal microscopy. *European journal of pharmaceuticals and biopharmaceutics : official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik e.V* **2011**, *77* (3), 469-88.
131. Booksh, K. S.; Kowalski, B. R., Theory of Analytical Chemistry. *Analytical Chemistry* **2012**, *66* (15), 782A-791A.

132. Li, B.; Ryan, P. W.; Shanahan, M.; Leister, K. J.; Ryder, A. G., Fluorescence EEM Spectroscopy for Rapid Identification and Quality Evaluation of Cell Culture Media Components. *Applied spectroscopy* **2011**, *65* (11), 1240-1249.
133. 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.
134. Burstein, E. A.; Vedenkin, N. S.; Ivkova, M. N., Fluorescence and the Location of Tryptophan Residues in Protein Molecules. *Photochemistry and Photobiology* **1973**, *18* (4), 263-279.
135. Ryan, P. W.; Li, B.; Shanahan, M.; Leister, K. J.; Ryder, A. G., Prediction of Cell Culture Media Performance Using Fluorescence Spectroscopy. *Analytical Chemistry* **2010**, *82* (4), 1311-1317.
136. Georgakoudi, I.; Tsai, I.; Greiner, C.; Wong, C.; Defelice, J.; Kaplan, D., Intrinsic fluorescence changes associated with the conformational state of silk fibroin in biomaterial matrices. *Optics Express* **2007**, *15* (3), 1043-1053.
137. Zhongqi, H.; Uchimiya, M.; Cao, H., Intrinsic Fluorescence Excitation-Emission matrix spectral features of cottonseed protein fractions and the effects of denaturants. *Journal of the American Oil Chemists' Society* **2014**, *91* (9), 1489-1497.
138. Obeidat, S. M.; Glasser, T.; Landau, S. Y.; Anderson, D. M.; Rayson, G. D., Application of multi-way data analysis on excitation-emission spectra for plant identification *Talanta* **2007**, *72*, 682-690.
139. Smilde, A.; Bro, R.; Geladi, P., *Multi-way analysis with applications in the chemical sciences*. John Wiley & Sons, Inc.: Chichester, 2004.
140. da Silva, J. C. G. E.; Tauler, R., Multivariate curve resolution of synchronous fluorescence spectra matrices of fulvic acids obtained as a function of pH. *Applied spectroscopy* **2006**, *60* (11), 1315-1321.
141. Antunes, M. C. G.; da Silva, J., Multivariate curve resolution analysis excitation-emission matrices of fluorescence of humic substances. *Analytica Chimica Acta* **2005**, *546* (1), 52-59.
142. Jiang, J. H.; Ozaki, Y., Self-modeling curve resolution (SMCR): Principles, techniques, and applications. *Applied Spectroscopy Reviews* **2002**, *37* (3), 321-345.
143. de Juan, A.; Tauler, R., Multivariate curve resolution (MCR) from 2000: Progress in concepts and applications. *Critical Reviews In Analytical Chemistry* **2006**, *36* (3-4), 163-176.
144. 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.
145. Andersen, C.; Bro, R., Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics* **2003**, *17* (4), 200-215.
146. Bro, R., PARAFAC. Tutorial and applications. *Chemometrics And Intelligent Laboratory Systems* **1997**, *38* (2), 149-171.
147. Amigo, J. M.; Marini, F., Chapter 7 - Multiway Methods. In *Data Handling in Science and Technology*, Federico, M., Ed. Elsevier: 2013; Vol. Volume 28, pp 265-313.
148. Hall, G. J.; Clow, K. E.; Kenny, J. E., Estuarial fingerprinting through multidimensional fluorescence and multivariate analysis. *Environmental Science & Technology* **2005**, *39* (19), 7560-7567.
149. Hall, G. J.; Kenny, J. E., Estuarine water classification using EEM spectroscopy and PARAFAC-SIMCA. *Analytica Chimica Acta* **2007**, *581* (1), 118-124.
150. Stedmon, C. A.; Bro, R., Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnology And Oceanography-Methods* **2008**, *6*, 572-579.

151. Stedmon, C. A.; Markager, S., Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnology and Oceanography* **2005**, *50* (2), 686-697.
152. Ohno, T.; Bro, R., Dissolved organic matter characterization using multiway spectral decomposition of fluorescence landscapes. *Soil Science Society of America Journal* **2006**, *70* (6), 2028-2037.
153. Ohno, T.; Amirbahman, A.; Bro, R., Parallel factor analysis of excitation-emission matrix fluorescence spectra of water soluble soil organic matter as basis for the determination of conditional metal binding parameters. *Environmental Science & Technology* **2008**, *42* (1), 186-192.
154. Saurina, J.; Leal, C.; Compano, R.; Granados, M.; Tauler, R.; Prat, M. D., Determination of triphenyltin in sea-water by excitation-emission matrix fluorescence and multivariate curve resolution. *Analytica Chimica Acta* **2000**, *409* (1-2), 237-245.
155. SádeCka, J.; ToThoVa, J., Fluorescence Spectroscopy and Chemometrics in the Food Classification- a Review. *Czech Journal of Food Sciences* **2007**, *25* (4), 159-173.
156. Callejon, R. M.; Amigo, J. M.; Pairo, E.; Garmon, S.; Ocana, J. A.; Morales, M. L., Classification of sherry vinegars by combining multidimensional fluorescence, PARAFAC and different classification approaches. *Talanta* **2012**, *88*, 456-462.
157. Calvet, A.; Li, B.; Ryder, A. G., A rapid fluorescence based method for the quantitative analysis of cell culture media photo-degradation. *Analytica Chimica Acta* **2014**, *807* (0), 111-119.
158. Destrampe, K. A.; Hieftje, G. M., New Instrumentation for Use in Excitation-Emission Fluorescence-Polarization Measurements. *Applied spectroscopy* **1993**, *47* (10), 1548-1554.
159. Nishiyama, S.; Tajima, M.; Yoshida, Y., Excitation Emission Matrix of Spin-coated Polymeric Thin films and its application to Polarized Fluorescence. *Journal of Photopolymer Science and Technology* **2006**, *19* (1), 21-28.
160. 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.
161. Gonzalez Flecha, F. L.; Levi, V., Determination of the molecular size of BSA by fluorescence anisotropy. *Biochemistry and Molecular Biology Education* **2003**, *31* (5), 319-322.
162. Garcia Sanchez, F.; Navas Diaz, A.; Lopez Guerrero, M. M., Simultaneous determination of glycols based on fluorescence anisotropy. *Analytica Chimica Acta* **2007**, *582* (1), 92-97.
163. Groenzin, H.; Mullins, O. C., Molecular size and structure of asphaltenes. *Petroleum Science and Technology* **2001**, *19* (1-2), 219-230.
164. Mazzeo, B. A.; Busath, D. D., From molecular dynamics to fluorescence anisotropy of fluorophores bound to oriented structures. *Journal of Computational Physics* **2013**, *232* (1), 482-497.
165. Barkley, M. D.; Kowalczyk, A. A.; Brand, L., Fluorescence decay studies of anisotropic rotations of small molecules. *The Journal of Chemical Physics* **1981**, *75*, 3581-3593.
166. Balter, A.; Szubiakowski, J., Fluorescence probes of viscosity: a comparative study of the fluorescence anisotropy decay of perylene and 3,9-Dibromoperylene in glycerol. *Journal of Fluorescence* **1993**, *3* (4), 247-249.
167. Goldstein, D. H., Chapter 23. Polarization Optical Elements. In *Polarized Light, third edition*, Press, C., Ed. Taylor & Francis Group: 2011.
168. Weber, T.; Fuchs, H.-J.; Schmidt, H.; Kley, E.-B.; Tunnermann, A., Wire-grid polarizer for the UV spectral region. *Proceedings of SPIE* **2009**, *7205*.
169. Pang, Y. T.; Meng, G. W.; Zhang, Y.; Fang, Q.; Zhang, L. D., Copper nanowire arrays for infrared polarizer. *Applied Physics A* **2003**, *76*, 533-536.

170. Resch-Genger, U.; Derose, P. C., Fluorescence standards: Classification, terminology, and recommendations on their selection, use, and production (IUPAC Technical Report). *Pure and Applied Chemistry* **2010**, *82* (12), 2315-2335.
171. Thompson, R. B.; Gryczynski, I.; Malicka, J., Fluorescence Polarization Standards for High-Throughput Screening and Imaging. *Biotechniques* **2002**, *32* (1), 34-42.
172. Valeur, B., Chapter 2. Absorption of UV-Visible Light. In *Molecular Fluorescence: Principles and Applications*, Sons, J. W. a., Ed. Wiley-VCH Verlag GmbH: 2002.
173. Kasha, M., Characterization of electronic transitions in complex molecules. *Discussions of the Faraday Society* **1950**, *9*, 14-19.
174. Prazeres, T. J. V.; Fedorov, A.; Barbosa, S. P.; Martinho, J. M. G.; Berberan-Santos, M. N., Accurate Determination of the Limiting Anisotropy of Rhodamine 101. Implications for Its Use as a Fluorescence Polarization Standard. *The Journal of Physical Chemistry A* **2008**, *112*, 5034-5039.
175. Wagner, B. D.; Tittelback-Helmrich, D.; Steer, R. P., Radiationless decay of the S2 states of azulene and related compounds: solvent dependance and the energy gap law. *Journal of Physical Chemistry* **1992**, *96*, 7904-7908.
176. Plummer, B. F.; Al-Saigh, Z. Y., Electronic excited states and anomalous fluorescence of Cyclopenta[cd]pyrene. *Journal of Physical Chemistry* **1983**, *87*, 1579-1582.
177. Huber, J. R.; Mahaney, M., S2 -> S0 fluorescence in an aromatic thioetone, xanthione. *Chemical Physics Letters* **1975**, *30* (3), 410-412.
178. Itoh, T., Fluorescence and Phosphorescence from Higher Excited States of Organic Molecules. *Chemical reviews* **2012**, *112*, 4541-4568.
179. Kohler, A.; Bassler, H., *Electronic Processes in Organic Semiconductors, an Introduction*. Wiley-VCH: 2015.
180. Mazurenko, Y. T.; Bakhshiev, N. G., Effect of orientation dipole relaxation on spectral, time, and polarization characteristics of the luminescence of solutions. *Optics and Spectroscopy* **1970**, *28*, 490-494.
181. Matayoshi, E. D.; Kleinfeld, A. M., Emission wavelength-dependant decay of the 9-anthroyloxy-fatty acid membrane probes. *Biophysical journal* **1981**, *35* (1), 215-235.
182. Szubiakowski, J.; Balter, A.; Nowak, W.; Kowalczyk, A.; Wisniewski, K.; Wierzbowska, M., Anisotropic reorientation of perylene and 3,9-dibromoperylene in glycerol: fluorescence anisotropy decay and quantum-mechanical study. *Chemical Physics* **1996**, *208*, 283-296.
183. Lakowicz, J. R., Chapter 12. Advanced anisotropy concepts. In *Principles of Fluorescence Spectroscopy, 3rd Edition*, Springer, Ed. 2006.
184. Johansson, J. B.-A., Limiting Fluorescence Anisotropies of Perylene and Xanthene Derivatives. *Journal of the Chemical Society, Faraday Transactions* **1990**, *86* (12), 2103-2107.
185. Prazeres, T. J. V.; Fedorov, A.; Barbosa, S. P.; Martinho, J. M. G.; Berberan-Santos, M. N., Accurate determination of the limiting anisotropy of rhodamine 101. Implications for its use as a fluorescence polarization standard. *Journal of Physical Chemistry A* **2008**, *112* (23), 5034-5039.
186. Pantke, E. R.; Labhart, H., Influence of excess excitation energy on rotational relaxation of molecules in solution. *Chemical Physics Letters* **1973**, *23* (4), 476-481.
187. Lawaetz, A. J.; Stedmon, C. A., Fluorescence Intensity Calibration Using the Raman Scatter Peak of Water. *Applied spectroscopy* **2009**, *63* (8), 936-940.
188. Zhang, J. Z.; Grant, C. D., Chapter 1. Optical and dynamic properties of undoped and doped semiconductor nanostructures. In *Annual Review of Nano Research*, Cao, G.; Brinker, C. J., Eds. World Scientific: 2008; Vol. 2.
189. Bohren, C. F.; Huffman, D., *Absorption and scattering of light by small particles*. John Wiley: 1983.

190. Mie, G., On the optics of turbid media, special colloidal metallic solutions. *Annalen Der Physik* **1908**, *25*, 377-445.
191. Tyndall, J., On the blue color of the sky, the polarization of sky light, and on the polarization of light by cloudy matter generally. *Philosophical Magazine* **1869**, *37*, 384-394.
192. Tyndall, J., Notes on the formation and phenomena of clouds. *Philosophical Magazine* **1869**, *38*, 156-158.
193. Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R., Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Analytical Methods* **2013**, *5* (23), 6557-6566.
194. Matthews, B. J. H.; Jones, A. C.; Theodorou, N. K.; Tudhope, A. W., Excitation-emission-matrix fluorescence spectroscopy applied to humic acid bands in coral reefs. *Marine Chemistry* **1996**, *55*, 317-332.
195. Bennett, J. M., Part 1. Optical Elements. Chapter 3: Polarizers. In *Handbook of Optics. Second Edition. Volume II: Devices, Measurements, and Properties*, Bass, M., Ed. McGraw-Hill Professional: 1994.
196. Sharma, K. K., *Optics: Principles and Applications*. Academic Press / Elsevier: 2006.
197. Lipson, A.; Lipson, S. G.; Lipson, H., *Optical Physics, 4th edition*. Cambridge University Press: 2011.
198. Dorband, B.; Muller, H.; Gross, H., Metrology of Optical Components and Systems. In *Handbook of Optical Systems*, Gross, H., Ed. Wiley-VCH: 2012; Vol. 5.
199. Yeh, P.; Gu, C., Chapter 3. Electromagnetic propagation in anisotropic media. In *Optics of liquid crystal displays, second edition*, Boreman, G. D., Ed. Wiley: 2010.
200. Yang, D.-K.; Wu, S.-T., Chapter 2. Propagation of Light in Anisotropic Optical Media. In *Fundamentals of liquid crystal devices, second edition*, Wiley: 2015.
201. Bennett, J. M., Part 3. Polarized Light. Chapter 12. Polarization. In *Handbook of Optics. Third Edition. Volume I. Geometrical and Physical Optics, Polarized Light, Components and Instruments*, Bass, M., Ed. McGraw-Hill: 2010.
202. Dai, M.; Wan, W.; Zhu, X.; Song, B.; Liu, X.; Lu, M.; Cui, B.; Chen, Y., Broadband and wide angle infrared wire-grid polarizer. *Optics Express* **2015**, *23* (12), 15390-15397.
203. Lane, L. B., Freezing points of glycerol and its aqueous solutions. *Industrial and Engineering Chemistry* **1925**, *17* (9), 924-.
204. Gregory, S. R., Chapter 6. Physical properties of glycerine. In *Glycerine, a key cosmetic ingredient*, Jungermann, E.; Sonntag, N. O. V., Eds. Marcel Dekker Inc: 1991.
205. Ryabov, Y. E.; Hayashi, Y.; Gutina, A.; Feldman, Y., Features of supercooled glycerol dynamics. *Physical review B* **2003**, *67*, 132002.
206. Zondervan, R.; Kulzer, F.; Berkhout, G. C. G.; Orrit, M., Local viscosity of supercooled glycerol near T<sub>g</sub> probed by rotational diffusion of ensembles and single dye molecules. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104* (31), 12628-12633.
207. Kalogerias, I. M.; Hagg Lobland, H. E., The nature of the glassy state: structure and glass transitions. *Journal of Materials Education* **2012**, *34*, 69-94.
208. Angell, C. A., Relaxation in liquids, polymers and plastic crystals - strong/fragile patterns and problems. *Journal of Non-Crystalline Solids* **1991**, *131-133*, 13-31.
209. Richert, R.; Angell, C. A., Dynamics of glass-forming liquids. V. On the link between molecular dynamics and configurational entropy. *Journal of Chemical Physics* **1998**, *108* (21), 9016-9026.
210. Schroter, K.; Donth, E., Viscosity and shear response of the dynamic glass transition of glycerol. *Journal of Chemical Physics* **2000**, *113* (20), 9101-9108.
211. Wuttke, J.; Chang, I.; Fujara, F.; Petry, W., Viscous glycerol. *Physica B: Condensed Matter* **1997**, *234-236*, 431-432.

212. Sokolov, A. P.; Steffen, W.; Rossler, E., Dynamics of glass-forming liquids at high temperatures: a general scenario. *Journal of Physics: Condensed Matter* **1996**, *8*, 9587-9591.
213. Wuttke, J.; Hernandez, J.; Li, G.; Coddens, G.; Cummins, Z.; Fujara, F.; Petry, W.; Sillescu, H., Neutron and light scattering study of supercooled glycerol. *Physical Review Letters* **1994**, *72* (19), 3052-3055.
214. Lunkenheimer, P.; Loidl, A., Dielectric spectroscopy of glass-forming materials: alpha-relaxation and excess wing. *Chemical Physics* **2002**, *284* (1-2), 205-219.
215. Vera, L.; Czarny, B.; Georgiadis, D.; Dive, V.; Stura, E. A., Practical use of glycerol in protein crystallization. *Crystal Growth Design* **2011**, *11* (7), 2755-2762.
216. Van Koningsveld, H., The crystal structure of glycerol and its conformation. *Recueil des Travaux Chimiques des Pays-Bas* **1968**, *87* (3), 243-254.
217. Duarte, P.; Ferreira, D. P.; Ferreira Machado, I.; Vieira Ferreira, L. F.; Rodriguez, H. B.; Sam Roman, E., Phloxine B as a probe for entrapment in microcrystalline cellulose. *Molecules* **2012**, *17*, 1602-1616.
218. Koziar, J. C.; Cowan, D. O., Photochemical heavy-atom effects. *Accounts of Chemical Research* **1977**, 334-341.
219. Lakowicz, J. R., Chapter 9. Mechanisms and dynamics of fluorescence quenching. In *Principles of Fluorescence Spectroscopy, 3rd Edition*, Springer, Ed. 2006.
220. Fleming, G. R.; Knight, A. W. E.; Morris, J. M.; S., M. R. J.; Robinson, G. W., Picosecond fluorescence studies of xanthene dyes. *Journal of the American Chemical Society* **1977**, *99* (13), 4306-4311.
221. Mann, T. L.; Krull, U. J., Fluorescence polarization spectroscopy in protein analysis. *Analyst* **2003**, *128*, 313-317.
222. Rinnan, Å.; 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.
223. Tauler, R., Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *Journal of Chemometrics* **2001**, *15* (8), 627-646.
224. Bro, R.; Smilde, A. K., Principal component analysis. *Analytical Methods* **2014**, *6*, 2812-2831.
225. Abdi, H.; Williams, L. J., Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* **2010**, *2* (4), 433-459.
226. Massart, D.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics: Part A*. Elsevier: Amsterdam, 1997.
227. Gemperline, P. J., Chapter 4. Principal component analysis. In *Practical guide to chemometrics, second edition*, Gemperline, P. J., Ed. Taylor and Francis: 2006.
228. Tauler, R.; Smilde, A.; Kowalski, B., Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *Journal of Chemometrics* **1995**, *9* (1), 31-58.
229. Tauler, R.; Kowalski, B., Multivariate Curve Resolution applied to spectra data from multiple runs of an industrial process. *Analytical Chemistry* **1993**, *65*, 2040-2047.
230. Craig Hamilton, J.; Gemperline, P. J., Mixture analysis using factor analysis. II: Self-modeling curve resolution. *Journal of Chemometrics* **1990**, *4* (1), 1-13.
231. Lawton, W. H.; Sylvestre, E. A., Self Modeling Curve Resolution. *Technometrics* **1971**, *13* (3), 617-633.
232. Osten, D. W.; Kowalski, B. R., Multivariate Curve Resolution in Liquid Chromatography. *Analytical Chemistry* **1984**, *56*, 991-995.
233. Borgen, O. S.; Kowalski, B. R., An extension of the multivariate component-resolution method to three components. *Analytica Chimica Acta* **1985**, *174*, 1-26.

234. Tauler, R.; Barcelo, D., Multivariate curve resolution applied to liquid chromatography - diode array detection. *TrAC Trends in Analytical Chemistry* **1993**, *12* (8), 319-327.
235. Bortolato, S. A.; Olivieri, A. C., Chemometric processing of second-order liquid chromatographic data with UV-vis and fluorescence detection. A comparison of multivariate curve resolution and parallel factor analysis 2. *Analytica Chimica Acta* **2014**, *842*, 11-19.
236. Amigo, J. M.; Skov, T.; Bro, R., ChroMATHography: solving chromatographic issues with mathematical models and intuitive graphics. *Chemical reviews* **2010**, *110* (8), 4582-4605.
237. De Juan, A.; Jaumot, J.; Tauler, R., Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Analytical Methods* **2014**, *6*, 4964-4976.
238. Azzouz, T.; Tauler, R., Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples. *Talanta* **2008**, *74* (5), 1201-10.
239. Kessler, W.; Kessler, R. W., Multivariate curve resolution: a method of evaluating the kinetics of biotechnological reactions. *Analytical and bioanalytical chemistry* **2006**, *384* (5), 1087-1095.
240. Tauler, R.; De Juan, A., Chapter 5. Multivariate Curve Resolution for Quantitative Analysis. In *Fundamentals and Analytical Applications of Multiway Calibration*, de la Pena, A. M.; Goicoechea, H. C.; Escandar, G. M.; Olivieri, A. C., Eds. Elsevier: Amsterdam, 2015; Vol. 29.
241. Joliffe, I. T., *Principal Component Analysis*. Springer-Verlag: New York, 1986.
242. Beer, A., Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten *Annalen der Physik und Chemie* **1852**, *162*, 78-88.
243. Schenone, A. V.; de Araujo Gomes, A.; Culzoni, M. J.; Campiglia, A. D.; de Araujo, M. C.; Goicoechea, H. C., Modeling nonbilinear total synchronous fluorescence data matrices with a novel adapted partial least squares method. *Analytica Chimica Acta* **2015**, *859*, 20-28.
244. Jaumot, J.; Tauler, R.; Gargallo, R., Exploratory data analysis of DNA microarrays by multivariate curve resolution. *Analytical Biochemistry* **2006**, *358* (1), 76-89.
245. Jones, H. D. T.; Haaland, D. M.; Sinclair, M. B.; Melgaard, D. K.; Collins, A. M.; Timlin, J. A., Preprocessing strategies to improve MCR analyses of hyperspectral images. *Chemometrics and Intelligent Laboratory Systems* **2012**, *117*, 149-158.
246. de Juan, A.; Maeder, M.; Hanczewicz, T.; Duponchel, L.; Tauler, R., Chapter 2. Chemometric Tools for Image Analysis. In *Infrared and Raman Spectroscopic Imaging*, Salzer, R.; Siesler, H. W., Eds. Wiley-VCH: 2009.
247. Zhang, X.; de Juan, A.; Tauler, R., Multivariate Curve Resolution applied to hyperspectral imaging analysis of Chocolate samples. *Applied spectroscopy* **2015**, *69* (8), 993-1003.
248. Amigo, J. M.; Cruz, J.; Bautista, M.; Maspoch, S.; Coello, J.; Blanco, M., Study of pharmaceutical samples by NIR chemical-image and multivariate analysis. *TrAC Trends in Analytical Chemistry* **2008**, *27* (8), 696-713.
249. Kvalheim, O. M.; Liang, Y. Z., Heuristic evolving latent projections: resolving two-way multicomponent data. 1. Selectivity, latent-projective graph, datascope, local rank, and unique resolution. *Analytical Chemistry* **1992**, *64* (8), 936-946.
250. Malinowski, E. R., Window factor analysis: Theoretical derivation and application to flow injection analysis data. *Journal of Chemometrics* **1992**, *6*, 29-40.
251. Manne, R.; Shen, H.; Liang, Y., Subwindow factor analysis. *Chemometrics and Intelligent Laboratory Systems* **1999**, *45*, 171-176.
252. de Juan, A.; Rutan, S. C.; Tauler, R., Two-Way Data Analysis: Multivariate Curve Resolution - Iterative Resolution Methods. In *Comprehensive Chemometrics*, Brown, S.; Tauler, R.; Walczak, B., Eds. Elsevier: 2009.

253. Vandeginste, B. G. M.; Derks, W.; Kateman, G., Multicomponent self-modeling curve resolution in high performance liquid chromatography by iterative target transformation factor analysis. *Analytica Chimica Acta* **1985**, *173*, 253-264.
254. Gemperline, P. J., A priori estimates of the elution profiles of the pure components in overlapped liquid chromatography peaks using target factor analysis. *Journal of Chemical Information and Computer Sciences* **1984**, *24*, 206-212.
255. Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R., A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometrics and Intelligent Laboratory Systems* **2005**, *76* (1), 101-110.
256. Tauler, R., Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems* **1995**, *30* (1), 133-146.
257. Gemperline, P., *Practical Guide to Chemometrics*. Second ed.; Taylor and Francis: 2006.
258. Zhu, Z.-L.; Cheng, W.-Z.; Zhao, Y., Iterative target transformation factor analysis for the resolution of kinetic-spectral data with an unknown kinetic model. *Chemometrics and Intelligent Laboratory Systems* **2002**, *64* (2), 157-167.
259. Ruckebusch, C.; Blanchet, L., Multivariate curve resolution: a review of advanced and tailored applications and challenges. *Analytica Chimica Acta* **2013**, *765*, 28-36.
260. Maeder, M., Evolving Factor Analysis for the resolution of overlapping chromatographic peaks. *Analytical Chemistry* **1987**, *59*, 517-530.
261. Windig, W.; Guilment, J., Interactive Self-Modeling Mixture Analysis. *Analytical Chemistry* **1991**, *63*, 1425-1432.
262. Maeder, M.; Zuberbuehler, A. D., The resolution of overlapping chromatographic peaks by evolving factor analysis. *Analytica Chimica Acta* **1986**, *181*, 287-291.
263. Keller, H. R.; Massart, D. L., Peak purity control in liquid chromatography with photodiode array detection by fixed-size-moving-window-evolving factor analysis. *Analytica Chimica Acta* **1991**, *246*, 379-390.
264. Goicoechea, H. C.; Olivieri, A. C.; Tauler, R., Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data. *Analyst* **2010**, *135* (3), 636-42.
265. Lozano, V. A.; Ibanez, G. A.; Olivieri, A. C., A novel second-order standard addition analytical method based on data processing with multidimensional partial least-squares and residual bilinearization. *Analytica Chimica Acta* **2009**, *651* (2), 165-172.
266. Schenone, A. V.; Culzoni, M. J.; Campiglia, A. D.; Goicoechea, H. C., Total synchronous fluorescence spectroscopic data modeled with first- and second-order algorithms for the determination of doxorubicin in human plasma. *Analytical and bioanalytical chemistry* **2013**, *405* (26), 8515-23.
267. Rajko, R.; Istvan, K., Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry. *Journal of Chemometrics* **2006**, *19* (8), 448-463.
268. Sawall, M.; Kubis, C.; Selent, D.; Borner, A.; Neymeyr, K., A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. I: concepts and applications. *Journal of Chemometrics* **2013**, *27* (5), 106-116.
269. Beyramysoltan, S.; Rajko, R.; Abdollahi, H., Investigation of the equality constraint effect on the reduction of the rotational ambiguity in three-component system using a novel grid search method. *Analytica Chimica Acta* **2013**, *791*, 25-35.
270. Jaumot, J.; Tauler, R., MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution. *Chemometrics and Intelligent Laboratory Systems* **2010**, *103* (2), 96-107.

271. de Juan, A.; Tauler, R., Chemometrics applied to unravel multicomponent processes and mixtures: Revisiting latest trends in multivariate resolution. *Analytica Chimica Acta* **2003**, *500* (1-2), 195-210.
272. de Juan, A.; Vander Heyden, Y.; Tauler, R.; Massart, D. L., Assessment of new constraints applied to the alternating least squares method. *Analytica Chimica Acta* **1997**, *346*, 307-318.
273. Bro, R.; Jong, S., A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics* **1997**, *11*, 393-401.
274. Bro, R.; Nicholaos, D., Least squares algorithms under unimodality and non-negativity constraints. *Journal of Chemometrics* **1998**, *12* (4), 223-247.
275. Tauler, R.; Izquierdo-Ridorsa, A.; Casassas, E., Simultaneous analysis of several spectroscopic titrations with self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems* **1993**, *18*, 293-300.
276. Mendieta, J.; Diaz-Cruz, M. S.; Esteban, M.; Tauler, R., Multivariate Curve Resolution: a possible tool in the detection of intermediate structures in protein folding. *Biophysical journal* **1998**, *74* (6).
277. Navea, S.; de Juan, A.; Tauler, R., Detection and resolution of intermediate species in protein folding processes using fluorescence and circular dichroism spectroscopies and multivariate curve resolution. *Analytical Chemistry* **2002**, *74*, 6031-6039.
278. Gampp, H.; Maeder, M.; Meyer, C. J.; Zuberbuehler, A. D., Quantification of a known component in an unknown mixture. *Analytica Chimica Acta* **1987**, *193*, 287-293.
279. Tauler, R.; Maeder, M.; de Juan, A., Multiset Data Analysis: Extended Multivariate Curve Resolution. In *Comprehensive Chemometrics*, Brown, S.; Tauler, R.; Walczak, B., Eds. Elsevier: 2009.
280. de Juan, A.; Maeder, M.; Martínez, M.; Tauler, R., Combining hard- and soft-modelling to solve kinetic problems. *Chemometrics and Intelligent Laboratory Systems* **2000**, *54* (2), 123-141.
281. Mas, S.; de Juan, A.; Lacorte, S.; Tauler, R., Photodegradation study of decabromodiphenyl ether by UV spectrophotometry and a hybrid hard- and soft-modelling approach. *Analytica Chimica Acta* **2008**, *618* (1), 18-28.
282. Manouchehri, F.; Izadmanesh, Y.; Aghaee, E.; Ghasemi, J. B., Experimental, computational and chemometrics studies of BSA-vitamin B6 interaction by UV-Vis, FT-IR, fluorescence spectroscopy, molecular dynamics simulation and hard-soft modeling methods. *Bioorganic chemistry* **2016**, *68*, 124-136.
283. Garrido, M.; Rius, F. X.; Larrechi, M. S., Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Analytical and bioanalytical chemistry* **2008**, *390* (8), 2059-66.
284. Diewok, J.; de Juan, A.; Maeder, M.; Tauler, R.; Lendl, B., Application of a combination of hard and soft modeling for equilibrium systems to the quantitative analysis of pH-modulated mixture samples. *Analytical Chemistry* **2003**, *75* (3), 641-647.
285. Smilde, A.; Bro, R.; Geladi, P., *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons: 2005.
286. Navea, S.; de Juan, A.; Tauler, R., Modeling Temperature-Dependent Protein Structural Transitions by Combined Near-IR and Mid-IR Spectroscopies and Multivariate Curve Resolution. *Analytical Chemistry* **2003**, *75* (20), 5592-5601.
287. Navea, S.; de Juan, A.; Tauler, R., Three-way data analysis applied to multispectroscopic monitoring of protein folding. *Analytica Chimica Acta* **2001**, *446* (1-2), 185-195.
288. Tauler, R.; Kowalski, B.; Fleming, S., Multivariate curve resolution applied to spectra data from multiple runs of an industrial process. *Analytical Chemistry* **1993**, *65*, 2040-2047.

289. Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R., Fluorescence spectroscopy and multi-way techniques. *PARAFAC. Analytical Methods* **2013**, *5*, 6557-6566.
290. Jaumot, J.; Escaja, N.; Gargallo, R.; Gonzelez, C.; Pedroso, E.; Tauler, R., Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids. *Nucleic Acids Research* **2002**, *30* (17), e92.
291. Kroonenberg, P. M.; de Leeuw, J., Principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **1980**, *45*, 69-97.
292. ten Berge, J. M. F.; de Leeuw, J.; Kroonenberg, P. M., Some additional results on principal components analysis of three mode data by means of alternating least squares. *Psychometrika* **1987**, *52*, 183-191.
293. Acar, E.; Yener, B., Unsupervised multiway data analysis: a literature survey. *IEEE Transactions on Knowledge and Data Engineering* **2008**, *21* (1).
294. 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.
295. de Juan, A.; Tauler, R., Comparison of three-way resolution methods for non-trilinear chemical data sets. *Journal of Chemometrics* **2001**, *15* (10), 749-771.
296. Olivieri, A. C., Chapter 14. Application Example: U-PLS/RBL. In *Practical three-way calibration*, Elsevier: Escandar, G. M.
297. de Araujo Gomes, A.; Schenone, A. V.; Goicoechea, H. C.; de Araujo, M. C. U., Unfolded partial least squares/residual bilinearization combined with the Successive Projections Algorithm for interval selection: enhanced excitation-emission fluorescence data modeling in the presence of the inner filter effect. *Analytical Bioanalytical Chemistry* **2015**, *407*, 5649-5659.
298. Piccirilli, G. N.; Escandar, G. M., Partial least-squares with residual bilinearization for the spectrofluorimetric determination of pesticides. A solution of the problems of inner-filter effects and matrix interferences. *Analyst* **2006**, *131*, 1012-1020.
299. Kompany-Zareh, M.; Gholami, S., Soft and hard multiway FRET-based investigation of interaction between drug and QD labeled DNA. *Chemometrics and Intelligent Laboratory Systems* **2014**, *139*, 33-41.
300. Gholami, S.; Kompany-Zareh, M., Targeting human c-Myc promoter duplex DNA with actinomycin D by use of multi-way analysis of quantum-dot-mediated fluorescence resonance energy transfer. *Analytical and bioanalytical chemistry* **2013**, *405* (19), 6271-6280.
301. Förster, T., Intermolecular energy migration and fluorescence. *Annalen der Physik* **1948**, *437* (1), 55-75.
302. Ji, R. D.; Booksh, K. S., Mitigation of Rayleigh and Raman spectral interferences in multiway calibration of excitation-emission matrix fluorescence spectra. *Analytical Chemistry* **2000**, *72* (4), 718-725.
303. Rinnan, A.; Andersen, C. M., Handling of first-order Rayleigh scatter in PARAFAC modelling of fluorescence excitation-emission data. *Chemometrics And Intelligent Laboratory Systems* **2005**, *76* (1), 91-99.
304. 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.
305. Wentzell, P. D.; Nair, S. S.; Guy, R. D., Three-way analysis of fluorescence spectra of polycyclic aromatic hydrocarbons with quenching by nitromethane. *Analytical Chemistry* **2001**, *73*, 1408-1415.
306. 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.

307. Kunitani, M.; Wolfe, S.; Rana, S.; Apicella, C.; Levi, V.; Dollinger, G., Classical light scattering quantitation of protein aggregates: off-line spectroscopy versus HPLC detection. *Journal of pharmaceutical and biomedical analysis* **1997**, *16* (4), 573-586.
308. Dollinger, G.; Cunico, B.; Kunitani, M.; Johnson, D.; Jones, R., Practical on-line determination of biopolymer molecular weights by high-performance liquid chromatography with classical light-scattering detection. *Journal of Chromatography A* **1992**, *592* (1-2), 215-228.
309. Andersen, C. M.; Bro, R., Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics* **2003**, *17* (4), 200-215.
310. Harshman, R. A.; Lundy, M. E., PARAFAC: Parallel factor analysis. *Computational Statistics & Data Analysis* **1994**, *18* (1), 39-72.
311. 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.
312. 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) dataset of dilute aqueous solutions of fluorophores. *Chemometrics and Intelligent Laboratory Systems* **2012**, *116*, 78-86.
313. Pacheco, M. E.; Bruzzone, L., Synchronous fluorescence spectrometry: Conformational investigation or inner filter effect? *Journal of Luminescence* **2013**, *137*, 138-142.
314. Jaumot, J.; de Juan, A.; Tauler, R., MCR-ALS GUI 2.0: New features and applications. *Chemometrics and Intelligent Laboratory Systems* **2015**, *140*, 1-12.
315. Wall, M. E.; Rechtsteiner, A.; Rocha, L. M., Chapter 5. Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis*, Berrar, D. P.; Dubitzky, W.; Granzow, M., Eds. Springer: 2003.
316. Windig, W.; Guilment, J., Interactive self-modeling mixture analysis. *Analytical Chemistry* **1991**, *63* (14), 1425-1432.
317. Li, B.; Ryder, A. G. Similarity index: a rapid classification method for multivariate data arrays. 2009.
318. Gaiduk, A.; Ruijgrok, P. V.; Yorulmaz, M.; Orrit, M., Making gold nanoparticles fluorescent for simultaneous absorption and fluorescence detection on the single particle level. *Physical Chemistry Chemical Physics* **2001**, *13*, 149-153.
319. Freire, S.; de Araujo, M. H.; Al-Soufi, W.; Novo, M., Photophysical study of Thioflavin T as fluorescence marker of amyloid fibrils. *Dyes and Pigments* **2014**, *110*, 97-105.
320. Gafni, A.; Brand, L., Fluorescence decay studies of reduced nicotinamide adenine dinucleotide in solution and bound to liver alcohol dehydrogenase. *Biochemistry* **1976**, *15* (15), 3165-3171.
321. Matczuk, A.; Bojarski, P.; Gryczyński, I.; Kuśba, J.; Kułak, L.; Bojarski, C., The influence of water structure on the rotational depolarization of fluorescence. *Journal of Photochemistry and Photobiology A: Chemistry* **1995**, *90* (2-3), 91-94.
322. Smyth, S.; Heron, A., Diabetes and obesity: the twin epidemics. *Nature medicine* **2006**, *12* (1), 75-80.
323. Montague, W., 3. Insulin Synthesis, Storage and Secretion. In *Diabetes and the Endocrine Pancreas: A Biochemical Approach*, Montague, W., Ed. Croom Helm Ltd: 1983.
324. Chan, S. J.; Cao, Q. P.; Steiner, D. F., Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87* (23), 9319-9323.

325. Adams, M. J.; Blundell, T. L.; Dodson, E. J.; Dodson, G. G.; Vijayan, M.; Baker, E. N.; Harding, M. M.; Hodgkin, D. C.; Rimmer, B.; Sheat, S., Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **1969**, *224* (5218), 491-495.
326. Blundell, T. L.; Cutfield, J. F.; Cutfield, S. M.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. C.; Mercola, D. A., Three-Dimensional Atomic Structure of Insulin and Its Relationship to Activity. *Diabetes* **1972**, *21* (Supplement\_2), 492-505.
327. Baker, E. N.; Blundell, T. L.; Cutfield, J. F.; Cutfield, S. M.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. M. C.; Hubbard, R. E.; Isaacs, N. W.; Reynolds, C. D.; Sakabe, K.; Sakabe, N.; Vijayan, N. M., The Structure of 2Zn Pig Insulin Crystals at 1.5 Å Resolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **1988**, *319* (1195), 369-456.
328. De Meyts, P., Insulin and its receptor: structure, function and evolution. *BioEssays : news and reviews in molecular, cellular and developmental biology* **2004**, *26* (12), 1351-62.
329. Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D., Insulin: The Structure in the Crystal and its Reflection in Chemistry and Biology by. *Advances in Protein Chemistry* **1972**, *26*, 279-402.
330. Li, Y. V., Zinc and insulin in pancreatic beta-cells. *Endocrine* **2014**, *45* (2), 178-89.
331. Rorsman, P.; Renstrom, E., Insulin granule dynamics in pancreatic beta cells. *Diabetologia* **2003**, *46* (8), 1029-45.
332. Howell, S. L.; Fink, C. J.; Lacy, P. E., Isolation and properties of secretory granules from rat islets of Langerhans. I. Isolation of a secretory granule fraction. *The Journal of Cell Biology* **1969**, *41* (1), 154-161.
333. Blundell, T. L.; Cutfield, J. F.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. C.; Mercola, D. A., The crystal structure of rhombohedral 2 Zinc Insulin. *Cold Spring Harbor Symposia on Quantitative Biology* **1972**, *36*, 233-241.
334. Derewenda, U.; Derewenda, Z. S.; Godson, G. G.; Hubbard, R. E., Insulin Structure. In *Insulin*, Cuatrecasas, P.; Jacobs, S., Eds. Springer: 1990.
335. Chothia, C.; Lesk, A. M.; Dodson, G. G.; Hodgkin, D. C., Transmission of conformational change in insulin. *Nature* **1983**, *302*, 500-505.
336. Knegt, R. M. A.; Boelens, R.; Ganadu, M. L.; Kaptein, R., The solution structure of a monomeric insulin. A two-dimensional 1H-NMR study of des-(B26-B30)-insulin in combination with distance geometry and restrained molecular dynamics. *European Journal of Biochemistry* **1991**, *202* (2), 447-458.
337. Dodson, E. J.; Dodson, G. G.; Lewitova, A.; Sabesan, M., Zinc-free cubic pig insulin: Crystallization and structure determination. *Journal of Molecular Biology* **1978**, *125* (3), 387-396.
338. Badger, J.; Haris, M. R.; Reynolds, C. D.; Evans, A. C.; Dodson, E. J.; Dodson, G. G.; North, A. C. T., Structure of the pig insulin dimer in the cubic crystal. *Acta Crystallographica Section B* **1991**, *47* (1), 127-136.
339. Cutfield, J. F.; Cutfield, S. M.; Dodson, E. J.; Dodson, G. G.; Emdin, S. F.; Reynolds, C. D., Structure and biological activity of hagfish insulin. *Journal of Molecular Biology* **1979**, *132* (1), 85-100.
340. Fu, Z.; Gilbert, E. R.; Liu, D., Regulation of Insulin Synthesis and Secretion and Pancreatic Beta-Cell Dysfunction in Diabetes. *Current Diabetes Reviews* **2012**, *9* (1), 25-53.
341. Weiss, M.; Steiner, D. F.; Philipson, L. H., Insulin Biosynthesis, Secretion, Structure, and Structure-Activity Relationships. In *Endotext*, De Groot, L. J.; Chrousos, G.; Dungan, K.; Feingold, K. R.; Grossman, A.; Hershman, J. M.; Koch, C.; Korbonits, M.; McLachlan, R.; New, M.; Purnell, J.; Rebar, R.; Singer, F.; Vinik, A., Eds. 2014.

342. Blundell, T. L.; Pitts, J. E.; Wood, S. P., The Conformation and Molecular Biology of Pancreatic Hormones and Homologous Growth Factor. *Critical Reviews in Biochemistry* **1982**, *13* (2), 141-213.
343. Krüger, P.; Straßburger, W.; Wollmer, A.; van Gunsteren, W. F.; Dodson, G. G., The simulated dynamics of the insulin monomer and their relationship to the molecule's structure. *European Biophysics Journal* **1987**, *14* (8), 449-459.
344. Wodak, S. J.; Alard, P.; Delhaise, P.; Renneboog-Squilbin, C., Simulation of conformational changes in 2 Zn insulin. *Journal of Molecular Biology* **1985**, *181* (2), 317-322.
345. Schmitz, O.; Rungby, J.; Edge, L.; Juhl, C. B., On high-frequency insulin oscillations. *Ageing research reviews* **2008**, *7* (4), 301-305.
346. Jonassen, I.; Havelund, S.; Hoeg-Jensen, T.; Steensgaard, D. B.; Wahlund, P. O.; Ribel, U., Design of the novel protraction mechanism of insulin degludec, an ultra-long-acting basal insulin. *Pharmaceutical Research* **2012**, *29*, 2104-2114.
347. Dodson, G.; Steiner, D., The role of assembly in insulin's biosynthesis. *Current Opinion in Structural Biology* **1998**, *8* (2), 189-194.
348. Pertusa, J. A. G.; Leon-Quinto, T.; Berna, G.; Tejedo, J. R.; Hmadcha, A.; Bedoya, F. J.; Martin, F.; Soria, B., Zn<sup>2+</sup> chelation by serum albumin improves hexameric Zn<sup>2+</sup>-insulin dissociation into monomers after exocytosis. *PLoS one* **2017**, *12* (11), e0187547.
349. Frank, B. H.; Pekar, A. H.; Veros, A. J., Insulin and Proinsulin Conformation in Solution. *Diabetes* **1972**, *21* (Supplement 2), 486-491.
350. Bi, R. C.; Dauter, Z.; Dodson, E.; Dodson, G.; Giordano, F.; Reynolds, C. D., Insulin's structure as a modified and monomeric molecule. *Biopolymers* **1984**, *23* (3), 391-395.
351. Hou, J. C.; Min, L.; Pessin, J. E., Insulin Granule Biogenesis, Trafficking and Exocytosis. *Vitamins and Hormones* **2009**, *80*, 473-506.
352. McDonald, J. M.; Pershadsingh, H. A., The Role of Calcium in the Transduction of Insulin Action. In *Molecular Basis of Insulin Action*, Czech, M. P., Ed. Springer: 1985; pp 103-118.
353. Montague, W., 4. Insulin Action. In *Diabetes and the Endocrine Pancreas: A Biochemical Approach*, Montague, W., Ed. Croom Helm Ltd: 1983.
354. Kurtzhals, P.; Ribel, U., Action Profile of Cobalt(III)-Insulin: A Novel Principle of Protraction of Potential Use for Basal Insulin Delivery. *Diabetes* **1995**, *44* (12), 1381-1385.
355. Pullen, R. A.; Lindsay, D. G.; Wood, S. P.; Tickle, I. J.; Blundell, T. L.; Wollmer, A.; Krail, G.; Brandenburg, D.; Zahn, H.; Gliemann, J.; Gammeltoft, S., Receptor-binding region of insulin. *Nature* **1976**, *259*, 369-373.
356. Menting, J. G.; Whittaker, J.; Margetts, M. B.; Whittaker, L. J.; Kong, G. K.; Smith, B. J.; Watson, C. J.; Zakova, L.; Kletvikova, E.; Jiracek, J.; Chan, S. J.; Steiner, D. F.; Dodson, G. G.; Brzozowski, A. M.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C., How insulin engages its primary binding site on the insulin receptor. *Nature* **2013**, *493* (7431), 241-245.
357. Kaarsholm, N. C.; Ludvigsen, S., The high resolution solution structure of the insulin monomer determined by NMR. *Receptor* **1995**, *5* (1), 1-8.
358. Whittingham, J. L.; Scott, D. J.; Chance, K.; Wilson, A.; Finch, J.; Brange, J.; Dodson, G., Insulin at pH 2: Structural Analysis of the Conditions Promoting Insulin Fibre Formation. *Journal of Molecular Biology* **2002**, *318* (2), 479-490.
359. Hua, Q. X.; Shoelson, S. E.; Kochoyan, M.; Weiss, M. A., Receptor binding redefined by a structural switch in a mutant human insulin. *Nature* **1991**, *354* (6350), 238-41.
360. Goldman, J.; Carpenter, F. H., Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives. *Biochemistry* **1974**, *13* (22), 4566-4574.

361. Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R., Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology* **1984**, *179* (1), 125-142.
362. Ward, C. W.; Lawrence, M. C., Landmarks in insulin research. *Frontiers in endocrinology* **2011**, *2*, 76.
363. Xu, B.; Hua, Q. X.; Nakagawa, S. H.; Jia, W.; Chu, Y. C.; Katsoyannis, P. G.; Weiss, M. A., A cavity-forming mutation in insulin induces segmental unfolding of a surrounding alpha-helix. *Protein science : a publication of the Protein Society* **2002**, *11* (1), 104-16.
364. Hua, Q. X.; Jia, W.; Frank, B. H.; Weiss, M. A., Comparison of the dynamics of an engineered insulin monomer and dimer by acid-quenched amide proton exchange. Non-local stabilization of interchain hydrogen bonds by dimerization. *Journal of molecular biology* **1993**, *230* (2), 387-94.
365. Wan, Z. L.; Huang, K.; Hu, S. Q.; Whittaker, J.; Weiss, M. A., The structure of a mutant insulin uncouples receptor binding from protein allostery. An electrostatic block to the TR transition. *The Journal of biological chemistry* **2008**, *283* (30), 21198-210.
366. Hua, Q. X.; Nakagawa, S.; Hu, S. Q.; Jia, W.; Wang, S.; Weiss, M. A., Toward the active conformation of insulin: stereospecific modulation of a structural switch in the B chain. *The Journal of biological chemistry* **2006**, *281* (34), 24900-9.
367. Hua, Q. X.; Weiss, M. A., Mechanism of insulin fibrillation: the structure of insulin under amyloidogenic conditions resembles a protein-folding intermediate. *The Journal of biological chemistry* **2004**, *279* (20), 21449-60.
368. Antolikova, E.; Zakova, L.; Turkenburg, J. P.; Watson, C. J.; Hanclova, I.; Sanda, M.; Cooper, A.; Kraus, T.; Brzozowski, A. M.; Jiracek, J., Non-equivalent role of inter- and intramolecular hydrogen bonds in the insulin dimer interface. *The Journal of biological chemistry* **2011**, *286* (42), 36968-77.
369. Zeng, Z.-H.; Liu, Y.-S.; Jin, L.; Zhang, Y.; Havelund, S.; Markussen, J.; Wang, D.-C., Conformational correlation and coupled motion between residue A21 and B25 side chain observed in crystal structures of insulin mutants at position A21. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **2000**, *1479* (1-2), 225-236.
370. Rhodes, C. J.; Shoelson, S. E.; Halban, P. A., 5. Insulin biosynthesis, processing, and chemistry. In *Joslin's Diabetes Mellitus*, Kahn, C. R.; LKing, G. L.; Moses, A. L.; Weir, G. C.; Jacobson, A. M.; Smith, R. J., Eds. Lippincott Williams & Wilkins: 2005.
371. Murray, J. E.; Laurieri, N.; Delgoda, R., Chapter 24 - Proteins. In *Pharmacognosy. Fundamentals, Applications and Strategies*, Badal, S.; Delgoda, R., Eds. Academic Press: 2017; pp 477-494.
372. Ciszak, E.; Beals, J. M.; Frank, B. H.; Baker, J. C.; Carter, N. D.; Smith, G. D., Role of C-terminal B-chain residues in insulin assembly: the structure of hexameric LysB28ProB29-human insulin. *Structure* **1995**, *3* (6), 615-22.
373. Lee, C. C.; Nayak, A.; Sethuraman, A.; Belfort, G.; McRae, G. J., A three-stage kinetic model of amyloid fibrillation. *Biophysical journal* **2007**, *92* (10), 3448-58.
374. Drablos, F., Clustering of non-polar contacts in proteins. *Bioinformatics* **1999**, *15* (6), 501-509.
375. Betz, S. F., Disulfide bonds and the stability of globular proteins. *Protein science : a publication of the Protein Society* **1993**, *2* (10), 1551-8.
376. Schlichtkrull, J.; Årtun, T.; Waterman, H. I.; Eliasson, N. A.; Thorell, B., Insulin Crystals. I. The Minimum Mole-Fraction of Metal in Insulin Crystals Prepared with Zn<sup>++</sup>, Cd<sup>++</sup>, Co<sup>++</sup>, Ni<sup>++</sup>, Cu<sup>++</sup>, Mn<sup>++</sup>, or Fe<sup>++</sup>. *Acta Chemica Scandinavica* **1956**, *10*, 1455-1458.
377. Dodson, E.; Harding, M. M.; Hodgkin, D. C.; Rossmann, M. G., The crystal structure of insulin. *Journal of Molecular Biology* **1966**, *16* (1), 227-241.

378. Bentley, G.; Dodson, E.; Dodson, G. U. Y.; Hodgkin, D.; Mercola, D. A. N., Structure of insulin in 4-zinc insulin. *Nature* **1976**, *261* (5556), 166-168.
379. Smith, G. D.; Blessing, R. H., Lessons from an aged, dried crystal of T6 human insulin. *Acta Crystallographica Section D Biological Crystallography* **2003**, *59* (8), 1384-1394.
380. Smith, G. D.; Pangborn, W. A.; Blessing, R. H., The structure of T6 human insulin at 1.0 Å resolution. *Acta Crystallographica Section D Biological Crystallography* **2003**, *59* (3), 474-482.
381. Krishna, N. R. S.; Pattabhi, V.; Rajan, S. S., Metal Induced Conformational Changes in Human Insulin: Crystal Structures of Sr<sup>2+</sup>, Ni<sup>2+</sup> and Cu<sup>2+</sup> Complexes of Human Insulin. *Protein & Peptide Letters* **2011**, *18* (5), 457-466.
382. Favero-Retto, M. P.; Palmieri, L. C.; Souza, T. A.; Almeida, F. C.; Lima, L. M., Structural meta-analysis of regular human insulin in pharmaceutical formulations. *European journal of pharmaceutics and biopharmaceutics : official journal of Arbeitsgemeinschaft fur Pharmazeutische Verfahrenstechnik e.V* **2013**, *85* (3 Pt B), 1112-1121.
383. Ciszak, E.; Smith, G. D., Crystallographic Evidence for Dual Coordination Around Zinc in the T3R3 Human Insulin Hexamer. *Biochemistry* **1994**, *33* (6), 1512-1517.
384. Smith, G. D.; Swenson, D. C.; Dodson, E. J.; Dodson, G. G.; Reynolds, C. D., Structural stability in the 4-zinc human insulin hexamer. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81*, 7093-7097.
385. Kaarsholm, N. C.; Ko, H. C.; Dunn, M. F., Comparison of solution structural flexibility and zinc binding domains for insulin, proinsulin, and miniproinsulin. *Biochemistry* **1989**, *28* (10), 4427-4435.
386. Weiss, M. A., The Structure and Function of Insulin: Decoding the TR Transition. *Vitamins and Hormones* **2009**, *80*, 33-49.
387. Bentley, G.; Dodson, G.; Lewitova, A., Rhombohedral insulin crystal transformation. *Journal of Molecular Biology* **1978**, *126* (4), 871-875.
388. Kosinova, L.; Veverka, V.; Novotna, P.; Collinsova, M.; Urbanova, M.; Moody, N. R.; Turkenburg, J. P.; Jiracek, J.; Brzozowski, A. M.; Zakova, L., Insight into the structural and biological relevance of the T/R transition of the N-terminus of the B-chain in human insulin. *Biochemistry* **2014**, *53* (21), 3392-3402.
389. Bryant, C.; Spencer, D. B.; Miller, A.; Bakaysa, D. L.; McCune, K. S.; Maple, S. R.; Pekar, A. H.; Brems, D. N., Acid stabilization of insulin. *Biochemistry* **1993**, *32* (32), 8075-82.
390. Whittingham, J. L.; Youshang, Z.; Zakova, L.; Dodson, E. J.; Turkenburg, J. P.; Brange, J.; Dodson, G. G., I222 crystal form of despentapeptide (B26-B30) insulin provides new insights into the properties of monomeric insulin. *Acta crystallographica. Section D, Biological crystallography* **2006**, *62* (Pt 5), 505-11.
391. Brange, J.; Andersen, L.; Laursen, E. D.; Meyn, G.; Rasmussen, E., Toward understanding insulin fibrillation. *Journal of pharmaceutical sciences* **1997**, *86* (5), 517-525.
392. Waugh, D. F., The linkage of corpuscular protein molecules. I. A fibrous modification of insulin. *Journal of the American Chemical Society* **1944**, *66*, 663.
393. Waugh, D. F., A Fibrous Modification of Insulin. I. The Heat Precipitate of Insulin. *Journal of the American Chemical Society* **1946**, *68* (2), 247-250.
394. Waugh, D. F.; Wilhelmson, D. F.; Commerford, S. L.; Sackler, M. L., Studies of the Nucleation and Growth Reactions of Selected Types of Insulin Fibrils. *Journal of the American Chemical Society* **1953**, *75* (11), 2592-2600.
395. Jean, L.; Lee, C. F.; Vaux, D. J., Enrichment of amyloidogenesis at an air-water interface. *Biophysical journal* **2012**, *102* (5), 1154-1162.
396. Sluzky, V.; Tamada, J. A.; Klibanov, A. M.; Langer, R., Kinetics of insulin aggregation in aqueous solutions upon agitation in the presence of hydrophobic surfaces. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88* (21), 9377-9381.

397. Choi, T. S.; Lee, J. W.; Jin, K. S.; Kim, H. I., Amyloid fibrillation of insulin under water-limited conditions. *Biophysical journal* **2014**, *107* (8), 1939-1949.
398. Grudzielanek, S.; Jansen, R.; Winter, R., Solvational tuning of the unfolding, aggregation and amyloidogenesis of insulin. *Journal of molecular biology* **2005**, *351* (4), 879-94.
399. Muzaffar, M.; Ahmad, A., The mechanism of enhanced insulin amyloid fibril formation by NaCl is better explained by a conformational change model. *PloS one* **2011**, *6* (11), e27906.
400. Ahmad, A.; Millett, I. S.; Doniach, S.; Uversky, V. N.; Fink, A. L., Partially folded intermediates in insulin fibrillation. *Biochemistry* **2003**, *42* (39), 11404-11416.
401. Rambaran, R. N.; Serpell, L. C., Amyloid fibrils: Abnormal protein assembly. *Prion* **2008**, *2* (3), 112-117.
402. Jeffrey, P. D.; Coates, J. H., An Equilibrium Ultracentrifuge Study of the Self-Association of Bovine Insulin. *Biochemistry* **1966**, *5* (2), 489-498.
403. Pekar, A. H.; Frank, B. H., Conformation of proinsulin. Comparison of insulin and proinsulin self-association at neutral pH. *Biochemistry* **1972**, *11* (22), 4013-4016.
404. Hvidt, S., Insulin association in neutral solutions studied by light scattering. *Biophysical chemistry* **1991**, *39* (2), 205-213.
405. Holladay, L. A.; Ascoli, M.; Puett, D., Conformational stability and self-association of zinc-free bovine insulin at neutral pH. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1977**, *494* (1), 245-254.
406. Jeffrey, P. D.; Milthorpe, B. K.; Nichol, L. W., Polymerization pattern of insulin at pH 7.0. *Biochemistry* **1976**, *15* (21), 4660-4665.
407. Pocker, Y.; Biswas, S. B., Self-association of insulin and the role of hydrophobic bonding: a thermodynamic model of insulin dimerization. *Biochemistry* **1981**, *20* (15), 4354-4361.
408. De Meyts, P.; Sajid, W.; Palsgaard, J.; Theede, A. M.; Gauguin, L.; Aladdin, H.; Whittaker, J., Insulin and IGF-I Receptor Structure and Binding Mechanism. In *Madame Curie Science database*, Landes Bioscience: 2013.
409. Baudyš, M.; Uchio, T.; Mix, D.; Kim, S. W.; Wilson, D., Physical Stabilization of Insulin by Glycosylation. *Journal of pharmaceutical sciences* **1995**, *84* (1), 28-33.
410. Derewenda, U.; Derewenda, Z.; Dodson, G. G.; Hubbard, R. E.; Korber, F., Molecular Structure of insulin: The insulin monomer and its assembly. *British Medical Bulletin* **1989**, *45* (1), 4-18.
411. Pocker, Y.; Biswas, S. B., Conformational dynamics of insulin in solution. Circular dichroic studies. *Biochemistry* **1980**, *19* (22), 5043-5049.
412. Hassiepen, U.; Stahl, J.; Motzka, B.; Federwisch, M., The association/dissociation equilibria of insulin in the presence of metal ions: a fluorescence energy transfer and circular dichroism study. In *Insulin & Related Proteins - Structure to Function and Pharmacology*, Federwisch, M.; Leyck Dieken, M.; De Meyts, P., Eds. Klumer academic: 2002; pp 41-51.
413. Wood, S. P.; Blundell, T. L.; Wollmer, A.; Lazarus, N. R.; Neville, R. W. J., The Relation of Conformation and Association of Insulin to Receptor Binding; X-Ray and Circular-Dichroism Studies on Bovine and Hystricomorph Insulins. *European Journal of Biochemistry* **1975**, *55* (3), 531-542.
414. Melberg, S. G.; Johnson, W. C. J., Changes in secondary structure follow the dissociation of human insulin hexamers: a circular dichroism study. *Proteins* **1990**, *8* (3), 280-286.
415. Rupley, J. A.; Renthal, R. D.; Praissman, M., Concentration difference spectra in the dimerization of insulin. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1967**, *140* (1), 185-187.

416. Lord, R. S.; Gubensek, F.; Rupley, J. A., Insulin self-association. Spectrum changes and thermodynamics. *Biochemistry* **1973**, *12* (22), 4385-91.
417. Strazza, S.; Hunter, R.; Walker, E.; Darnall, D. W., The thermodynamics of bovine and porcine insulin and proinsulin association determined by concentration difference spectroscopy. *Archives of Biochemistry and Biophysics* **1985**, *238* (1), 30-42.
418. Jensen, M. H.; Wahlund, P. O.; Toft, K. N.; Jacobsen, J. K.; Steensgaard, D. B.; van de Weert, M.; Havelund, S.; Vestergaard, B., Small angle X-ray scattering-based elucidation of the self-association mechanism of human insulin analogue lys(B29)(N(epsilon)omega-carboxyheptadecanoyl) des(B30). *Biochemistry* **2013**, *52* (2), 282-294.
419. Hjorth, C. F.; Norrman, M.; Wahlund, P. O.; Benie, A. J.; Petersen, B. O.; Jessen, C. M.; Pedersen, T. A.; Vestergaard, K.; Steensgaard, D. B.; Pedersen, J. S.; Naver, H.; Hubalek, F.; Poulsen, C.; Otzen, D., Structure, Aggregation, and Activity of a Covalent Insulin Dimer Formed During Storage of Neutral Formulation of Human Insulin. *Journal of pharmaceutical sciences* **2016**, *105* (4), 1376-86.
420. Kadima, W.; Ogendal, L.; Bauer, R.; Kaarsholm, N.; Brodersen, K.; Hansen, J. F.; Porting, P., The influence of ionic strength and pH on the aggregation properties of zinc-free insulin studied by static and dynamic laser light scattering. *Biopolymers* **1993**, *33* (11), 1643-1657.
421. Bohidar, H. B.; Geissler, E., Static and dynamic light scattering from dilute insulin solutions. *Biopolymers* **1984**, *23*, 2407-2417.
422. Roy, M.; Lee, R. W.; Brange, J.; Dunn, M. F., <sup>1</sup>H NMR spectrum of the native human insulin monomer. Evidence for conformational differences between the monomer and aggregated forms. *Journal of Biological Chemistry* **1990**, *265*, 5448-5452.
423. Ludvigsen, S.; Roy, M.; Thøgersen, H.; Kaarsholm, N. C., High-Resolution Structure of an Engineered Biologically Potent Insulin Monomer, B16 Tyr --> His, As Determined by Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **1994**, *33* (26), 7998-8006.
424. Kline, A. D.; Justice, R. M., Complete sequence-specific proton NMR assignments for human insulin. *Biochemistry* **1990**, *29* (12), 2906-2913.
425. Kadima, W.; Roy, M.; Lee, R. W. K.; Kaarsholm, N. C.; Dunn, M. F., Studies of the association and conformational properties of metal-free insulin in alkaline sodium chloride solutions by one- and two-dimensional <sup>1</sup>H NMR. *Journal of Biological Chemistry* **1992**, *267*, 8983-8970.
426. Nettleton, E. J.; Tito, P.; Sunde, M.; Bouchard, M.; Dobson, C. M.; Robinson, C. V., Characterization of the Oligomeric States of Insulin in Self-Assembly and Amyloid Fibril Formation by Mass Spectrometry. *Biophysical journal* **2000**, *79* (2), 1053-1065.
427. Fabris, D.; Fenselau, C., Characterization of Allosteric Insulin Hexamers by Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* **1999**, *71* (2), 384-387.
428. Chitta, R. K.; Rempel, D. L.; Grayson, M. A.; Remsen, E. E.; Gross, M. L., Application of SIMSTEX to oligomerization of insulin analogs and mutants. *Journal of the American Society for Mass Spectrometry* **2006**, *17* (11), 1526-1534.
429. Hansen, J. F., The self-association of zinc-free human insulin and insulin analogue B13-glutamine. *Biophysical chemistry* **1991**, *39* (1), 107-110.
430. Lin, M. F.; Larive, C. K., Detection of Insulin Aggregates with Pulsed-Field Gradient Nuclear Magnetic Resonance Spectroscopy. *Analytical Biochemistry* **1995**, *229* (2), 214-220.
431. Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F., Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharmaceutical Research* **2003**, *20* (9), 1325-1336.
432. Hiroyuki, H.; Tsutomu, A.; Kentaro, S., Effect of Additives on Protein Aggregation *Current Pharmaceutical Biotechnology* **2009**, *10* (4), 400-407.

433. Wang, W.; Nema, S.; Teagarden, D., Protein aggregation--pathways and influencing factors. *International journal of pharmaceuticals* **2010**, *390* (2), 89-99.
434. Wang, W., Protein aggregation and its inhibition in biopharmaceuticals. *International journal of pharmaceuticals* **2005**, *289* (1-2), 1-30.
435. Bhatnagar, B. S.; Bogner, R. H.; Pikal, M. J., Protein stability during freezing: separation of stresses and mechanisms of protein stabilization. *Pharmaceutical development and technology* **2007**, *12* (5), 505-523.
436. Dobson, C. M., Getting out of shape. *Nature* **2002**, *418* (6899), 729-730.
437. Philo, J. S., Is any measurement method optimal for all aggregate sizes and types? *The AAPS journal* **2006**, *8* (3), E564-E571.
438. Convention, T. U. S. P., USP 36-NF 31. In <788> *Particulate matter in injections*, Rockville, 2013.
439. HealthCare, E. D. f. t. Q. o. M., European pharmacopoeia 7th Edition. In *Ph.Eur. 2.9.19. General, particulate contamination: sub-visible particles.*, Strasbourg, 2011.
440. Beltzer, J., Measuring and Characterizing Protein Aggregates. *Genetic Engineering & Biotechnology News* **2012**, *32* (3), 28-29.
441. Philo, J., A Critical Review of Methods for Size Characterization of Non-Particulate Protein Aggregates. *Current Pharmaceutical Biotechnology* **2009**, *10* (4), 359-372.
442. Ejima, D.; Yumioka, R.; Arakawa, T.; Tsumoto, K., Arginine as an effective additive in gel permeation chromatography. *Journal of chromatography. A* **2005**, *1094* (1-2), 49-55.
443. Arakawa, T.; Ejima, D.; Li, T.; Philo, J. S., The critical role of mobile phase composition in size exclusion chromatography of protein pharmaceuticals. *Journal of pharmaceutical sciences* **2010**, *99* (4), 1674-1692.
444. Štulík, K.; Pacáková, V.; Tichá, M., Some potentialities and drawbacks of contemporary size-exclusion chromatography. *Journal of Biochemical and Biophysical Methods* **2003**, *56* (1-3), 1-13.
445. Carpenter, J. F.; Randolph, T. W.; Jiskoot, W.; Crommelin, D. J.; Middaugh, C. R.; Winter, G., Potential inaccurate quantitation and sizing of protein aggregates by size exclusion chromatography: essential need to use orthogonal methods to assure the quality of therapeutic protein products. *Journal of pharmaceutical sciences* **2010**, *99* (5), 2200-2208.
446. Convention, T. U. S. P., USP 36-NF 31. In *Official monographs for USP 36 / Insulin*, Rockville, 2013; pp 3911-3913.
447. HealthCare, E. D. f. t. Q. o. M., European pharmacopoeia 7th Edition. In *Insulin, human*, Strasbourg, 2011; pp 2241-2243.
448. Brange, J.; Langkjoer, L., Insulin structure and stability. *Pharmaceutical biotechnology* **1993**, *5*, 315-350.
449. Yumioka, R.; Sato, H.; Tomizawa, H.; Yamasaki, Y.; Ejima, D., Mobile phase containing arginine provides more reliable SEC condition for aggregation analysis. *Journal of pharmaceutical sciences* **2010**, *99* (2), 618-620.
450. Slew, A., Analyzing protein aggregation in biopharmaceuticals. *BioPharm International* **2015**, *28* (1), 40-43.
451. FDA, Guidance for industry: Immunogenicity assessment for therapeutic protein products. In *V. Patient and product-specific factors that affect immunogenicity*, 2014.
452. Bertoncini, C. W.; Celej, M. S., Small Molecule Fluorescent Probes for the Detection of Amyloid Self-Assembly In Vitro and In Vivo. *Current Protein & Peptide Science* **2011**, *12* (3), 206-220.
453. Lindgren, M.; Sorgjerd, K.; Hammarstrom, P., Detection and characterization of aggregates, prefibrillar amyloidogenic oligomers, and protofibrils using fluorescence spectroscopy. *Biophysical journal* **2005**, *88* (6), 4200-4212.

454. Bekard, I. B.; Dunstan, D. E., Tyrosine Autofluorescence as a Measure of Bovine Insulin Fibrillation. *Biophysical journal* **2009**, *97* (9), 2521-2531.
455. Frid, P.; Anisimov, S. V.; Popovic, N., Congo red and protein aggregation in neurodegenerative diseases. *Brain research reviews* **2007**, *53* (1), 135-160.
456. Chan, F. T. S.; Pinotsi, D.; Gabriele, S.; Schierle, K.; Kaminski, C. F., Chapter 13. Structure-Specific Intrinsic Fluorescence of Protein Amyloids Used to Study their Kinetics of Aggregation. In *Bio-nanoimaging. Protein Misfolding and Aggregation*, Uversky, V. N.; Lyubchenko, Y. L., Eds. Academic Press: 2014; pp 147-155.
457. Weber, G., Fluorescence-polarization spectrum and electronic-energy transfer in tyrosine, tryptophan and related compounds. *Biochemical Journal* **1960**, *75* (2), 335-345.
458. Searcy, D. G.; Montenay-Garestier, T.; Helene, C., Phenylalanine-to-tyrosine singlet energy transfer in the archaebacterial histone-like protein HTa. *Biochemistry* **1989**, *28* (23), 9058-9064.
459. Platt, J. R., Classification of Spectra of Cata-Condensed Hydrocarbons. *The Journal of Chemical Physics* **1949**, *17* (5), 484-495.
460. Martinez, S. J.; Alfano, J. C.; Levy, D. H., The electronic spectroscopy of the amino acids tyrosine and phenylalanine in a supersonic jet. *Journal of Molecular Spectroscopy* **1992**, *156* (2), 421-430.
461. Alexander Ross, J. B.; Laws, W. R.; Rousslang, K. W.; Wyssbrod, H. R., Tyrosine Fluorescence and Phosphorescence from Proteins and Polypeptides. In *Topics in Fluorescence Spectroscopy Volume 3: Biochemical Applications*, Lakowicz, J. R., Ed. Plenum Press: 1992; Vol. 3, pp 1-63.
462. Antosiewicz, J. M.; Shugar, D., UV-Vis spectroscopy of tyrosine side-groups in studies of protein structure. Part 1: basic principles and properties of tyrosine chromophore. *Biophysical Reviews* **2016**, *8* (2), 151-161.
463. Fornander, L. H.; Feng, B.; Beke-Somfai, T.; Norden, B., UV transition moments of tyrosine. *The journal of physical chemistry. B* **2014**, *118* (31), 9247-9257.
464. Lee, J.; Ross, R. T., Absorption and Fluorescence of Tyrosine Hydrogen-Bonded to Amide-like Ligands. *The Journal of Physical Chemistry B* **1998**, *102* (23), 4612-4618.
465. Rayner, D. M.; Krajcarski, D. T.; Szabo, A. G., Excited state acid-base equilibrium of tyrosine. *Canadian Journal of Chemistry* **1978**, *56* (9), 1238-1245.
466. Willis, K. J.; Szabo, A. G., Fluorescence decay kinetics of tyrosinate and tyrosine hydrogen-bonded complexes. *The Journal of Physical Chemistry* **1991**, *95* (4), 1585-1589.
467. Pundak, S.; Roche, R. S., Tyrosine and tyrosinate fluorescence of bovine testes calmodulin: calcium and pH dependence. *Biochemistry* **2002**, *23* (7), 1549-1555.
468. Szabo, A. G.; Lynn, K. R.; Krajcarski, D. T.; Rayner, D. M., Tyrosinate fluorescence maxima at 345 nm in proteins lacking tryptophan at pH 7. *FEBS letters* **1978**, *94* (2), 249-252.
469. Aeschbach, R.; Amado, R.; Neukom, H., Formation of dityrosine cross-links in proteins by oxidation of tyrosine residues. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1976**, *439* (2), 292-301.
470. Amado, R.; Aeschbach, R.; Neukom, H., Dityrosine: In vitro production and characterization. *Methods in Enzymology* **1984**, *107*, 377-388.
471. Malencik, D. A.; Anderson, S. R., Dityrosine as a product of oxidative stress and fluorescent probe. *Amino acids* **2003**, *25* (3-4), 233-247.
472. Giulivi, C.; Davies, K. J. A., Dityrosine: A marker for oxidatively modified proteins and selective proteolysis. *Methods in Enzymology* **1994**, *233*, 363-371.
473. Correia, M.; Neves-Petersen, M. T.; Jeppesen, P. B.; Gregersen, S.; Petersen, S. B., UV-light exposure of insulin: pharmaceutical implications upon covalent insulin dityrosine dimerization and disulphide bond photolysis. *PloS one* **2012**, *7* (12), e50733.

474. Ahmad, A.; Millett, I. S.; Doniach, S.; Uversky, V. N.; Fink, A. L., Stimulation of insulin fibrillation by urea-induced intermediates. *The Journal of Biological Chemistry* **2004**, *279* (15), 14999-5013.
475. Biancalana, M.; Koide, S., Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochimica et biophysica acta* **2010**, *1804* (7), 1405-1412.
476. Groenning, M., Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils-current status. *Journal of chemical biology* **2010**, *3* (1), 1-18.
477. Groenning, M.; Norrman, M.; Flink, J. M.; van de Weert, M.; Bukrinsky, J. T.; Schluckebier, G.; Frokjaer, S., Binding mode of Thioflavin T in insulin amyloid fibrils. *Journal of structural biology* **2007**, *159* (3), 483-97.
478. Manno, M.; Craparo, E. F.; Podesta, A.; Bulone, D.; Carrotta, R.; Martorana, V.; Tiana, G.; San Biagio, P. L., Kinetics of different processes in human insulin amyloid formation. *Journal of molecular biology* **2007**, *366* (1), 258-74.
479. Fodera, V.; Librizzi, F.; Groenning, M.; van de Weert, M.; Leone, M., Secondary nucleation and accessible surface in insulin amyloid fibril formation. *The journal of physical chemistry. B* **2008**, *112* (12), 3853-8.
480. Hassiepen, U.; Federwisch, M.; Mulders, T.; Lenz, V. J.; Gattner, H.-G.; Kruger, P.; Wollmer, A., Analysis of protein self-association at constant concentration by fluorescence-energy transfer. *European Journal of Biochemistry* **1998**, *255* (3), 580-587.
481. Karreman, G.; Steele, R. H.; Szent-Györgyi, A., On resonance transfer of excitation energy between aromatic aminoacids in proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1958**, *44* (2), 140-143.
482. Eisinger, J.; Feuer, B.; Lamola, A. A., Intramolecular singlet excitation transfer. Applications to polypeptides. *Biochemistry* **1969**, *8* (10), 3908-3915.
483. Demchenko, A. P., Chapter 9. Nonradiative transfer of electronic excitation energy. In *Ultraviolet Spectroscopy of Proteins*, Demchenko, A. P., Ed. Springer-Verlag: 1986.
484. Weber, G., Dependence of the polarization of the fluorescence on the concentration. *Transactions of the Faraday Society* **1954**, *50*, 552-555.
485. Tournon, J.; Kuntz, E.; El-Bayoumi, M. A., Fluorescence Quenching in Phenylalanine and Model Compounds. *Photochemistry and Photobiology* **1972**, *16* (5), 425-433.
486. Chen, R. F.; Cohen, P. F., Quenching of tyrosine fluorescence in proteins by phosphate. *Archives of Biochemistry and Biophysics* **1966**, *114* (3), 514-522.
487. Noronha, M.; Lima, J. C.; Bastos, M.; Santos, H.; Macanita, A. L., Unfolding of ubiquitin studied by picosecond time-resolved fluorescence of the tyrosine residue. *Biophysical journal* **2004**, *87* (4), 2609-2620.
488. 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.
489. Verdian-Doghaei, A.; Housaindokht, M. R., Spectroscopic study of the interaction of insulin and its aptamer – sensitive optical detection of insulin. *Journal of Luminescence* **2015**, *159*, 1-8.
490. Kliger, D. S.; Lewis, J. W.; Randall Einterz, C., Chapter 6. Spectroscopy with polarized light. In *Polarized light in optics and spectroscopy*, Kliger, D. S.; Lewis, J. W.; Randall Einterz, C., Eds. Academic Press, Inc.: 1990.
491. Feitelson, J., On the Mechanism of Fluorescence Quenching. Tyrosine and Similar Compounds. *The Journal of Physical Chemistry* **1964**, *68* (2), 391-397.
492. Cornog, J. L.; Adams, W. R., The fluorescence of tyrosine in alkaline solution. *Biochimica et biophysica acta* **1963**, *66*, 356-365.
493. Bro, R.; Harshman, R. A.; Sidiropoulos, N. D.; Lundy, M. E., Modeling multi-way data with linearly dependent loadings. *Journal Of Chemometrics* **2009**, *23* (7-8), 324-340.

494. Demchenko, A. P., Chapter 2. Display of intramolecular and intermolecular interactions in electronic spectra of amino acids and proteins. In *Ultraviolet Spectroscopy of Proteins*, Demchenko, A. P., Ed. Springer-Verlag: 1986.
495. Schmid, F. X., Biological Macromolecules: UV-visible spectrophotometry. In *Encyclopedia of Life Sciences*, Macmillan Publishers Ltd. : 2001.
496. Li, Y. K.; Kuliopulos, A.; Mildvan, A. S.; Talalay, P., Environments and mechanistic roles of the tyrosine residues of  $\Delta^5$ -3-ketosteroid isomerase. *Biochemistry* **1993**, *32* (7), 1816-1824.
497. Poveda, J. A.; Prieto, M.; Encinar, J. A.; Gonzalez-Ros, J. M.; Mateo, C. R., Intrinsic tyrosine fluorescence as a tool to study the interaction of the shaker B "ball" peptide with anionic membranes. *Biochemistry* **2003**, *42* (23), 7124-7132.
498. Smerdon, M. J.; Isenberg, I., Conformational changes in subfractions of calf thymus histone H1. *Biochemistry* **1976**, *15* (19), 4233-4242.
499. Wetlaufer, D. B., Ultraviolet spectra of proteins and amino acids. *Advances in Protein Chemistry* **1963**, *17*, 303-390.
500. Herskovits, T. T.; Sorensen, S. M., Studies of the location of tyrosyl and tryptophanyl residues in protein. II. Applications of model data to solvent perturbation studies of proteins rich in both tyrosine and tryptophan. *Biochemistry* **1968**, *7* (7), 2533-2542.
501. Li, L.; Li, C.; Zhang, Z.; Alexov, E., On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *Journal of chemical theory and computation* **2013**, *9* (4), 2126-2136.
502. Ryan, R. O.; Oikawa, K.; Kay, C. M., Conformational, thermodynamic, and stability properties of manduca sexta apolipophorin III. *The journal of Biological Chemistry* **1993**, *268* (3), 1525-1530.
503. Ichikawa, T.; Terada, H., Second derivative spectrophotometry as an effective tool for examining phenylalanine residues in proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1977**, *494* (1), 267-270.
504. Tay, T.; Agius, C.; Hamilton, R.; Bodle, J.; Rockman, S., Investigation into alternative testing methodologies for characterization of influenza virus vaccine. *Human vaccines & immunotherapeutics* **2015**, *11* (7), 1673-1684.
505. 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.
506. Harshman, R. A.; Lundy, M. E., The PARAFAC model for three-way factor analysis and multidimensional scaling, . In *Research methods for Multimode data analysis*, Law, H. G.; Snyder, C. W.; Hattie, J. A.; McDonald, R. P., Eds. Praeger: 1984.
507. Bahram, M.; Bro, R., A novel strategy for solving matrix effect in three-way data using parallel profiles with linear dependencies. *Analytica Chimica Acta* **2007**, *584*, 387-402.
508. Chen, H.; Zheng, B.; Song, Y., Comparison of PARAFAC and PARALIND in modeling three-way fluorescence data array with special linear dependences in three modes: a case study in 2-naphthol. *Journal of Chemometrics* **2011**, *25* (1), 20-27.
509. Ross, R. T.; Leurgans, S. E., Component resolution using multilinear models. *Methods in Enzymology* **1995**, *246*, 679-700.
510. Kiers, H. A. L.; Smilde, A. K., Constrained three-mode factor analysis as a tool for parameter estimation with second-order instrumental data. *Journal of Chemometrics* **1998**, *12*, 125-147.
511. Garrido, M.; Lázaro, I.; Larrechi, M. S.; Rius, F. X., Multivariate resolution of rank-deficient near-infrared spectroscopy data from the reaction of curing epoxy resins using the rank augmentation strategy and multivariate curve resolution alternating least squares approach. *Analytica Chimica Acta* **2004**, *515* (1), 65-73.

512. Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M. M., On the rank deficiency and rank augmentation of the spectral measurement matrix. *Chemometrics and Intelligent Laboratory Systems* **1996**, *33* (1), 17-33.
513. 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*, 180-189.
514. Nominé, Y.; Ristriani, T.; Laurent, C.; Lefèvre, J.-F.; Weiss, É.; Travé, G., A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. *Protein Engineering, Design and Selection* **2001**, *14* (4), 297-305.
515. Mori, Y., Chapter 2. Fundamentals properties of particles. In *Powder technology. Fundamentals of particles, powder beds, and particle generation.*, Masuda, H.; Higashitani, K.; Yoshida, H., Eds. CRC Press: 2006.
516. Brahma, A.; De, D.; Bhattacharyya, D., Rayleigh scattering technique as a method to study protein-protein interaction using spectrofluorimeters. *Current Science* **2009**, *96* (7), 940-946.
517. Gualandi-Signorini, A. M.; Giorgi, G., Insulin formulations - a review. *European Review for Medical and Pharmacological Sciences* **2001**, *5*, 73-83.
518. Harding, S. E., Chapter 9. Hydrodynamic properties of proteins. In *Protein structure - a practical approach*, Creighton, T. E., Ed. IRL Press: 1997.
519. Libertini, L. J.; Small, E. W., The intrinsic tyrosine fluorescence of histone H1. Steady state and fluorescence decay studies reveal heterogeneous emission. *Biophysical journal* **1985**, *47* (6), 765-772.
520. Runnels, L. W.; Scarlata, S. F., Theory and application of fluorescence homotransfer to melittin oligomerization. *Biophysical journal* **1995**, *69* (4), 1569-1583.
521. Yeow, E. K.; Clayton, A. H., Enumeration of oligomerization states of membrane proteins in living cells by homo-FRET spectroscopy and microscopy: theory and application. *Biophysical journal* **2007**, *92* (9), 3098-3104.
522. Longworth, J. W., Chapter 6. Luminescence of Polypeptides and Proteins. In *Excited States of Proteins and Nucleic Acids*, Steiner, R. F.; Weinryb, I., Eds. Plenum Press: 1971.
523. Danho, W.; Sasaki, A.; Büllsbach, E.; Föhles, J.; Gattner, H.-G., [A19-Phenylalanine] Insulin: A New Synthetic Analogue. *Hoppe-Seyler's Zeitschrift für physiologische Chemie* **1980**, *361* (1), 735-746.
524. Iwamoto, G. K.; Van Wagenen, R. A.; Andrade, J. D., Insulin adsorption: Intrinsic tyrosine interfacial fluorescence. *Journal of Colloid and Interface Science* **1982**, *86* (2), 581-585.
525. Giancotti, V.; Quadrifoglio, F.; Cowgill, R. W.; Crane-Robinson, C., Fluorescence of buried tyrosine residues in proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1980**, *624* (1), 60-65.
526. O'Neil, J. D.; Dorrington, K. J.; Kells, D. I. C.; Hofmann, T., Fluorescence and circular-dichroism properties of pig intestinal calciumbinding protein (Mr = 9000), a protein with a single tyrosine residue. *Biochemical Journal* **1982**, *207* (3), 389-396.
527. 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.
528. FDA, Guidance for industry PAT – A framework for innovative pharmaceutical manufacturing and quality assurance. Guidance for industry PAT – A framework for innovative pharmaceutical manufacturing and quality assurance.: 2004.
529. Rudt, M.; Briskot, T.; Hubbuch, J., Advances in downstream processing of biologics - Spectroscopy: An emerging process analytical technology. *Journal of chromatography. A* **2017**, *1490*, 2-9.

530. Stefani, M.; Dobson, C. M., Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med (Berl)* **2003**, *81* (11), 678-99.
531. Uversky, V. N.; Fink, A. L., Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochimica et biophysica acta* **2004**, *1698* (2), 131-53.
532. Cowgill, R. W., Fluorescence and the structure of proteins. III. Effects of denaturation on fluorescence of insulin and ribonuclease. *Archives of Biochemistry and Biophysics* **1964**, *104* (1), 84-92.
533. Jiracek, J.; Zakova, L.; Antolikova, E.; Watson, C. J.; Turkenburg, J. P.; Dodson, G. G.; Brzozowski, A. M., Implications for the active form of human insulin based on the structural convergence of highly active hormone analogues. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (5), 1966-70.
534. Xu, B.; Huang, K.; Chu, Y. C.; Hu, S. Q.; Nakagawa, S.; Wang, S.; Wang, R. Y.; Whittaker, J.; Katsoyannis, P. G.; Weiss, M. A., Decoding the cryptic active conformation of a protein by synthetic photoscanning: insulin inserts a detachable arm between receptor domains. *The Journal of biological chemistry* **2009**, *284* (21), 14597-608.
535. Pullen, R. A.; Lindsay, D. G.; Wood, S. P.; Tickle, I. J.; Blundell, T. L.; Wollmer, A.; Krail, G.; Brandenburg, D.; Zahn, H.; Gliemann, J.; Gammeltoft, S., Receptor-binding region of insulin. *Nature* **1976**, *259* (5542), 369-373.
536. Banting, F. G.; Best, C. H., Pancreatic extracts. *The journal of laboratory and clinical medicine* **1922**, *7*, 464-472.
537. Banting, F. G.; Best, C. H., The internal secretion of the pancreas. *The journal of laboratory and clinical medicine* **1922**, *7*, 256-271.
538. Rosenfield, L., Insulin: Discovery and Controversy. *Clinical Chemistry* **2002**, *48* (12), 2270-2288.
539. Leung, P. S.; Ip, S. P., Pancreatic acinar cell: its role in acute pancreatitis. *The international journal of biochemistry & cell biology* **2006**, *38* (7), 1024-30.
540. Barron, M., The relation of the islets of Langerhans to diabetes with special reference to cases of pancreatic lithiasis. *Surgery, Gynecology & Obstetrics* **1920**, *31*, 437-448.
541. Best, C. H., The First Clinical Use of Insulin. *Diabetes* **1956**, *5* (1), 65-67.
542. Hagedorn, H. C., Protamine Insulin. *Journal of the American Medical Association* **1936**, *106* (3), 177-180.
543. Felig, P., Protamine Insulin. *Journal of the American Medical Association* **1984**, *251* (3), 393-396.
544. Oakley, W.; Hill, D.; Oakley, N., Combined Use of Regular and Crystalline Protamine (NPH) Insulins in the Treatment of Severe Diabetes. *Diabetes* **1966**, *15* (3), 219-222.
545. Hallas-Mo, K., The Lente Insulins. *Diabetes* **1956**, *5* (1), 7-14.
546. Brown, L., Part 9. Diabetes and the heart. In *Molecular Pathomechanisms and New Trends in Drug Research*, Keri, G.; Toth, I., Eds. Taylor & Francis: 2003.
547. Owens, D. R., Insulin preparations with prolonged effect. *Diabetes technology & therapeutics* **2011**, *13*, S5-14.
548. Morihara, K.; Oka, T.; Tsuzuki, H., Semi-synthesis of human insulin by trypsin-catalysed replacement of Ala-B30 by Thr in porcine insulin. *Nature* **1979**, *280* (5721), 412-413.
549. Inouye, K.; Watanabe, K.; Morihara, K.; Tochino, Y.; Kanaya, T.; Emura, J.; Sakakibara, S., Enzyme-assisted semisynthesis of human insulin. *Journal of the American Chemical Society* **1979**, *101* (3), 751-752.
550. Markussen, J.; Damgaard, U.; Jørgensen, K. H.; Sørensen, E.; Thim, L., Human Monocomponent Insulin. *Acta Medica Scandinavica* **2009**, *213* (S671), 99-105.

551. Morihara, K.; Ueno, Y., A new procedure for enzymatic semisynthesis of human insulin by hydrolysis of single-chain des-(b-30)-Insulin precursor with lysyl endopeptidase. *Biotechnology and bioengineering* **1991**, *37* (7), 693-5.
552. Morihara, K., Enzymatic semisynthesis of human insulin: an update. *Journal of molecular recognition : JMR* **1990**, *3* (5-6), 181-6.
553. Griffiths, A. J. F.; Miller, J. H.; Suzuki, D. T.; Lewontin, R. C.; Gelbart, W. M., Chapter 12. Recombinant DNA Technology. In *An Introduction to Genetic Analysis, 7th edition*, Freeman, W. H., Ed. 2000.
554. Kamionka, M., Engineering of Therapeutic Proteins Production in Escherichia coli. *Current Pharmaceutical Biotechnology* **2011**, *12* (2), 268-274.
555. Johnson, I., Human insulin from recombinant DNA technology. *Science* **1983**, *219* (4585), 632-637.
556. Baeshen, N. A.; Baeshen, M. N.; Sheikh, A.; Bora, R. S.; Ahmed, M. M.; Ramadan, H. A.; Saini, K. S.; Redwan, E. M., Cell factories for insulin production. *Microbial cell factories* **2014**, *13*, 141.
557. Thim, L.; Hansen, M. T.; Norris, K.; Hoegh, I.; Boel, E.; Forstrom, J.; Ammerer, G.; Fiil, N. P., Secretion and processing of insulin precursors in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83* (18), 6766-70.
558. Thim, L.; Hansen, M. T.; Sorensen, A. R., Secretion of human insulin by a transformed yeast cell. *FEBS letters* **1987**, *212* (2), 307-12.
559. Lalonde, M. E.; Durocher, Y., Therapeutic glycoprotein production in mammalian cells. *Journal of biotechnology* **2017**, *251*, 128-140.
560. Bandaranayake, A. D.; Almo, S. C., Recent advances in mammalian protein production. *FEBS letters* **2014**, *588* (2), 253-60.
561. Dingermann, T., Recombinant therapeutic proteins: production platforms and challenges. *Biotechnology journal* **2008**, *3* (1), 90-7.
562. Boyhan, D.; Daniell, H., Low-cost production of proinsulin in tobacco and lettuce chloroplasts for injectable or oral delivery of functional insulin and C-peptide. *Plant biotechnology journal* **2011**, *9* (5), 585-98.
563. Nykiforuk, C. L.; Boothe, J. G.; Murray, E. W.; Keon, R. G.; Goren, H. J.; Markley, N. A.; Moloney, M. M., Transgenic expression and recovery of biologically active recombinant human insulin from Arabidopsis thaliana seeds. *Plant biotechnology journal* **2006**, *4* (1), 77-85.
564. Bill, R. M., Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. *Frontiers in microbiology* **2014**, *5*, 85.
565. Kuroda, A.; Kaneto, H.; Yasuda, T.; Matsuhisa, M.; Miyashita, K.; Fujiki, N.; Fujisawa, K.; Yamamoto, T.; Takahara, M.; Sakamoto, F.; Matsuoka, T. A.; Shimomura, I., Basal insulin requirement is ~30% of the total daily insulin dose in type 1 diabetic patients who use the insulin pump. *Diabetes care* **2011**, *34* (5), 1089-90.
566. Kalra, S., Basal insulin analogues in the treatment of diabetes mellitus: What progress have we made? *Indian journal of endocrinology and metabolism* **2015**, *19* (Suppl 1), S71-3.
567. Rossetti, P.; Ampudia-Blasco, F. J.; Ascaso, J. F., Old and new basal insulin formulations: understanding pharmacodynamics is still relevant in clinical practice. *Diabetes, obesity & metabolism* **2014**, *16* (8), 695-706.
568. Bolli, G. B.; Di Marchi, R. D.; Park, G. D.; Pramming, S.; Koivisto, V. A., Insulin analogues and their potential in the management of diabetes mellitus. *Diabetologia* **1999**, *42* (10), 1151-67.
569. Berenson, D. F.; Weiss, A. R.; Wan, Z. L.; Weiss, M. A., Insulin analogs for the treatment of diabetes mellitus: therapeutic applications of protein engineering. *Annals of the New York Academy of Sciences* **2011**, *1243*, E40-E54.

570. Home, P. D.; Ashwell, S. G., An overview of insulin glargine. *Diabetes/metabolism research and reviews* **2002**, *18 Suppl 3*, S57-63.
571. Soran, H.; Younis, N., Insulin detemir: a new basal insulin analogue. *Diabetes, obesity & metabolism* **2006**, *8* (1), 26-30.
572. Tahrani, A. A.; Bailey, C. J.; Barnett, A. H., Insulin degludec: a new ultra-longacting insulin. *The Lancet* **2012**, *379* (9825), 1465-1467.
573. Homko, C.; Deluzio, A.; Jimenez, C.; Kolaczynski, J. W.; Boden, G., Comparison of Insulin Aspart and Lispro: Pharmacokinetic and metabolic effects. *Diabetes care* **2003**, *26* (7), 2027-2031.
574. Plank, J.; Wutte, A.; Brunner, G.; Siebenhofer, A.; Semlitsch, B.; Sommer, R.; Hirschberger, S.; Pieber, T. R., A Direct Comparison of Insulin Aspart and Insulin Lispro in Patients With Type 1 Diabetes. *Diabetes care* **2002**, *25* (11), 2053-2057.
575. Caparrotta, T. M.; Evans, M., PEGylated insulin Lispro, (LY2605541)--a new basal insulin analogue. *Diabetes, obesity & metabolism* **2014**, *16* (5), 388-95.
576. Mbanya, J. C.; Sandow, J.; Landgraf, W.; Owens, D. R., Recombinant Human Insulin in Global Diabetes Management - Focus on Clinical Efficacy. *European endocrinology* **2017**, *13* (1), 21-25.
577. American Diabetes, A., Economic costs of diabetes in the U.S. in 2012. *Diabetes care* **2013**, *36* (4), 1033-46.
578. American Diabetes, A., Economic Costs of Diabetes in the U.S. in 2017. *Diabetes care* **2018**, *41* (5), 917-928.
579. Boyle, J. P.; Thompson, T. J.; Gregg, E. W.; Barker, L. E.; Williamson, D. F., Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Population health metrics* **2010**, *8*, 29.
580. Heinemann, L., Biosimilar Insulin and Costs: What Can We Expect? *Journal of diabetes science and technology* **2015**, *10* (2), 457-62.
581. Nielsen, J., Production of biopharmaceutical proteins by yeast: advances through metabolic engineering. *Bioengineered* **2013**, *4* (4), 207-11.
582. Abel, J. J., Crystalline Insulin. *Proceedings of the National Academy of Sciences* **1926**, *12* (2), 132-136.
583. Vigneaud, V.; Jensen, H.; Wintersteiner, O., Studies on crystalline insulin III. Further observations on the crystallization of insulin and on the nature of the sulfur linkage. The isolation of cystine and tyrosine from hydrolyzed crystalline insulin. *Journal of Pharmacology and Experimental Therapeutics* **1928**, *32* (5), 367-385.
584. Sjogren, B.; Svedberg, T., The Molecular Weight of Insulin. *Journal of the American Chemical Society* **1931**, *53* (7), 2657-2661.
585. Gutfreund, H., The osmotic pressure of insulin solutions. *The Biochemical journal* **1948**, *42* (1), 156-60.
586. Harfenist, E. J.; Craig, L. C., The Molecular Weight of Insulin. *Journal of the American Chemical Society* **1952**, *74* (12), 3087-3089.
587. Stretton, A. O., The first sequence. Fred Sanger and insulin. *Genetics* **2002**, *162* (2), 527-32.
588. Sanger, F., The free amino groups of insulin. *Biochemical Journal* **1945**, *39* (5), 507-515.
589. Sanger, F., The terminal peptides of insulin. *Biochemical Journal* **1949**, *45*, 563-574.
590. Sanger, F., Fractionation of oxidized insulin. *Biochemical Journal* **1948**, *44*, 126-128.
591. Sanger, F.; Tuppy, H., The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochemical Journal* **1951**, *49*, 463-481.

592. Sanger, F.; Tuppy, H., The amino-acid sequence in the phenylalanyl chain of insulin. II. The identification of peptides from enzymic hydrolysates. *Biochemical Journal* **1951**, *49*, 481-490.
593. Sanger, F.; Thompson, O. P., The amino-acid sequence in the glycyl chain of insulin. I. The investigation of lower peptides from partial hydrolysates. *Biochemical Journal* **1953**, *53*, 353-366.
594. Sanger, F.; Thompson, O. P., The amino-acid sequence in the glycyl chain of insulin. II. The investigation of peptides from enzymic hydrolysates. *Biochemical Journal* **1953**, *53*, 366-374.
595. Ryle, A. P.; Sanger, F.; Smith, L. F.; Kitai, R., The disulphide bonds of insulin. *Biochemical Journal* **1954**, *60*, 541-556.
596. Howard, J. A., Dorothy Hodgkin and her contributions to biochemistry. *Nature reviews. Molecular cell biology* **2003**, *4* (11), 891-6.
597. Crowfoot, D., X-Ray Single Crystal Photographs of Insulin. *Nature* **1935**, *135* (3415), 591-592.
598. Scott, D. A., Crystalline insulin. *Biochemical Journal* **1934**, *28* (4), 1592-1602.
599. Low, B. W.; Einstein, J. R., Symmetry of Insulin Dimers and Hexamers. *Nature* **1960**, *186* (4723), 470-470.
600. Harding, M. M.; Hodgkin, D. C.; Kennedy, A. F.; O'Connor, A.; Weitzmann, P. D. J., The crystal structure of insulin. *Journal of Molecular Biology* **1966**, *16* (1), 212-IN30.
601. Blundell, T. L.; Cutfield, J. F.; Cutfield, S. M.; Dodson, E. J.; Dodson, G. G.; Hodgkin, D. C.; Mercola, D. A.; Vijayan, M., Atomic Positions in Rhombohedral 2-Zinc Insulin Crystals. *Nature* **1971**, *231* (5304), 506-511.
602. Group, B. I. S. R., Studies on crystal structure of insulin at 2.5 Å. *Scientia Sinica* **1972**, 3-30.
603. Zhang, Y., Insulin research in China and the U.K. *Biochemical Society transactions* **2011**, *39* (5), 1323-6.
604. Wang, J.-H., The insulin connection: Dorothy Hodgkin and the Beijing Insulin Group. *Trends in Biochemical Sciences* **1998**, *23* (12), 497-500.
605. Crowfoot Hodgkin, D., Chinese work on insulin. *Nature* **1975**, *255* (5504), 103-103.
606. Group, P. I. S. R., Studies on the insulin crystal structure: the molecule at 1.8 Å resolution. *Scientia Sinica* **1974**, *17*, 752-757.
607. Bi, R. U. C.; Dauter, Z.; Dodson, E.; Dodson, G.; Gordiano, F.; Hubbard, R.; Reynolds, C., Structural changes in the monomeric despentapeptide (B30-B26) insulin crystal. *Proceedings of the Indian Academy of Sciences (Chemical Sciences)* **1983**, *92* (4-5), 473-483.
608. Gattner, H. G., B-chain shortening of matrix-bound insulin by pepsin, I : Preparation and properties of bovine des-pentapeptide(B26-30) insulin. *Hoppe-Seyler's Zeitschrift Fur Physiologische Chemie* **1975**, *356*, 1397-1404.
609. Yun-yu, S.; Ru-huai, Y.; van Gunsteren, W. F., Molecular dynamics simulation of despentapeptide insulin in a crystalline environment. *Journal of Molecular Biology* **1988**, *200* (3), 571-577.
610. Bradbury, J. H.; Ramesh, V.; Dodson, G., <sup>1</sup>H nuclear magnetic resonance study of the histidine residues of insulin. *Journal of Molecular Biology* **1981**, *150* (4), 609-613.
611. Weiss, M. A.; Nguyen, D. T.; Khait, I.; Inouye, K.; Frank, B. H.; Beckage, M.; O'Shea, E.; Shoelson, S. E.; Karplus, M.; Neuringer, L. J., Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: assignment of aromatic resonances with application to protein folding, structure, and dynamics. *Biochemistry* **1989**, *28* (25), 9855-73.
612. Kowalsky, A., Nuclear Magnetic Resonance Studies of Proteins. *The journal of Biological Chemistry* **1962**, *237* (6), 1807-1819.

613. Williamson, K. L.; Williams, R. J. P., Conformational analysis by nuclear magnetic resonance: insulin. *Biochemistry* **1979**, *18* (26), 5966-5972.
614. Cheshovsky, D.; Neuringer, L. J.; Williamson, K. L., 500-MHz<sup>1</sup>H NMR studies of insulin: Complete assignment of histidine resonances. *Journal of Protein Chemistry* **1983**, *2* (4), 335-339.
615. Boelens, R.; Ganadu, M. L.; Verheyden, P.; Kaptein, R., Two-dimensional NMR studies on des-pentapeptide-insulin. Proton resonance assignments and secondary structure analysis. *European Journal of Biochemistry* **1990**, *191* (1), 147-153.
616. Weiss, M. A.; Frank, B. H.; Khait, I.; Pekar, A.; Heiney, R.; Shoelson, S. E.; Neuringer, L. J., NMR and photo-CIDNP studies of human proinsulin and prohormone processing intermediates with application to endopeptidase recognition. *Biochemistry* **1990**, *29* (36), 8389-8401.
617. Hua, Q. X.; Weiss, M. A., Comparative 2D NMR studies of human insulin and des-pentapeptide insulin: sequential resonance assignment and implications for protein dynamics and receptor recognition. *Biochemistry* **1991**, *30* (22), 5505-15.
618. Hua, Q. X.; Weiss, M. A., Toward the solution structure of human insulin: sequential 2D proton NMR assignment of a des-pentapeptide analog and comparison with crystal structure. *Biochemistry* **1990**, *29* (46), 10545-10555.
619. Mirmira, R. G.; Tager, H. S., Disposition of the phenylalanine B25 side chain during insulin-receptor and insulin-insulin interactions. *Biochemistry* **1991**, *30* (33), 8222-8229.
620. Nakagawa, S. H.; Tager, H. S., Role of the phenylalanine B25 side chain in directing insulin interaction with its receptor. Steric and conformational effects. *The journal of Biological Chemistry* **1986**, *261* (16), 7332-7341.
621. Derewenda, U.; Derewenda, Z.; Dodson, E. J.; Dodson, G. G.; Bing, X.; Markussen, J., X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. *Journal of Molecular Biology* **1991**, *220* (2), 425-433.
622. Chang, X.; Jorgensen, A. M.; Bardrum, P.; Led, J. J., Solution structures of the R6 human insulin hexamer. *Biochemistry* **1997**, *36* (31), 9409-22.
623. Jørgensen, A. M. M.; Kristensen, S. M.; Led, J. J.; Balschmidt, P., Three-dimensional solution structure of an insulin dimer. *Journal of Molecular Biology* **1992**, *227* (4), 1146-1163.
624. Mark, A. E.; Berendsen, H. J. C.; Van Gunsteren, W. F., Conformational flexibility of aqueous monomeric and dimeric insulin: a molecular dynamics study. *Biochemistry* **1991**, *30* (45), 10866-10872.
625. Weiss, M. A.; Hua, Q. X.; Lynch, C. S.; Frank, B. H.; Shoelson, S. E., Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective deuterium and carbon-13 labeling with application to protein design. *Biochemistry* **1991**, *30* (30), 7373-7389.
626. Rahuel-Clermont, S.; French, C. A.; Chou, C. I.; Kaarsholm, N. C.; Dunn, M. F., Mechanisms of Stabilization of the Insulin Hexamer through Allosteric Ligand Interactions. *Biochemistry* **1997**, *36* (32), 9950-9950.
627. Berchtold, H.; Hilgenfeld, R., Binding of phenol to R6 insulin hexamers. *Biopolymers* **1999**, *51* (2), 165-172.
628. Derewenda, U.; Derewenda, Z.; Dodson, E. J.; Dodson, G. G.; Reynolds, C. D.; Smith, G. D.; Sparks, C.; Swenson, D., Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. *Nature* **1989**, *338* (6216), 594-596.
629. Smith, G. D.; Dodson, G. G., Structure of a rhombohedral R6 insulin/phenol complex. *Proteins* **1992**, *14* (3), 401-8.
630. Bloom, C. R.; Choi, W. E.; Brzovic, P. S.; Ha Sheng-Tung Huang, J. J.; Kaarsholm, N. C.; Dunn, M. F., Ligand Binding to Wild-type and E-B13Q Mutant Insulins: A Three-state

- Allosteric Model System Showing Half-site Reactivity. *Journal of Molecular Biology* **1995**, *245* (4), 324-330.
631. Brzovic, P. S.; Choi, W. E.; Borchardt, D.; Kaarsholm, N. C.; Dunn, M. F., Structural Asymmetry and Half-Site Reactivity in the T to R Allosteric Transition of the Insulin Hexamer. *Biochemistry* **1994**, *33* (44), 13057-13069.
632. Brader, M. L.; Kaarsholm, N. C.; Lee, R. W. K.; Dunn, M. F., Characterization of the r-state insulin hexamer and its derivatives. The hexamer is stabilized by heterotropic ligand binding interactions. *Biochemistry* **1991**, *30* (27), 6636-6645.
633. Choi, W. E.; Brader, M. L.; Aguilar, V.; Kaarsholm, N. C.; Dunn, M. F., Allosteric transition of the insulin hexamer is modulated by homotropic and heterotropic interactions. *Biochemistry* **1993**, *32* (43), 11638-11645.
634. Roy, M.; Brader, M. L.; Lee, R. W.; Kaarsholm, N. C.; Hansen, J. F.; Dunn, M. F., Spectroscopic signatures of the T to R conformational transition in the insulin hexamer. *The journal of Biological Chemistry* **1989**, *264*, 19081-19085.
635. Whittingham, J. L.; Chaudhuri, S.; Dodson, E. J.; Moody, P. C. E.; Dodson, G. G., X-ray Crystallographic Studies on Hexameric Insulins in the Presence of Helix-Stabilizing Agents, Thiocyanate, Methylparaben and Phenol. *Biochemistry* **1995**, *34* (47), 15553-15563.
636. Vinther, T. N.; Norrman, M.; Ribel, U.; Huus, K.; Schlein, M.; Steensgaard, D. B.; Pedersen, T. A.; Pettersson, I.; Ludvigsen, S.; Kjeldsen, T.; Jensen, K. J.; Hubalek, F., Insulin analog with additional disulfide bond has increased stability and preserved activity. *Protein science : a publication of the Protein Society* **2013**, *22* (3), 296-305.
637. Nakagawa, S. H.; Zhao, M.; Hua, Q. X.; Hu, S. Q.; Wan, Z. L.; Jia, W.; Weiss, M. A., Chiral mutagenesis of insulin. Foldability and function are inversely regulated by a stereospecific switch in the B chain. *Biochemistry* **2005**, *44* (13), 4984-4999.
638. Brange, J.; Langkjoer, L., Chemical stability of insulin. 3. Influence of excipients, formulation, and pH. *Acta pharmaceutica Nordica* **1992**, *4* (3), 149-158.
639. Maltesen, M. J.; Bjerregaard, S.; Hovgaard, L.; Havelund, S.; van de Weert, M., Analysis of insulin allostery in solution and solid state with FTIR. *Journal of pharmaceutical sciences* **2009**, *98* (9), 3265-77.
640. Eftink, M. R.; Selvidge, L. A.; Callis, P. R.; Rehms, A. A., Photophysics of indole derivatives: experimental resolution of La and Lb transitions and comparison with theory. *The Journal of Physical Chemistry* **1990**, *94* (9), 3469-3479.
641. Valeur, B.; Weber, G., Resolution of the fluorescence excitation spectrum of indole into the 1La and 1Lb excitation bands. *Photochemistry and Photobiology* **1977**, *25* (5), 441-444.
642. Gryczynski, I.; Wicz, W.; Johnson, M. L.; Lakowicz, J. R., Lifetime distributions and anisotropy decays of indole fluorescence in cyclohexane/ethanol mixtures by frequency-domain fluorometry. *Biophysical chemistry* **1988**, *32* (2-3), 173-185.
643. Burstein, E. A.; Permyakov, E. A.; Yashin, V. A.; Burkhanov, S. A.; Finazzi Agro, A., The fine structure of luminescence spectra of azurin. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1977**, *491* (1), 155-159.
644. Pierce, D. W.; Boxer, S. G., Stark effect spectroscopy of tryptophan. *Biophysical journal* **1995**, *68* (4), 1583-1591.
645. Vivian, J. T.; Callis, P. R., Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophysical journal* **2001**, *80* (5), 2093-2109.
646. 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.
647. Callis, P. R.; Vivian, J. T., Understanding the variable fluorescence quantum yield of tryptophan in proteins using QM-MM simulations. Quenching by charge transfer to the peptide backbone. *Chemical Physics Letters* **2003**, *369* (3-4), 409-414.

648. Hansen, S. B.; Radic, Z.; Talley, T. T.; Molles, B. E.; Deerinck, T.; Tsigelny, I.; Taylor, P., Tryptophan fluorescence reveals conformational changes in the acetylcholine binding protein. *The Journal of biological chemistry* **2002**, *277* (44), 41299-41302.
649. Royer, C. A., Probing protein folding and conformational transitions with fluorescence. *Chemical reviews* **2006**, *106* (5), 1769-1784.
650. Zhang, L.; Kao, Y. T.; Qiu, W.; Wang, L.; Zhong, D., Femtosecond studies of tryptophan fluorescence dynamics in proteins: local solvation and electronic quenching. *The journal of physical chemistry. B* **2006**, *110* (37), 18097-18103.
651. Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M., Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420* (6911), 102-106.
652. Rawel, H. M.; Frey, S. K.; Meidtner, K.; Kroll, J.; Schweigert, F. J., Determining the binding affinities of phenolic compounds to proteins by quenching of the intrinsic tryptophan fluorescence. *Molecular nutrition & food research* **2006**, *50* (8), 705-713.
653. Sytnik, A.; Litvinyuk, I., Energy transfer to a proton-transfer fluorescence probe: Tryptophan to a flavonol in human serum albumin. *Proceedings of the National Academy of Sciences of the United States of America* **1996**, *93* (23), 12959-12963.
654. Ladokhin, A. S.; Jayasinghe, S.; White, S. H., How to measure and analyze tryptophan fluorescence in membranes properly, and why bother? *Analytical biochemistry* **2000**, *285* (2), 235-245.
655. Rajih, M.; Comon, P.; Harshman, R. A., Enhanced Line Search: A Novel Method to Accelerate PARAFAC. *SIAM Journal on Matrix Analysis and Applications* **2008**, *30* (3), 1128-1147.
656. Tomasi, G.; Bro, R., A comparison of algorithms for fitting the PARAFAC model. *Computational Statistics & Data Analysis* **2006**, *50* (7), 1700-1734.
657. Bro, R.; Kiers, H. A. L., A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics* **2003**, *17*, 274-286.
658. Leach, S. J.; Scheraga, H. A., Effect of Light Scattering on Ultraviolet Difference Spectra. *Journal of the American Chemical Society* **1960**, *82* (18), 4790-4792.

## List of scientific achievements related to this work

### Peer-reviewed publications

- Casamayou-Boucau, Y.; Ryder, A. G., Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR). *Analytica Chimica Acta* **2018**, 1000, 132-143.
- 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, 037001.

### Associated publications

- Steiner-Browne, M.; Elcoroaristizabal, S.; Casamayou-Boucau, Y., 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.

### International presentations

- Oral presentation: “Developing a better analytical methodology (ARMES) for accurately detecting and discriminating soluble insulin aggregates”, Y. Casamayou-Boucau\* and A.G. Ryder, *International Forum Process Analytical Chemistry*, Maryland, USA, 11-14 Feb., **2018**.
- Oral presentation: “Accurate anisotropy mapping of Perylene obtained after Rayleigh scattering correction”, Y. Casamayou-Boucau\* and A.G. Ryder, *XVI Chemometrics in Analytical Chemistry*, Barcelona, Spain, 6–10 Jun., **2016**.
- Poster: “Multidimensional Anisotropy Standards for ARMES”, Y. Casamayou-Boucau\* and A.G. Ryder, *15<sup>th</sup> Conference on Methods and Applications in Fluorescence*, Bruges, Belgium, 10-13 Sept., **2017**.

### National presentations

- Oral presentation (**1st prize**): “Developing a better analytical methodology (ARMES) for accurately measuring soluble insulin aggregates”, Y. Casamayou-Boucau\* and A.G. Ryder, *BOC Postgraduate prize 2018*, Galway, Ireland, 11 Apr., **2018**.

- Oral presentation (**Second prize**): “A new rapid analytical methodology to discriminate non-covalent soluble insulin aggregates”, Y. Casamayou-Boucau\* and A.G. Ryder, *Eli Lilly Postgraduate prize 2018*, Galway, **Ireland**, 6 Mar., 2018.
- Oral presentation (thesis in 3 minutes): “A new rapid analytical methodology to discriminate non-covalent soluble insulin aggregates”, Y. Casamayou-Boucau\* and A.G. Ryder, *2<sup>nd</sup> Edition of Chemistry Career’s Day*, Galway, **Ireland**, 7 Mar., **2018**.
- Poster: “A new analytical method to accurately discriminate soluble insulin aggregates.” Y. Casamayou-Boucau\* and A.G. Ryder, *Biopharma Ambition*, Dublin, **Ireland**, 21-22 Feb., **2018**.
- Poster, “Perylene as a multidimensional standard for ARMES”, Y. Casamayou-Boucau\* and A.G. Ryder, *7<sup>th</sup> Annual NUIG-UL Alliance Research Day*, Galway, **Ireland**, 19 Apr., **2017**.

#### **Additional research contributions**

All posters at the *15<sup>th</sup> Conference on Methods and Applications of Fluorescence: Spectroscopy, Imaging and Probes*, Bruges, **Belgium**, 10–13 Sept., **2017**.

- An Excitation Emission Fluorescence Lifetime Matrix (EEFLM) spectrometer based on a frequency doubled supercontinuum laser source. D. Melnikau\*, Y. Casamayou-Boucau, S. Elcoroaristizabal, and A.G. Ryder.
- 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, & AG Ryder.
- Characterizing IgG conjugation products using Anisotropy Resolved Multi-Dimensional Spectroscopy (ARMES). A.L. de Faria\*, S. Elcoroaristizabal, Y. Casamayou-Boucau, & AG Ryder.

All posters at the *Biopharma Ambition*, Dublin, **Ireland**, 21-22 Feb., **2018**.

- (**Best poster price**) Assessing Freeze-Thaw (FT) Effects in IgG using Anisotropy Resolved Multi-Dimensional Emission Spectroscopy (ARMES). Marina Steiner-Browne\*, Saioa Elcoroaristizabal, Yannick Casamayou-Boucau, and Alan G. Ryder.
- Characterizing PEGylation reactions using multidimensional fluorescence (MDF) spectroscopy. A.L. de Faria\*, S. Elcoroaristizabal, Y. Casamayou-Boucau, & AG Ryder