Analytica Chimica Acta DOI: 10.1016/j.aca.2020.09.007 .

Quantitative analysis of weakly bound insulin oligomers in solution using polarized multidimensional fluorescence spectroscopy.

Yannick Casamayou-Boucau and Alan G. Ryder.*
Nanoscale BioPhotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Galway, H91 CF50, Ireland.

* Corresponding author: **Email:** alan.ryder@nuigalway.ie, **Phone:** +353-91-492943.
Postal address: Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland, Galway, University Road, Galway, Ireland.

Citation: Quantitative analysis of weakly-bound insulin oligomers in solution using polarized multidimensional fluorescence spectroscopy. Y. Casamayou-Boucau and A.G. Ryder. *Analytica Chimica Acta*, 1138, 18-29, (**2020**). **DOI**: 10.1016/j.aca.2020.09.007

Running Title: Decomposition of multidimensional fluorescence measurements to obtain insulin size and oligomer composition.

Abstract: Being able to measure the size and distribution of oligomers in solution is a critical issue in the manufacture and stability of insulin and other protein formulations. Measuring oligomers can however be complicated, as these are fragile self-assembled structures held together by weak forces. This can cause issues in chromatographic based methods, where dissociation or re-equilibration of oligomer populations can occur *e.g.* upon dilution in a different eluting buffer, but also for light scattering based methods like dynamic light scattering (DLS) where the size range involved does not always generate effective scattering signals, or allow for mixtures of oligomers to be resolved. Intrinsic fluorescence offers an attractive alternative as it is non-invasive, sensitive but also because it contains scattered light when implemented via excitation emission matrix (EEM) measurements, that is sensitive to changes in particle size.

Here, using insulin at formulation level concentrations, we show for the first time how EEM can both discriminate and quantify the proportion of oligomeric states in solution. This was achieved by using the Rayleigh scatter (RS) band and the fluorescence signal contained in EEM. After validating size changes with DLS, we show in particular how the volume under the RS band correlated linearly with protein/oligomer molecular weight. This was true for the RS data from both EEM and polarized EEM (pEEM) measurements, the latter providing a more informative, stronger scatter signal, sensitive to particle size changes.

The fluorescence signal was then used with multivariate curve resolution (MCR) to quantify more precisely the soluble oligomer composition of insulin solutions. In conditions that promoted the formation of mainly one type of oligomer (monomer, dimer, or hexamer), pEEM-MCR helped identify the presence of small amounts of other oligomeric forms, while in conditions that were previously said to favour the insulin tetramer, we show that in the presence of zinc, these insulin samples were instead a heterogenous mixture composed of mostly dimers and hexamers. These MCR results correlated in all cases with the observed discrimination by principal component analysis (PCA), and deviations observed in the RS data. In conclusion, using pEEM scatter and emission components with chemometric data analysis provides a unique analytical method for characterising and monitoring changes in the soluble oligomeric state of proteins.

Keywords: Protein, Oligomers, Multidimensional Fluorescence, Rayleigh scatter, Polarization, Chemometrics.

1. Introduction

Protein aggregation remains one of the biggest challenges affecting the manufacture and safe use of biopharmaceuticals, altering bioavailability [1] and potentially increasing the risk of adverse immune responses [2]. While its underlying mechanisms are not yet fully understood, aggregation can occur at every bioprocess step, and can be triggered by changes in multiple factors (temperature, pH, salt, concentration, light, agitation etc.) [3, 4]. Measuring protein aggregation is challenging as "aggregates" can vary in size from nm to mm [4, 5]. The most difficult and important to measure are the small (nm) soluble oligomers, which can be formed by self-assembly from either a partially unfolded state as suspected in amyloid related diseases, or from a protein in its native state as frequently observed biologically via various environmental factors [6]. These often lack covalent linkages and instead are held together by a complex network of weak forces such as hydrophobic interactions, van der Waals forces, electrostatic, ionic interactions, and hydrogen bonds. Oligomers are thus fragile structures that exist in delicate equilibrium with other species, making their accurate measurement in solution difficult, especially if small size changes are involved.

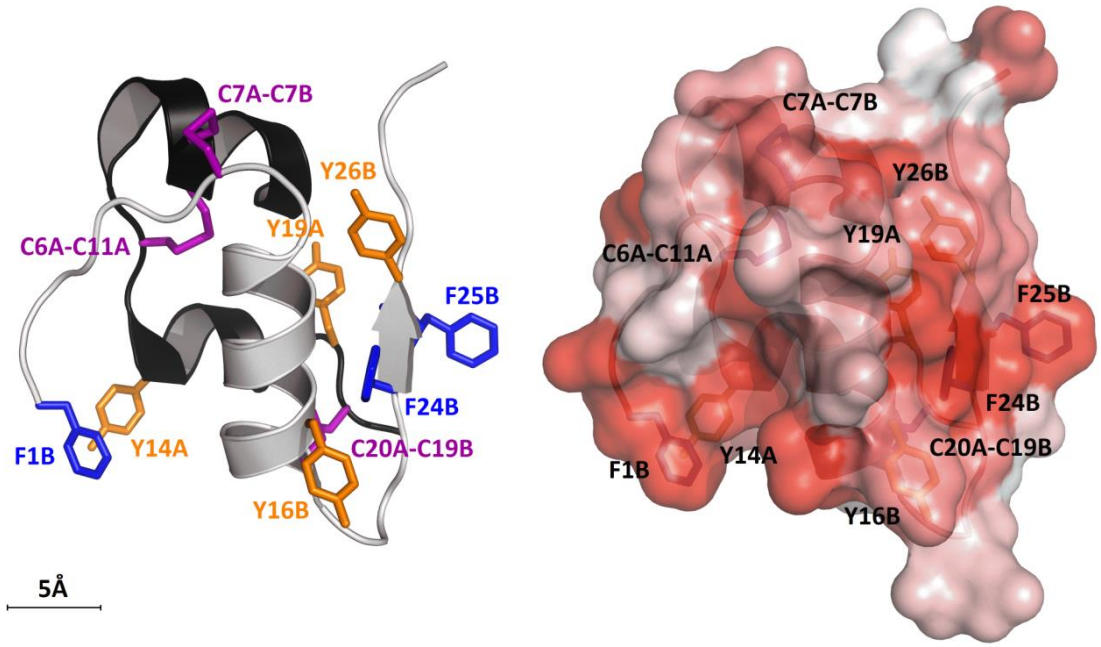

Figure 1: (Left) Representation of monomeric insulin (molecule 1 conformation) extracted from 2-Zn porcine hexamer crystal (PDB file 4INS) with tyrosine (Y), phenylalanine (F), and disulfide bridges (C) shown. The A and B chains are coloured in black and light grey respectively; (Right) Same molecule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

Insulin (Figure 1), a small disulphide-bridged peptide of 51 residues, for example has a natural tendency to form oligomers in solution. Insulin is critical to treating diabetes which is a major healthcare issue today, with patient number expected to rise ~ 693 million in 2045 [7], and a global healthcare cost estimated at ~700 B$ in 2017. This peptide, which is active and present in blood as monomer, can also exist as a dimer, and be stored as densely packed hexamer in the pancreas [8]. In solution, all these forms are in equilibrium, which is affected by multiple factors including primarily concentration and pH, but also temperature, ionic strength, and presence of

divalent cations (Zn^{2+} especially) [9-13]. Knowing the precise oligomeric state of insulin is important as it affects formulation stability and bioavailability.

Size Exclusion Chromatography (SEC) is the most common method for aggregate content analysis in protein therapeutics [4, 5, 14]. One of its main disadvantages for self-assembled oligomer analysis, apart from potential interactions with column surface, is that diluting a protein solution and exposing it to a different elution buffer can result in dissociation and/or dynamic re-equilibration of non-covalent oligomer populations, as previously shown for insulin [15]. SEC can thus be relatively insensitive to the original oligomer state (type and distribution), which is also the case for other QC methods like SDS-PAGE. Photo-induced cross-linking (PICUP) with SDS-PAGE could be a potential solution, unfortunately the use of EDTA as a reaction quencher [16] is problematic. This is because EDTA is a strong chelating agent that will complex the Zn^{2+} present this causing any insulin hexamer to dissociate to monomer and/or dimer. It is also suggested that the SDS-PAGE method itself can modulate oligomer distributions [17] and this needs to be carefully considered when comparing with other techniques. Other alternatives that do not require sample modification like DLS or analytical ultracentrifugation are thus more suitable to this type of analysis, but suffer respectively from lack of resolution or slow throughput/expensive equipment [5, 14].

Therefore there is a need for an alternative, minimally invasive method, and intrinsic fluorescence is advantageous in that it is non-destructive, can probe the protein native states without perturbation, while being also fast and sensitive [18]. Intrinsic emission from Insulin originates for example from four tyrosine (Tyr) residues, with a minor contribution from three phenylalanine (Phe) residues. When implemented in a multidimensional measurement like EEM, it offers a complete picture of protein emission space. EEM profiles in particular are sensitive to changes in fluorophore microenvironment, including changes in Förster Resonance Energy Transfer (FRET) and other photophysical process affecting emission. EEM are thus protein fingerprints, providing a sensitive tool for detecting protein structure changes. When coupled with factor-based methods like MCR or Parallel factor analysis (PARAFAC), the complex spectral overlap constituting this emission space can sometimes be resolved, and contributions of individual fluorophore identified. An additional information layer from emission anisotropy of macromolecules can also be added via pEEM measurements [19, 20].

The deconvolution of EEM/pEEM data by PARAFAC or MCR can however be ruined by the presence of RS (and Raman scatter to a lesser extent) that behave as non-bilinear elements during modelling [21, 22], which probably explains why most researchers discard this data and rarely use it. However, the RS data in EEM measurements is equivalent to that obtained in static light scattering (SLS) experiments, and as such it contains information about particle size changes. A recent study from our lab examined the possibility of using pEEM for the discrimination/analysis of Immunoglobulin (IgG) type proteins [23]. These proteins are much larger (155 kDa) than insulin and represent a different set of analytical challenges. IgG for which the non-reversible aggregates are a more important quality parameter, could be analysed by SEC. Changes in the population of non-reversible aggregates were induced by stressing, and the aggregate content (as determined by SEC) could be followed using pEEM and a Partial Least Squares (PLS) regression model. However, it was not possible to identify or quantify the individual IgG oligomers in this study.

Here the focus and sample system are both very different. First, we are specifically targeting the analysis and quantification of reversible oligomers in solution, that are much more fragile structures formed from proteins in their conserved native state. Second, the much smaller size of insulin (5.7 kDa, ~ 2 nm hydrodynamic radius, Rh) versus IgG (155 kDa, Rh 5-6 nm) represents a completely different sample class, with insulin being of a size that can show relatively weak scatter efficiencies, limiting DLS analysis or making it impractical if oligomer mixtures must be resolved. Third, due to the absence of tryptophan (and its strong environmental sensitivity), insulin and its oligomers are more difficult to characterise by fluorescence compared to other proteins. However, the low number of fluorophores present and their key engagement in oligomer formation suggests that curve resolution might here be more fruitful. We have previously demonstrated fluorophore resolution for non-interacting small molecules [20], but for IgG this was limited [19] even under optimal data pre-treatment due to the much larger number of interacting fluorophores and the presence of emission non-linearity, and rank deficiencies. Fourth, for the previous IgG study [23], the RS information was used in a very coarse fashion, as just a quick method to discriminate changes, whereas here we delve into much greater detail and show for the first time how RS band extracted from EEM/pEEM measurements can be exploited to monitor size change explicitly. Finally, insulin and peptide-based drugs are a different class of therapeutic agents compared to monoclonal antibodies, as such the analytical methods required are significantly different, particularly optical based methods based on light scattering or fluorescence emission.

Here we show how to use the full EEM/pEEM measurement to identify/discriminate the oligomeric state and quantify the composition of various reversible oligomers in solution by simultaneously using the RS and emission data. We show also that pEEM contains stronger and more informative scatter signal than EEM, more sensitive to particle size changes but also characterised by polarization data that ultimately provide an additional layer in the identification of protein oligomeric state. Finally, we show how the pure fluorescence part of the measurement, when coupled with MCR can easily be used to quantify oligomers content in solution, information that could not be obtained from pure scatter-based method like DLS.

2. Materials and Methods

2.1 Materials

Human Insulin (# I0908), hydrochloric acid, acetic acid, formic acid, ammonium formate, Tris-base, Tris-HCl, L-Tyrosine, sodium hydroxide, and sodium chloride were purchased from Sigma-Aldrich (Arklow, Ireland) and received without further purification. A single lot of insulin (batch number SLBP5090V, Sigma, Arklow, Ireland) was used to make all solutions. According to the certificate of analysis this lot contained 0.4% of zinc (w/w of insulin), which was enough to form hexamers, as it corresponded to two Zn^{2+} cations per insulin hexamer [10-12].

2.2 Insulin samples

Solutions which favour the formation of specific oligomers (monomers, dimers, *"tetramers"*, or hexamers) were prepared in various buffers, according to well-established procedures described in the literature [9-13]. This avoided the use of concentration driven aggregation, which would have complicated this fundamental fluorescence study by introducing very large intensity variation and inner-filter effects. Working at constant concentration also

formed a closed system, suitable for monitoring self-assembly evolution in fixed volumes. The *"tetramer"* case is unique, as its existence in solution is controversial and requires zinc free conditions (*vide infra*).

Briefly, monomers were obtained in 20% acetic acid (pH 2), dimers in 25 mM HCl (pH 1.6), *"tetramers"* in 20 mM formate buffer (pH 3), and hexamers in Tris buffer (pH 7.4). The solution ionic strength was controlled by adding 0.1 M NaCl, and the pH adjusted to within 0.05 units of the desired value, using a calibrated pH meter (Eutech instruments, model Cyberscan pH 10, Thermo Scientific, Dublin, Ireland). Considering that native wild-type human insulin is soluble in dilute acidic conditions (pH ~ 2/3, isoelectric point 5.3), solutions of monomers, dimers, and *"tetramers"* were prepared by dissolving insulin directly in the appropriate buffer. However due to low insulin solubility at neutral pH, hexamers could not be formed directly in the 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 reaching a final 20 mM Tris composition and then pH was then adjusted to 7.4 with 1.0 M NaOH.

All insulin solutions were freshly prepared for analysis in order to avoid changes in the sample composition and a potential re-distribution of oligomer populations [9-13]. At the beginning of each analysis day, ~10 mg of insulin was 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 ~ 30 min in the dark at room temperature. This stock solution was then sterile filtered (Minisart, 0.2 µm, Sartorius Göttingen, Germany) and the insulin concentration (2 mg/mL) determined by absorbance spectroscopy using an extinction coefficient of 1.0 $(mg/mL)^{-1}.cm^{-1}$ at 276 nm [24]. From this solution, three replicate samples were prepared per day for analysis, and this process was repeated to create 12 independent replicates per insulin form (four sets of three), designated R_m1-12 (monomer), R_d1-12 (dimer), R_t1-12 (tetramer), and R_h1-12 (hexamer). Data collection took several weeks using the same stock buffer solutions which were carefully stored at 2-8ºC.

2.3 L-Tyrosine solutions

A control experiment was performed to determine if changing the buffer affected the absorbance and fluorescence spectra of free L-Tyrosine in solution. The optical density had to be sufficiently low to avoid IFE that might otherwise distort the EEM spectra. For that reason, L-Tyr was measured at a concentration of ~ 1.8×10^{-4} M, which gave a maximum absorbance of~ 0.085, matching the absorbance of the 0.2 $mg.mL^{-1}$ insulin solutions.

2.4 Instrumentation

All samples were measured at 25°C and were thermally equilibrated prior to measurement. Absorbance and fluorescence spectra were measured in 0.4×1 cm quartz cuvettes (Lightpath Optical, Cranborne, UK), with the short pathlength used for excitation. Absorbance spectra (700-200 nm) were measured on a Cary 60 spectrometer (Agilent Technologies, Mulgrave, Australia) and were baseline corrected by measuring the corresponding buffer as blank. Most data were collected at 2 nm resolution, but some additional data was collected at higher resolution (0.15 nm) with an independent set of insulin samples.

Fluorescence measurements were performed on a Cary Eclipse spectrophotometer (Agilent Technologies, Mulgrave, Australia) fitted with a temperature controlled multi-cell holder. EEM data were collected over an excitation/emission range ($\lambda_{ex}/\lambda_{em}$) of 240–330/270-400 nm, with the

following fixed parameters: a resolution of 2 nm, excitation/emission slit widths of 10 nm, a PMT voltage of 650 V, and a scan rate of 1200 nm.min^{-1}. Polarized-EEM spectra were first measured using bespoke UV transmitting polarizers [25] and their alignment was verified using a diluted (0.1%) ludox solution (r > 0.98, Figure S-1, *Supplemental Information*, SI). pEEM data were collected using four different polarization configurations: vertical-vertical (VV), vertical-horizontal (VH), horizontal-vertical (HV) and horizontal-horizontal (HH). Non-polarized EEM spectra were also collected, by replacing the polarizer accessory with a 3%T attenuation filter (neutral density filter) in emission to avoid signal saturation. Spectra were not corrected for instrument response.

DLS measurements were performed on a Zeta Sizer NS (Malvern Panalytical, Malvern, UK) in back scatter mode (173° angle, λ_{ex} = 633 nm). Sample solutions were placed in 1×1 cm clean square polystyrene cuvettes (Sarstedt #67-754, Nümbrecht, Germany). Cuvettes were filled with ~ 1.2 mL of solution, and measured at 25°C, after two minutes of equilibration time. Each measurement consisted of 30 × 10 second runs (no attenuator used), and three measurements were averaged to create a single (less noisy) spectrum per sample.

2.5 Data Analysis

For the pEEM data, two measurements (EEM_{HH} and EEM_{HV}) were used to calculate the G-factor [26]. EEM_{VH}, once multiplied by the G-factor gave the perpendicularly polarized $EEM_{\perp}$. The parallel polarized $EEM_{||}$ is the same as the EEM_{VV} spectra. The anisotropy (r) at each $\lambda_{ex}/\lambda_{em}$ point in an EEM can be calculated using [18, 26] :

Equation 1

$$EEM_r = \frac{EEM_{||} - EEM_{\perp}}{EEM_{||} + 2 \times EEM_{\perp}}$$

All EEM data were then subjected to RS removal with the goal being to separate the RS band from the fluorescence signal as accurately as possible, because this contains information about protein and particles size and is thus sensitive to self-assembly processes. We showed previously that due to spectral overlap, classical interpolation was not as efficient in extracting the RS band as modelling it using PCA or PARAFAC once shifted into a bilinear matrix [22]. This generated accurate anisotropy measurements over a wider emission range and produced less spectral distortion in the overlap region [25]. Here the RS band was modelled as a low rank bilinear matrix using PARAFAC which was faster than PCA. Raman scatter was removed using a buffer blank subtraction, but this must be done with caution *vide infra* [26].

Data analysis was performed using the PLS_Toolbox ver. 8.2.1 (Eigenvector Research Inc., Manson, WA, USA), MATLAB ver. 9.1.0 (The Mathworks Inc., Natick, MA, USA), and in-house written codes. EEM data were organized by sample (mode 1), λ_{em} (mode 2), and λ_{ex} (mode 3), and had to be unfolded due to the bilinear nature of MCR. EEM data were usually column wise augmented and arranged so that (sample × λ_{ex}) or (sample × λ_{em}) formed the "concentration" matrix (C), while λ_{em} (or λ_{ex} respectively) formed the pure "spectral" matrix S^T. This enabled fluorophore resolution with excitation and emission spectra being resolved separately in C and S^T [20]. Here one important objective was to fit sample EEM (*i.e.* the suspect "tetramer" samples, *vide infra*) with the spectra of the monomer, dimer, and hexamer for which we had "pure" EEM spectra. The entire spectral information (λ_{ex} × λ_{em}) had thus to be kept together in S^T (as per PCA), so that the sample EEM could be fitted via the application of equality

constraints (hard modelling). This left the concentration block (C) with the characteristics of a closed system, as all samples had the same insulin concentration. A closure constraint could thus be imposed in C to maintain mass balance, thus enabling direct quantification of species in solution.

MCR model convergence was ensured by an ALS routine [27], and a non-negativity constraint was used on both concentration and pure spectra profiles. The closure constraint also served as normalization to avoid intensity ambiguity during ALS, and all models reached the convergence criterion quickly with few iterations. This criterion was set so that the relative difference between the root mean square of the residuals matrix between consecutive iterations was <0.1%. Estimation of rotational ambiguities was done using MCR-BANDS [28].

Model quality was assessed using the percentage of variance explained, the lack of fit (LOF) [27], investigating the quantity of information left in the residuals, and by estimating the extent of rotational ambiguities. LOF estimates the difference between input data (D) and data reproduced by MCR-ALS ($C \times S^T$). LOF was calculated using equation 2, where d_{ij} correspond to an element of D and e_{ij} of the model residuals.

Equation 2

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

3. Results and Discussion.

3.1 DLS analysis

DLS data were collected on three fresh replicate samples of insulin, acquired several months after the other spectroscopic data (the DLS system was only acquired at the end of the study). These samples were prepared using the same methods and the same source insulin lot, and were as such, representative of the previous solutions made. Using the original solutions would have otherwise introduced a freeze-thaw cycle which can alter the oligomer population distribution in solution [29] or in some cases induce some degree of protein denaturation [30], often followed by aggregation.

Figure 2A shows the raw intensity based DLS data, which is preferred here as it is the closest to what DLS measures (fluctuations in scattered light intensity) and to the RS measurements made in EEM. The insulin data showed clearly that all distributions were monomodal and that the auto-correlation function (Figure 2B) was shifted to the right on going from monomer to hexamer containing solutions.

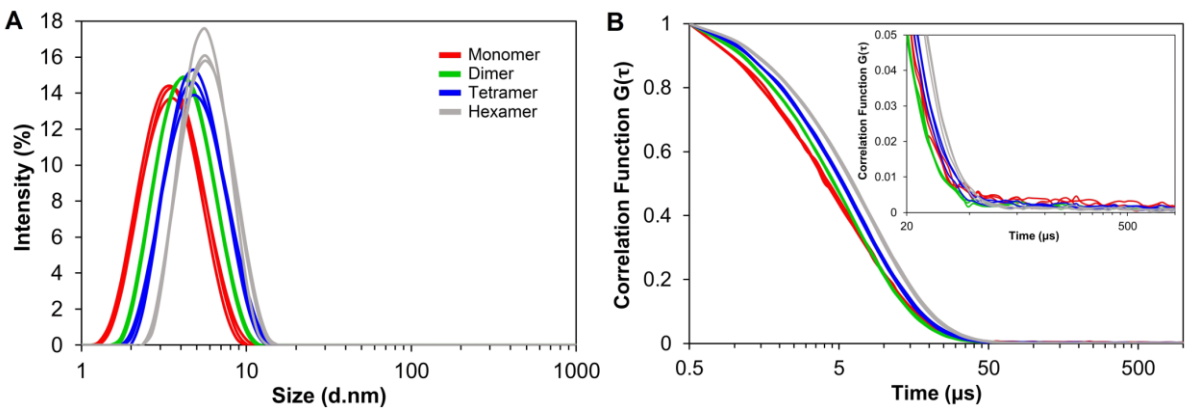

Figure 2: (A) Intensity based DLS data obtained from insulin monomer, dimer, "tetramer" and hexamer solutions (2 mg.mL^{-1}, triplicate sample solutions); (B) Normalised correlation functions with inset focusing on baseline noise.

The monomodal distributions were analysed using cumulant fit analysis (Table 1), to produce a Z-average size value (or cumulants mean) which represents the hydrodynamic diameter (D_H) and is the most acceptable value as defined by ISO standard 22412. Cumulant fitting produces also a second parameter, the polydispersity index (PDI) related to the width of the size distribution, and this is also sensitive to any factor (including noise) that would cause the correlation function to deviate from an ideal mono-exponential decay.

Table 1. Summary of the most important DLS measurements characteristics obtained on insulin (2 mg/mL). All measurements were averaged from triplicate samples.

Insulin Form	Z-Average (nm)	Reported DLS sizes in literature		Polydispersity Index	Count rate (kcps)	Intercept	Fit error $\times 10^{-3}$
		D_H (nm)	R_H (nm)				
Monomer	2.77 ± 0.06	2.75[31] 3.0[32]	1.3[33] 1.37[34]	0.21 ± 0.01	73.20 ± 0.62	0.63 ± 0.01	3.7 ± 0.3
Dimer	3.94 ± 0.04	3.9[32] 4.0[35]	1.8/1.9[36]	0.12 ± 0.01	112.53 ± 0.35	0.80 ± 0.01	1.0 ± 0.2
"Tetramer"	4.56 ± 0.02	5.0[32]	/	0.13 ± 0.01	164.43 ± 4.56	0.82 ± 0.02	0.7 ± 0.3
Hexamer	5.39 ± 0.05	5.6[37] 5.43[38]	2.6/2.7[36, 39]	0.09 ± 0.02	286.30 ± 1.83	0.88± 0.01	0.5 ± 0.1

D_H values (Table 1) all agreed with literature values, but the measurement quality data, raises issues over the monomer and dimer results, which is largely due to the low count rates (the manufacturer recommendation is >100 kcps). These very small particles, which do not scatter effectively, produced noisy correlation functions which resulted in cumulant fit error values that were relatively high even though the data were clearly monomodal. PDI was affected in particular by noise in the gradient and baseline (Figure 2B-inset), and values were higher for samples with noisier DLS data (Table 1), with PDI varying proportionally with cumulant fit error, and inversely to the count rate and intercept value. This was true for all samples, apart from the "tetramer" samples, where PDI slightly increased. It reached a value of 0.09 for the hexamer which is probably the most mono-disperse sample, or at least the one that scattered the most, generating the most accurate value presented. The minimum PDI value below which non-spherical real samples are considered monodisperse have not been officially defined, and vary depending on the sample type, for example <0.2 for polymer nanoparticles and <0.3 in liposome-based drug delivery systems being commonly accepted [40]. Values < 0.05 are never seen except for well-defined latex standards, and it seems quite reasonable that PDI <0.1 is representative of monodisperse, quasi-spherical proteins.

It is important to note however that the literature was quite sparse regarding the size of insulin tetramer measured by DLS. Kodima *et al.* reported a value of 5.0 nm in zinc free buffer which was significantly higher than the value measured here. This is a critical condition for tetramer formation [10-12, 41], that was not fulfilled here as the same amount of zinc was present in all samples. Finally, it should be noted that for DLS measurements, insulin monomers and hexamers only differ by a factor of two in size which is much less that the three or more

recommended for DLS to clearly resolve different species. Thus, we can expect mixtures of insulin oligomers to not be resolvable by DLS.

3.2 Absorbance Spectroscopy

Absorbance spectra were measured to first measure insulin concentrations, and second to determine the reproducibility of replicate solution preparation (Table S-2, post scatter correction). Absorbance spectra potentially contain information about oligomer state and thus the spectra of the insulin and control were compared (Figure 3A/B).

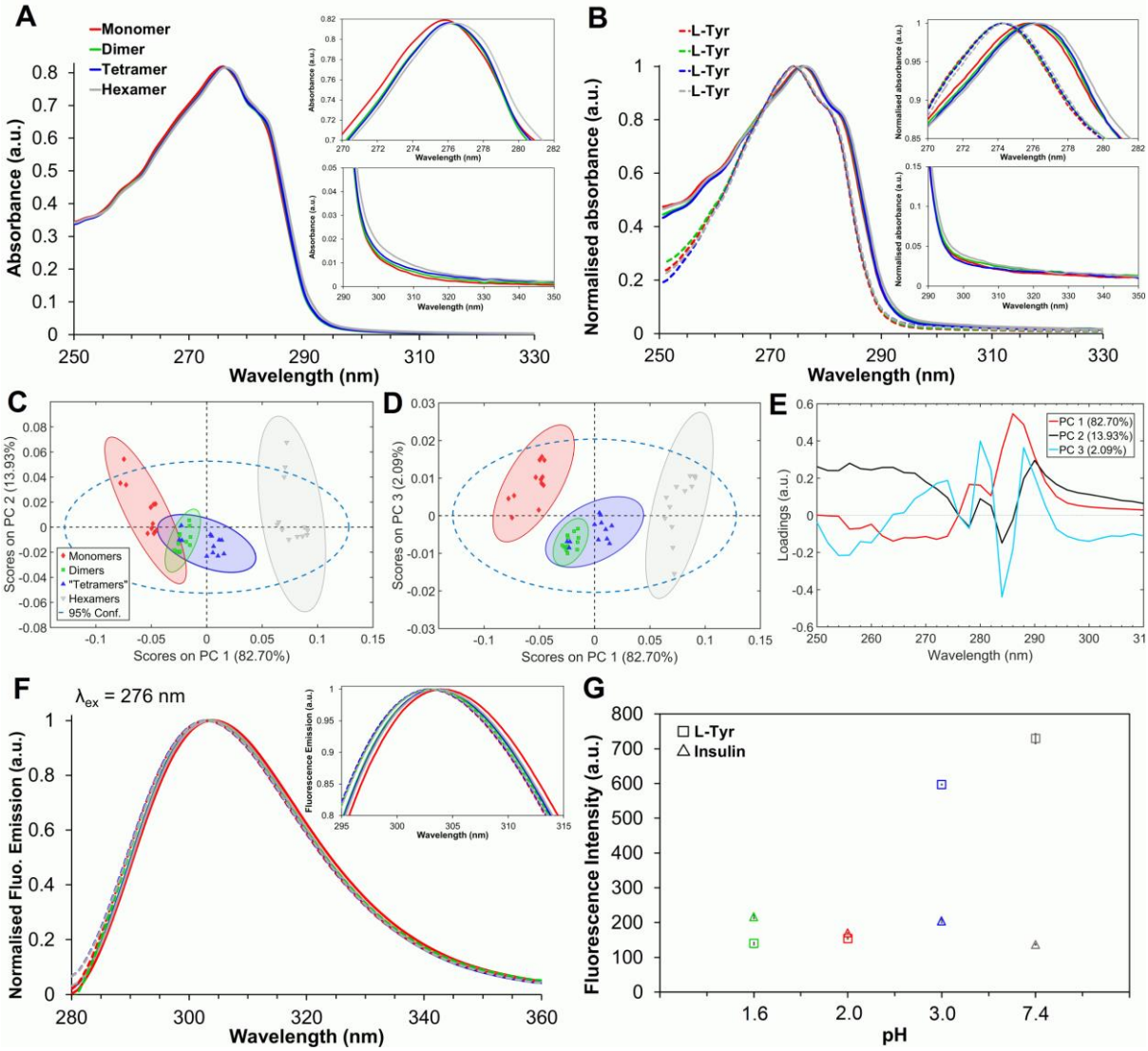

Figure 3: ***(A)*** Blank-corrected absorbance spectra of the insulin (2 mg.mL^{-1}) oligomers, 2 nm resolution, averaged over 12 replicate measurements; ***(B)*** Normalised absorbance spectra of diluted 0.2 mg.mL^{-1} insulin solutions (full line), and L-Tyr (dash, 1.8×10^{-4} M), measured with a higher resolution (0.15 nm) and averaged over 3 replicate measurements (each measured 3 times); ***(C/D)*** Scores scatter plots and ***(E)*** loadings for the 3 component PCA model built using absorbance spectra of insulin oligomers (2 mg.mL^{-1}), after mean-centring & normalisation; ***(F)*** Fluorescence emission (λ_{ex} = 276 nm) spectra (blank corrected, normalised) of the diluted solutions. ***(G)*** Fluorescence intensity variation of the diluted solutions upon change of pH buffer.

In terms of obvious differences, all insulin spectra (Figure 3A) had shoulders at 258 and 264 nm due to Phe residues which were absent in L-Tyr spectra. Furthermore, insulin absorption spectra had maxima at ~276 nm (as previously reported [42, 43]), whereas L-Tyr spectra showed maxima ~ 274 nm typical for polar environments (Figure 3B). The insulin red-shift of ~ 1.5 nm was due to some residues like TyrA19 (Figure 1) being buried in hydrophobic regions of insulin [44-46]. Changes in Tyr environment is known to only produce small spectral shifts which are more evident in the absorbance/excitation rather than in emission spectra. For example, NAYA, often used to model Tyr in proteins, only has a λ_{max} shift of 4 nm (to 278 nm) on going from a completely polar (water) to a non-polar (dioxane) solvent [47]. For proteins, shifts of <2 nm are thus considered real and indicate significant changes in Tyr environments, similar to other class A protein studies [45, 46].

Figure 3 also showed that absorbance red shift slightly increased for dimers and hexamers, while L-Tyr spectra remained unchanged in these buffers. λ_{max} values increased from 275.7 (monomer) $\rightarrow$ 276 (dimer) $\rightarrow$ 276.2 (tetramer) $\rightarrow$ 276.4 nm (hexamer). This shift is small, but agrees with the fact that in monomeric insulin, three Tyr residues are already either completely (TyrA19) or partially (TyrB16 and TyrB26) buried in hydrophobic environments. This progressive λ_{max} shift is characteristic of Tyr being located in more hydrophobic (or non-polar) environments [44]. This oligomer dependant shift is caused by TyrB16/B26 and TyrA14 that become progressively buried in hydrophobic interfaces upon dimer and hexamer formation respectively. Absorbance spectra Figure 3A, also showed the presence of scattered light as an increased baseline at longer wavelengths. Using the ratio of absorbance at 350 nm (where most proteins do not absorb) and at 280 nm, an aggregation index (AI) useful for identifying large concentrations of aggregates, was calculated (Equation S-1, SI) which showed an increase upon oligomer formation (Table S-3, SI).

These simultaneous changes suggested that multivariate analysis might be useful in better discriminating these oligomers. To build the PCA model, all spectra were normalised (using infinity norm to remove signal intensity changes) and restricted to 250-320 nm to decrease the amount of noise being modelled. A three-component model was built on mean-centred data, that explained over 98.7% of data variance. The model (Figure 3C/D/E) showed that apart from two outliers (due to mainly scatter contamination, Figure S-2, SI), oligomer separation was along PC1 (82.7%). The loadings showed a peak (270-290 nm) representing the spectral red-shift and a tail (290-310 nm) due to increased light scatter. Therefore, absorbance spectra, even based on small spectral difference, were able to separate monomers from dimers/tetramers and hexamers, enabling an assessment of changes in protein quaternary structure. However, this is not a particularly robust measurement as the spectral changes were small. We thus looked at a more sensitive measurement with larger spectral changes, such as EEM & pEEM which also provide size information.

3.3 2-D fluorescence analysis of L-Tyr and insulin:

The buffer effect on L-Tyr and insulin emission was checked, using diluted samples (absorbance < 0.1). Figure 3F shows that L-Tyr emission maximum was 303 nm in all buffers, while insulin emission was slightly red shifted to ~303.6 nm (~304 nm for monomer) [42] with no evidence (at longer wavelength, data not shown) for tyrosinate or dityrosine emission. Compared to tryptophan (Trp) which can have polarity induced Stokes shifts of ~ 50 nm, an

obvious observation is that L-Tyr emission is rather insensitive to change in its local environment [18]. Tryptophan, which has two nearly isoenergetic excited states (^1L_a and ^1L_b) can emit from ^1L_a (as seen in proteins) due to the direct involvement of the pyrrole's ring nitrogen atom, increasing the sensitivity of that dipole to hydrogen bonding, and stabilising it at energy levels lower than ^1L_b. Tyr owes its fluorescence to the presence of a phenol group where the excited states have too large an energy gap to overlap [48]. For that reason, above λ_{ex} > 250 nm, Tyr emits solely from ^1L_b, with no involvement from a nitrogen which means less sensitivity to hydrogen bonding.

Instead, a variety of effects involving hydrogen bonding to the phenol group can produce small spectral shifts [47]. For example, a small Tyr emission red shift can appear if the solvent has a low dielectric constant and high hydrogen bond acceptor strength, whereas a blue shift would appear with a proton donor. The fact that proteins were recently shown to have lower dielectric constants in hydrophobic as opposed to exposed domains [49] explains why insulin spectra were slightly red-shifted compared to L-Tyr. The small blue shift observed upon dimerization might then have more to do with multiple interactions occurring at the dimer interface, that actively involve TyrB26 and B16 [8, 50]. In any case, these effects were small, specific to insulin, and the low sensitivity to changes in local environment agrees with the behaviour of NAYA in the presence of various solvents and ligands [47, 51].

The L-Tyr fluorescence parameter most affected by changing buffer is the quantum yield (Figure 3G), which is pH dependant and constant only between pH 4-8 [52]. In alkaline conditions, it drops due to ground state ionization of the phenolic hydroxyl group (pKa=10.3), while below pH 4, it drops due to protonation of the carboxyl group (pKa = 2.3) [52]. Figure 3G shows that L-Tyr intensity drops with decreasing pH, and acetic acid is known to not quench Tyr emission [53]. For insulin, the effects were different and rather unpredictable, with intensity tending to drop upon oligomer formation and with increasing pH. This would suggest that in insulin, intra/inter molecular mechanisms probably govern Tyr emission rather than simple solvent quenching (consistent with existence of extensive FRET in the molecule).

There is thus a buffer effect on L-Tyr emission intensity, and even though insulin seems to show intensity variations that were unrelated to the L-Tyr case, we decided that it was safer to build all models using normalised (to the maximum by infinity-norm) spectra. This had a major advantage of focusing only on spectral shape variation, which in practice should be more robust than absolute intensity changes, and thus less sensitive to measurement errors. This also enables building models for highly concentration insulin solutions, where inner filter effect (IFE) correction is not feasible.

3.4 Measuring protein oligomerization using Rayleigh scatter

Oligomerization produces size increase, and as such can be characterised using Rayleigh scatter intensity. An aggregation index (AI-F), based on the ratio (using λ_{ex} = 280 nm) between the light intensity at 280 and 340 nm, was applied here (Table S4, SI), but similarly to the absorbance-based AI, the resolution was poor as only hexamers could be discriminated. AI-F can also suffer from the pH dependency of Tyr emission (Figure 3G), meaning that a more robust tool was needed.

Because most proteins (especially insulin) are much smaller than the wavelength of incident light (typically << 1/10), the light scattered by a single protein (Equation S-2, SI) obey

the Rayleigh scatter regime, a simplification of Mie theory. For dilute protein solutions, the total scattered light intensity measured can be modelled by the Debye-Zimm relationship which can be approximated for macromolecules (Mw < 50×10^6 g.mol^{-1}) by [54]:

Equation 3

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

Where c is concentration, R_g the radius of gyration of the protein, M_r the weight-average molecular weight, λ the measurement wavelength, θ the angle between the scattered and incident light, B a virial coefficient used to represent non-ideality, K an experiment constant and R_θ is the Rayleigh excess ratio. The scattered light intensity is thus dependant on protein M_r and solution concentration, which is what is typically measured by SLS.

A conventional right angle fluorometer can also be used as a simple SLS photometer [54] if excitation and emission are set to the same wavelength. Measurements must however be performed in a region of minimal absorption to avoid the *so-called* colour effect, resulting otherwise in abnormally low scattering intensity by reabsorption, as observed here between λ_{ex}=260-280 nm (Figure 4D/E). This can be overcome using multidimensional EEM, and if the objective was to study only the RS light, one could simply select the regions of pure RS signal that appear above λ_{ex} >280 nm. However, because the method involves MCR at a later stage which requires complete RS removal, it made more sense to extract the RS band in its entirety.

This was done accurately (see section 7, SI for more details), so that the residual scatter left in $EEM_{||}$ was < 0.4% of fluorescence intensity (Table S-5, SI). Also because the blank sample (non-absorbing) tended to scatter more light than the protein containing solutions for λ_{ex} < 280 nm (Figure 4D/E), it was preferable to extract RS data prior to blank subtraction and avoid negative spectral artefacts. More details about this procedure, showing the two equivalent methods for extracting RS and fluorescence data that include Raman correction, can be found in section 8, SI.

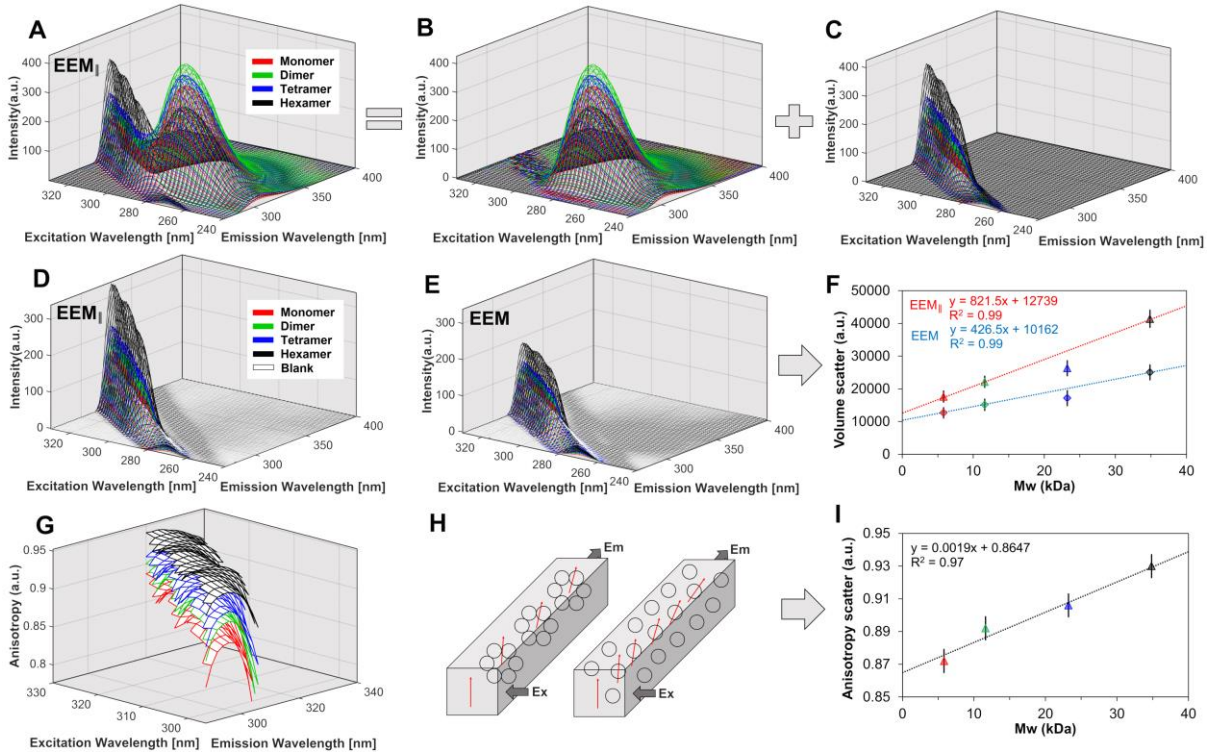

Figure 4: Illustration of how $EEM_{||}$ spectra ***(A)***, were separated into its fluorescence ***(B)*** and Rayleigh scattered (RS) light ***(C)*** components; RS bands from $EEM_{||}$ ***(D)*** and EEM ***(E)*** spectra respectively; ***(F)*** the correlation obtained between RS volume and average molecular weight (*Mw* in kDa); ***(G)*** *aniso*-EEM spectra (10% threshold) calculated for RS band upon oligomerization; ***(H)*** illustrates how the light scattered by a particle can be more depolarized (at equivalent concentration, volume) by smaller particles (right, monomer) than bigger assemblies (left, hexamer); *(I)* Correlation between scatter anisotropy (not blank corrected) and *Mw*. All data presented were averaged over 12 replicate measurements per insulin form, and the colour code is the same for all figures.

Once the complete RS band data was extracted, a simplification of Equation 3 could be applied [54]. First with a concentration of 2 $mg.mL^{-1}$ (= 3.5×10^{-4} M) and *B* being in $mL.mol.g^{-2}$, the term *2Bc* could be ignored. R_g values for different insulin states were obtained from reference [12], a small angle x-ray scattering (SAXS) study of insulin performed in the exact same experimental conditions (buffers and concentration) as here. Using R_g values of 11.6 (monomer), 14.9, 17.8, and 19.8 (hexamer) Å, the $16\pi^2R_g^2/3\lambda^2$ terms at 250 nm were 0.0011, 0.0019, 0.0027, and 0.0033 respectively. Furthermore, with a 90° angle fluorometer, $\sin(\theta/2)^2$ = 0.5, which meant that Equation 3 simplified to:

Equation 4 $$\mathbf{R_\theta = KcM_r}$$

In this equation, R_θ (Rayleigh excess ratio) is defined for polarized light, as the ratio between the excess scattered light intensity (after blank correction) and the incident light intensity. In SLS measurements, blank subtraction is required however there are cases where dilution is not feasible and instead one wants to monitor changing particle size at a fixed concentration. In this case, R_θ does not need blank subtraction as the blank is the same for all samples, and R_θ can simply be replaced by the measured scatter. This is what was applied here, and instead of considering the scatter at a single wavelength, the volume under the Rayleigh scatter curve (V) was calculated with the objective to have a value less sensitive to noise and instrumental errors. In theory, V is equal to:

Equation 5 $$\mathbf{V = \int_{ex}\int_{em} I_{\lambda_{ex},\lambda_{em}}\, d\lambda_{ex} d\lambda_{em}}$$

But taking into account the discrete nature of the data, V was instead calculated as the sum of all variables:

Equation 6 $$\mathbf{V = \sum_{ex}\sum_{em} I_{\lambda_{ex}\lambda_{em}} \Delta\lambda_{ex}\Delta\lambda_{em}}$$

Considering that each buffer promotes a particular insulin form, solutions should be relatively homogeneous, and M_r was replaced by the *Mw* of human monomer insulin, dimer, tetramer and hexamer. Also, because *c* was constant here, V could be directly plotted against *Mw*.

Figure 4F clearly showed for both EEM and $EEM_{||}$ that the monomer, dimer, and hexamer samples satisfied this relationship. Only the "tetramer" samples did not agree and only scattered slightly more light (~15%) than the dimer solutions. This level of scattering could correspond to a homogeneous solution of trimers, however in the presence of zinc at least, monomers, dimers, and hexamers are the forms preferably adopted by wild-type insulin, in preference to oligomers composed of odd numbers of molecules, in agreement with numerous studies using

SAXS, mass spectrometry, analytical ultracentrifugation or other light scattering techniques [9-12, 37, 38, 41, 55, 56].

This observation suggests that the R_t1-12 samples, because they were prepared in the presence of zinc, were instead a heterogeneous mixture of monomer, dimer, and hexamer. The RS data (Figure 4F) and DLS measurements (R_h< 5 nm) suggested a majority of dimers with some hexamer, while monomers were expected to be a minor component considering the absence of acetic acid to stabilise it at this concentration and pH. This is mostly supported by the fact that the insulin tetramer, for which no crystal structure exists, was only recently observed in presence of zinc as a metastable intermediate in the dissociation of hexamer to dimer [56]. Thus formation of the tetramer requires the absence of zinc [11, 12, 32, 41] and even then, it would require a concentration of ~ an order of magnitude greater than used here under identical pH (3.3) conditions [41]. The reason being that at room temperature and for acidic to neutral pH, the dimer to tetramer association constant (K_{24}) was weak (~10^2/10^3 M^{-1}) whereas the monomer to dimer association constant was much larger, K_{12} ~10^4 to 10^5 M^{-1} [38, 41, 57]. This means that the dimer is the main insulin species in solution, except in presence of acetic acid (pH 2) where it is monomeric, or in presence of zinc at neutral pH were hexamerization is favoured (K_{26} ~ 10^{10} M^{-2}) [37, 38, 41, 55, 57]. These are also the three oligomers of most interest and the only ones measured when looking at insulin formulations (*e.g.* rapid versus long acting) as highlighted in a recent study [58], reinforcing the relevance of our study.

The scatter volume correlated thus with oligomer size and could also be used to identify when solutions of a specific oligomer become more heterogenous due to variations in sample preparation or handling. A similar correlation can be implemented with DLS (Figure S-5, SI) using the mean and derived count rate, providing another set of useful information to complement the size and distribution widths obtained from fitting the correlation curves.

Another point to note was the steeper regression slope observed for $EEM_{||}$ in Figure 4F compared to EEM, indicative of its greater sensitivity to changes in particle size, and thus to potential small contributions from HMW species. This was expected as Rayleigh scattered light is strongly polarized, but explaining fully this behaviour required further investigation. Many parameters such as particle size and concentration, refractive index (RI) ratio (particle/solvent), particle shape and distribution will affect the polarization of scattered light, with different effects in the Mie or Rayleigh regime [59, 60]. Here considering the fixed concentration, the similar particle shapes (globular proteins), and that the samples were populated mainly by one species (except the *"tetramer"*), the only remaining variables were particle size changes and RI ratio. However, in the Rayleigh regime and particularly with linearly polarised light, RI changes only have a small impact on scattered light polarization [60], leaving size change as being the most critical parameter.

RS anisotropy increased upon oligomerization (Figure 4G/I), but this also contained the buffer contribution. All blank buffer solutions had a similar RS anisotropy (Figure S-6A, SI) of r ~ 0.83, which contributed increasingly to the insulin solution anisotropies as particle size decreased, because less light was scattered from the smaller species. Thus, to get a more accurate look at the RS anisotropy change due to the protein scattering, the spectra had to be blank corrected (Figure S-6B, SI). As expected, the recovered values were larger, varying from r = 0.93 (monomer) to 0.97 (hexamer), and even though error increased as SNR decreased, the important point was that RS anisotropy still increased slightly with particle size. This is in

agreement with previous Monte-Carlo simulations [59, 60], that estimated smaller sized scatterers will depolarise RS more efficiently. In summary, changes in RS polarization provides thus an additional method for investigating changes in protein size in solution.

3.5 Quantifying protein oligomerization by fluorescence

The pure EEM/EEM$_{||}$ fluorescence which was separated from RS band (Figure 4B) are essentially a protein fingerprint, sensitive to both oligomer type and concentrations. In order to focus on EEM profile changes and eliminate noisy spectral regions, the normalised spectra were reduced to $\lambda_{ex}/\lambda_{em}$ = 244–296/280-360 nm (1107 variables), and lightly smoothed (Savitsky-Golay filter, width 5 and polynomial order 2) to decrease noise without affecting spectral shape. The 12 "tetramer" samples (R_t1-12), which were suspected to be instead mixtures of monomer dimer and hexamer, were used to test the method. Using MCR-ALS with spectral and closure constraints, EEM$_{||}$ data were fitted using the averaged-EEM$_{||}$ spectra from monomer, dimer, and hexamer (non-polarised EEM gave very similar results, as samples were fitted with known spectral profiles here).

Fitting was excellent (>99.98% of variance explained) with a LOF that remained small (~1.3%) even with the strength of constraints applied (Table S-6, SI). Similarly, the closure constraint only marginally increased LOF (1.3 *vs* 1.1%) showing its compatibility with the dataset modelled, with residuals <1000 times less intense than the emission. Another advantage of applying these strong constraints (spectral equality in particular), was that the MCR solutions were all free of rotational ambiguities (Table S-7, SI). Overall, this showed conclusively that these "tetramer" samples could be explained as mixtures of monomer, dimer, and hexamer with a high degree of certainty from the emission spectra (loadings are reported in Figure S-7, SI).

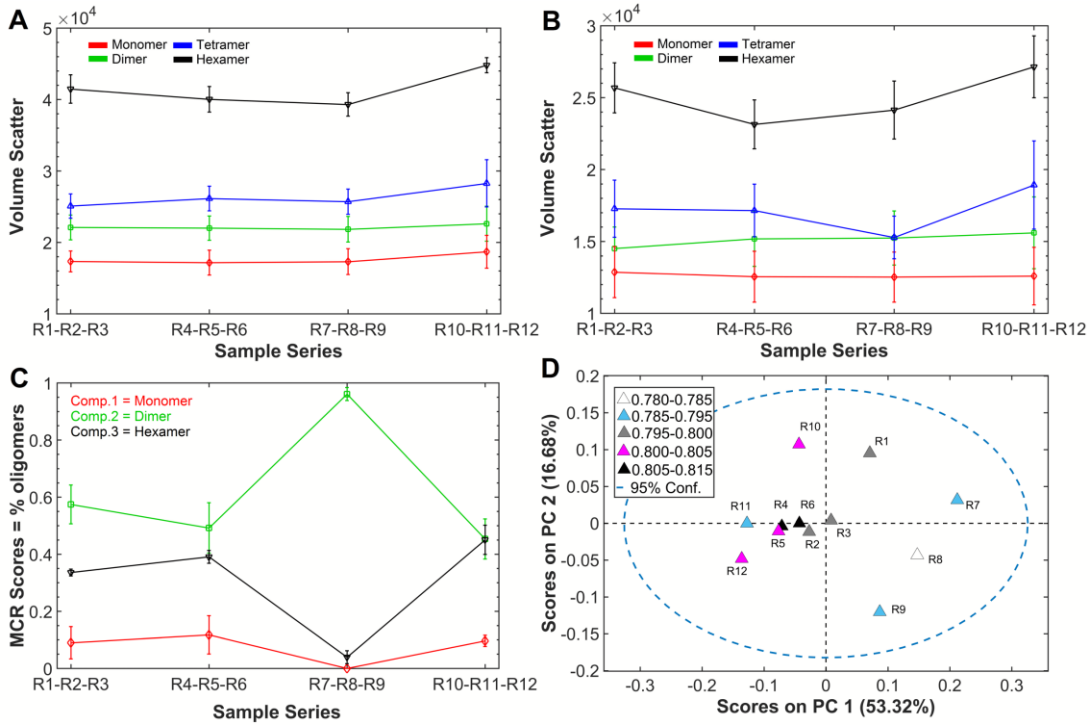

Figure 5: Volume under RS band for EEM$_{||}$ ***(A)*** and EEM ***(B)*** data, extracted for all insulin samples (*n=48*). Error bars were similar and due to the same 3 cuvettes used across the data collection which scattered light differently; ***(C)*** MCR scores obtained for the 12 "tetramer" normalised-EEM$_{||}$ samples (R_t1–12), using the averaged-EEM$_{||}$ spectra from monomer, dimer, and hexamer as spectral constraints; ***(D)*** Corresponding PCA scores scatter plot built using same "tetramer" data but mean-centred and coloured according to absorbance at 276 nm (no trend).

Looking at the results (Figure 5C), it appeared that most of the four "tetramer" sample series solutions (R_t1-12) were resolved as having more dimer than hexamer, with very little monomer (< 10% is best estimate based on accuracy of this MCR model), which made sense considering the concentration, pH, and data obtained from RS and DLS measurements. What was more interesting to see was that the different series were resolved with distinct oligomer amounts, which is tangible as solutions were made independently on different weeks, with unavoidable variation in their preparation (initial insulin dissolution, lab conditions, pH, glassware etc.). In particular, R_t7-9 (main outliers) were modelled as being >90% dimer with a small hexamer content, which was surprising, but agreed with the drop seen in RS signals (Figure 5B). The drop seen in RS (from $EEM_{||}$, Figure 5A) was lower as hexamers scattered proportionally more. Similarly, R_t10–12 were found to have a marginally higher hexamer content, which agreed with their larger RS signal (Figure 5A/B).

MCR resolution was also consistent with the trends seen in a PCA model (see Figure S-7 SI for the loadings plots), built using the same data (mean-centered) but with principal components constrained to orthogonality. Figure 5D shows that samples were distributed right to left according to decreasing dimer content (or increasing hexamer content), without displaying any trend related to absorbance at 276 nm, which was important as it ruled out the possibility of samples being spread according to an IFE induced distortion of the EEM. The MCR resolution thus made sense and highlighted the variability that can exist in oligomer content for replicate solutions which have been prepared very carefully.

The same methodology was then applied to the other monomer, dimer, and hexamer samples (Figure S-9, SI) which were assumed to be more homogeneous. Here the aim was to see if the method could be extended to quickly estimate the purity of self-assembly solutions and detect small traces of contaminating oligomers. The results were satisfying as R_m1–12 (insulin at pH 2 in 20% acetic acid) were found to be on average >95% monomer, with only 2% dimer, and 3% hexamer being resolved. Apart from agreeing with literature, this supports the point made previously about the DLS data from monomeric insulin where a relatively high PDI (~0.2) was due to poor quality data rather than true polydispersity. Similarly, R_d1–12 (insulin at pH 1.6) were found on average to be 89% dimer, with only 5% and 6% of monomer and hexamer being detected. MCR identified for both R_m1–12 and R_d1-12, the sample series that were slightly different in composition, matching the discrimination obtained by PCA models (the loadings plots for which are shown in Figure S-8, SI).

R_h1-12 (pH 7.4) were found to be on average >92% hexamer, with only 1% monomer being resolved, and up to 6% dimers. The fact that almost no monomer was resolved, and that some dimer remained, made sense as these samples were obtained by adjusting the pH from an initial dimer solution. Examination of individual samples revealed variations in RS signal, for which the lowest values were obtained with R_h4–6 and R_h7–9. This matched the MCR results as only these two sample series showed a dimer content (~12%, see Figure S-9E, SI). R_h10–12 on the other hand, which were expected to be pure hexamer, had a higher than expected RS signal, possibly due to the presence of higher-order species like dodecamers. Dodecamers could be present in small amounts [61], however, we did not have an EEM spectrum to use in the model. Yet knowing its Mw (~70 kDa), one can use the equation shown in Figure 4F to

extrapolate the RS volume for a 100% dodecamer solution, and thus predict that a 4% dodecamer content might explain the excess RS signal of the R_h10–12 samples.

4. Conclusions

Insulin oligomers, held together by weak inter-molecular forces, are difficult to analyse reliably by chromatographic means because dilution and mobile phase compositions impact the delicate equilibration between species in solution. Here we have shown that both absorbance and fluorescence spectroscopy can be used to discriminate between monomer, dimer, and hexamers of insulin in solution. Absorbance based discrimination using PCA worked on the basis of increasing both red-shift and scattered light, however, it could not discriminate the heterogenous "tetramer" samples from dimers. Furthermore, the measurement is based on a relatively small change in the absorbance band.

The use of pEEM measurements was better at discriminating both types and mixtures of oligomers, because both the fluorescence emission and the Rayleigh scatter were sensitive to changes in protein structure. We proved in particular that the RS volume correlated linearly with the average Mw of insulin oligomers in solution, offering an easy way to identify the main oligomeric state adopted by the protein. We also showed that $EEM_{||}$ was more sensitive than unpolarised EEM measurements to changes in particle size, with RS polarization providing additional information for monitoring self-assembly processes. The fluorescence component of the measurement was then used along with MCR-ALS and specific constraints (non-negativity, spectra equality and closure) to quantitatively estimate oligomer content in solution. This matched the discrimination obtained by PCA, proving that it was an efficient method to characterise and quantify changes in oligomer mixtures. The correlation between scatter and fluorescence-based analysis also showed that this pEEM based approach was better at characterising the variability of oligomer solutions in this size range compared to DLS. This was because DLS data was less reliable here due to weak scattering efficiencies, but more importantly because it cannot resolve mixtures of small particles with a size difference less than a factor of three [62]. Comparing the specific results of these insulin measurements with those from other labs using different techniques such as SDS-PAGE [16] or SAXS [39], have to be carefully considered as the sample handling [17], environmental chemistry [11, 12, 32, 41], and source insulin lots all have an impact on the specific oligomer compositions present in the sample under test. This variability in protein behaviour is well known and can make accurate comparison between different methods and literature results difficult unless measurements are made under very carefully controlled conditions. For solution state measurements, it is critical that orthogonal test methods be implemented on the exact same solution sample or at the very least on samples prepared under identical conditions from the same source lot. Even then, as we have shown here, variability in weakly bound oligomer composition in solution will probably still be present, and thus should be taken into consideration.

This work demonstrated thus the potential utility of pEEM measurements for soluble oligomer characterisation by using both the emission and scattered light signals. This is important in the context of recent recommendations by the FDA for the better analysis of subvisible aggregates in therapeutic protein products. Ideally, these methods should be fast, sensitive, non-destructive, and non-invasive which are all compatible with pEEM measurements. Here the ability to quantitatively assess in near-real time, protein oligomerization would be very

useful in *e.g.* purification or formulation steps (downstream), or towards monitoring other complex, self-assembly based processes involving proteins (such as the production of self-assembled particulate vaccines). This analytical methodology is also potentially applicable to highly concentrated protein formulations as long as the concentration (or rather the optical density) does not vary too much. Further work is required however, to generate more reproducible, lower noise pEEM data, which can be obtained using multichannel detectors, reducing sample acquisition time sufficiently to enable multiple spectral acquisition. In conclusion, we have shown that instead of discarding the Rayleigh scatter signal from EEM measurements, the complete emission and scatter signal should be used for the characterisation of proteins and their assemblies.

Supplemental information available

Supporting information is available and includes further details about spectral measurements and chemometric modelling.

Acknowledgements

YCB was supported by an 'EMBARK Initiative' Postgraduate Scholarship (Grant number GOIPG/2015/3826) from the Irish Research Council.

References:

[1] S. Hermeling, D.J.A. Crommelin, H. Schellekens, W. Jiskoot, Structure-Immunogenicity Relationships of Therapeutic Proteins, Pharmaceutical Research, 21 (2004) 897-903.

[2] A.S. Rosenberg, Effects of protein aggregates: an immunologic perspective, The AAPS journal, 8 (2006) E501-507.

[3] W. Wang, Protein aggregation and its inhibition in biopharmaceutics, International journal of pharmaceutics, 289 (2005) 1-30.

[4] H.C. Mahler, W. Friess, U. Grauschopf, S. Kiese, Protein Aggregation: Pathways, Induction Factors and Analysis, J. Pharm. Sci., 98 (2009) 2909-2934.

[5] J. den Engelsman, P. Garidel, R. Smulders, H. Koll, B. Smith, S. Bassarab, A. Seidl, O. Hainzl, W. Jiskoot, Strategies for the assessment of protein aggregates in pharmaceutical biotech product development, Pharm Res, 28 (2011) 920-933.

[6] S.E. Ahnert, J.A. Marsh, H. Hernandez, C.V. Robinson, S.A. Teichmann, Principles of assembly reveal a periodic table of protein complexes, Science, 350 (2015).

[7] N.H. Cho, J.E. Shaw, S. Karuranga, Y. Huang, J.D. da Rocha Fernandes, A.W. Ohlrogge, B. Malanda, IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045, Diabetes Research and Clinical Practice, 138 (2018) 271-281.

[8] T. Blundell, G. Dodson, D. Hodgkin, D. Mercola, Insulin: The Structure in the Crystal and its Reflection in Chemistry and Biology by, Advances in Protein Chemistry, 26 (1972) 279-402.

[9] A.K. Attri, C. Fernandez, A.P. Minton, pH-dependent self-association of zinc-free insulin characterized by concentration-gradient static light scattering, Biophysical chemistry, 148 (2010) 28-33.

[10] V.N. Uversky, L.N. Garriques, I.S. Millett, S. Frokjaer, J. Brange, S. Doniach, A.L. Fink, Prediction of the association state of insulin using spectral parameters, J Pharm Sci, 92 (2003) 847-858.

[11] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V.N. Uversky, A.L. Fink, Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism, Biochemistry, 40 (2001) 6036-6046.

[12] L. Nielsen, S. Frokjaer, J. Brange, V.N. Uversky, A.L. Fink, Probing the mechanism of insulin fibril formation with insulin mutants, Biochemistry, 40 (2001) 8397-8409.

[13] A. Ahmad, V.N. Uversky, D. Hong, A.L. Fink, Early events in the fibrillation of monomeric insulin, J Biol Chem, 280 (2005) 42669-42675.

[14] J.F. Carpenter, T.W. Randolph, W. Jiskoot, D.J. Crommelin, C.R. Middaugh, G. Winter, 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, 99 (2010) 2200-2208.
[15] R. Tantipolphan, S. Romeijn, J. Engelsman, R. Torosantucci, T. Rasmussen, W. Jiskoot, Elution behavior of insulin on high-performance size exclusion chromatography at neutral pH, Journal of pharmaceutical and biomedical analysis, 52 (2010) 195-202.
[16] M.T. Mawhinney, T.L. Williams, J.L. Hart, M.L. Taheri, B. Urbanc, Elucidation of insulin assembly at acidic and neutral pH: Characterization of low molecular weight oligomers, Proteins : Structure, Function, and Bioinformatics, 85 (2017) 2096-2110.
[17] R. Pujol-Pina, S. Vilaprinyo-Pascual, R. Mazzucato, A. Arcella, M. Vilaseca, M. Orozco, N. Carulla, SDS-PAGE analysis of A beta oligomers is disserving research into Alzheimer's disease: appealing for ESI-IM-MS, Scientific Reports, 5 (2015) 13.
[18] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 3rd Edition ed., Springer, New York, 2006.
[19] M. Steiner-Browne, S. Elcoroaristizabal, Y. Casamayou-Boucau, A.G. Ryder, Investigating native state fluorescence emission of Immunoglobulin G using polarized Excitation Emission Matrix (pEEM) spectroscopy and PARAFAC, Chemometrics Intellig. Lab. Syst., 185 (2019) 1-11.
[20] Y. Casamayou-Boucau, A.G. Ryder, Accurate anisotropy recovery from fluorophore mixtures using Multivariate Curve Resolution (MCR), Anal. Chim. Acta, 1000 (2018) 132-143.
[21] R.D. JiJi, K.S. Booksh, Mitigation of Rayleigh and Raman spectral interferences in multiway calibration of excitation-emission matrix fluorescence spectra, Anal. Chem., 72 (2000) 718-725.
[22] Å. Rinnan, K.S. Booksh, R. Bro, First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes, Analytica Chimica Acta, 537 (2005) 349-358.
[23] A.L. de Faria e Silva, S. Elcoroaristizabal, A.G. Ryder, Multi-attribute quality screening of immunoglobulin G using polarized Excitation Emission Matrix spectroscopy, Anal. Chim. Acta, 1101 (2020) 99-110.
[24] R.R. Porter, Partition chromatography of insulin and other proteins, Biochemical Journal, 53 (1953) 320-328.
[25] Y. Casamayou-Boucau, A.G. Ryder, Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters, validation standards, and Rayleigh scatter removal methods, Methods and applications in fluorescence, 5 (2017) 037001.
[26] M. Ameloot, M. vandeVen, A.U. Acuna, B. Valeur, Fluorescence anisotropy measurements in solution: Methods and reference materials (IUPAC Technical Report), Pure Appl. Chem., 85 (2013) 589-608.
[27] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemometrics Intellig. Lab. Syst., 76 (2005) 101-110.
[28] J. Jaumot, R. Tauler, MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, Chemometrics Intellig. Lab. Syst., 103 (2010) 96-107.
[29] E.S. Oh, A. Woods, J.R. Couchman, Multimerization of the cytoplasmic domain of syndecan-4 is required for its ability to activate protein kinase C, The journal of Biological Chemistry, 272 (1997) 11805-11811.
[30] E. Cao, Y. Chen, Z. Cui, P.R. Foster, Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions, Biotechnolology and Bioengineering, 82 (2003) 684-690.
[31] L.F. Pease, 3rd, M. Sorci, S. Guha, D.H. Tsai, M.R. Zachariah, M.J. Tarlov, G. Belfort, Probing the nucleus model for oligomer formation during insulin amyloid fibrillogenesis, Biophysical journal, 99 (2010) 3979-3985.

[32] W. Kadima, L. Ogendal, R. Bauer, N. Kaarsholm, K. Brodersen, J.F. Hansen, P. Porting, The influence of ionic strength and pH on the aggregation properties of zinc-free insulin studied by static and dynamic laser light scattering, Biopolymers, 33 (1993) 1643-1657.
[33] L. Bromberg, J. Rashba-Step, T. Scott, Insulin particle formation in supersaturated aqueous solutions of poly(ethylene glycol), Biophysical journal, 89 (2005) 3424-3233.
[34] S.-H. Wang, X.-Y. Dong, Y. Sun, Effect of (−)-epigallocatechin-3-gallate on human insulin fibrillation/aggregation kinetics, Biochemical Engineering Journal, 63 (2012) 38-49.
[35] T. Sneideris, D. Darguzis, A. Botyriute, M. Grigaliunas, R. Winter, V. Smirnovas, pH-Driven Polymorphism of Insulin Amyloid-Like Fibrils, PloS one, 10 (2015) e0136602.
[36] S.M. Patil, D.A. Keire, K. Chen, Comparison of NMR and Dynamic Light Scattering for Measuring Diffusion Coefficients of Formulated Insulin: Implications for Particle Size Distribution Measurements in Drug Products, The AAPS journal, 19 (2017) 1760-1766.
[37] S. Hvidt, Insulin association in neutral solutions studied by light scattering, Biophysical chemistry, 39 (1991) 205-213.
[38] R.K. Chitta, D.L. Rempel, M.A. Grayson, E.E. Remsen, M.L. Gross, Application of SIMSTEX to oligomerization of insulin analogs and mutants, Journal of the Americal Society for Mass Spectrometry, 17 (2006) 1526-1534.
[39] C.F. Hjorth, M. Norrman, P.O. Wahlund, A.J. Benie, B.O. Petersen, C.M. Jessen, T.A. Pedersen, K. Vestergaard, D.B. Steensgaard, J.S. Pedersen, H. Naver, F. Hubalek, C. Poulsen, D. Otzen, Structure, Aggregation, and Activity of a Covalent Insulin Dimer Formed During Storage of Neutral Formulation of Human Insulin, Journal of Pharmaceutical Sciences, 105 (2016) 1376-1386.
[40] M. Danaei, M. Dehghankhold, S. Ataei, F. Hasanzadeh Davarani, R. Javanmard, A. Dokhani, S. Khorasani, M.R. Mozafari, Impact of Particle Size and Polydispersity Index on the Clinical Applications of Lipidic Nanocarrier Systems, Pharmaceutics, 10 (2018) 1-17.
[41] E.J. Nettleton, P. Tito, M. Sunde, M. Bouchard, C.M. Dobson, C.V. Robinson, Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry, Biophys J, 79 (2000) 1053-1065.
[42] F.W.J. Teale, The ultraviolet fluorescence of proteins in neutral solution, Biochemical Journal, 76 (1960) 381-388.
[43] M. Correia, M.T. Neves-Petersen, P.B. Jeppesen, S. Gregersen, S.B. Petersen, UV-light exposure of insulin: pharmaceutical implications upon covalent insulin dityrosine dimerization and disulphide bond photolysis, PloS one, 7 (2012) e50733.
[44] F.X. Schmid, Biological Macromolecules: UV-visible spectrophotometry, Encyclopedia of Life Sciences, Macmillan Publishers Ltd. , 2001.
[45] Y.K. Li, A. Kuliopulos, A.S. Mildvan, P. Talalay, Environments and mechanistic roles of the tyrosine residues of Δ5-3-ketosteroid isomerase, Biochemistry, 32 (1993) 1816-1824.
[46] J.A. Poveda, M. Prieto, J.A. Encinar, J.M. Gonzalez-Ros, C.R. Mateo, Intrinsic tyrosine fluorescence as a tool to study the interaction of the shaker B "ball" peptide with anionic membranes, Biochemistry, 42 (2003) 7124-7132.
[47] M. Noronha, J.C. Lima, P. Lamosa, H. Santos, C. Maycock, R. Ventura, A.L. Maçanita, Intramolecular Fluorescence Quenching of Tyrosine by the Peptide α-Carbonyl Group Revisited, The Journal of Physical Chemistry A, 108 (2004) 2155-2166.
[48] S.J. Martinez, J.C. Alfano, D.H. Levy, The electronic spectroscopy of the amino acids tyrosine and phenylalanine in a supersonic jet, Journal of Molecular Spectroscopy, 156 (1992) 421-430.
[49] L. Li, C. Li, Z. Zhang, E. Alexov, On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi, Journal of chemical theory and computation, 9 (2013) 2126-2136.

[50] E.N. Baker, T.L. Blundell, J.F. Cutfield, S.M. Cutfield, E.J. Dodson, G.G. Dodson, D.M.C. Hodgkin, R.E. Hubbard, N.W. Isaacs, C.D. Reynolds, K. Sakabe, N. Sakabe, N.M. Vijayan, The Structure of 2Zn Pig Insulin Crystals at 1.5 Å Resolution, Philosophical Transactions of the Royal Society B: Biological Sciences, 319 (1988) 369-456.
[51] J. Lee, R.T. Ross, Absorption and Fluorescence of Tyrosine Hydrogen-Bonded to Amide-like Ligands, The Journal of Physical Chemistry B, 102 (1998) 4612-4618.
[52] J. Feitelson, On the Mechanism of Fluorescence Quenching. Tyrosine and Similar Compounds, The Journal of Physical Chemistry, 68 (1964) 391-397.
[53] D.M. Rayner, D.T. Krajcarski, A.G. Szabo, Excited state acid–base equilibrium of tyrosine, Canadian Journal of Chemistry, 56 (1978) 1238-1245.
[54] S.E. Harding, Chapter 9. Hydrodynamic properties of proteins, in: T.E. Creighton (Ed.) Protein structure - a practical approach, IRL Press1997.
[55] A.H. Pekar, B.H. Frank, Conformation of proinsulin. Comparison of insulin and proinsulin self-association at neutral pH, Biochemistry, 11 (1972) 4013-4016.
[56] D. Rimmerman, D. Leshchev, D.J. Hsu, J. Hong, B. Abraham, I. Kosheleva, R. Henning, L.X. Chen, Insulin hexamer dissociation dynamics revealed by photoinduced T-jumps and time-resolved X-ray solution scattering, Photochemical & Photobiological Sciences, 17 (2018) 874-882.
[57] J. Goldman, F.H. Carpenter, Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives, Biochemistry, 13 (1974) 4566-4574.
[58] C.L. Maikawa, A.A.A. Smith, L. Zou, C.M. Meis, J.L. Mann, M.J. Webber, E.A. Appel, Stable Monomeric Insulin Formulations Enabled by Supramolecular PEGylation of Insulin Analogues, Advanced Therapeutics, 3 (2020) 1900094.
[59] D. Bicout, C. Brosseau, A.S. Martinez, J.M. Schmitt, Depolarization of multiply scattered waves by spherical diffusers: Influence of the size parameter, Physical Review E, 49 (1994) 1767-1770.
[60] F. Shen, B. Zhang, K. Guo, Z. Yin, Z. Guo, The Depolarization Performances of the Polarized Light in Different Scattering Media Systems, IEEE Photonics Journal, 10 (2018) 1-12.
[61] N. Nagel, M.A. Graewert, M. Gao, W. Heyse, C.M. Jeffries, D. Svergun, H. Berchtold, The quaternary structure of insulin glargine and glulisine under formulation conditions, Biophysical chemistry, 253 (2019) 106226.
[62] S. Bhattacharjee, DLS and zeta potential - What they are and what they are not?, J. Controlled Release, 235 (2016) 337-351.