Characterization of Lysozyme PEGylation products using polarized Excitation Emission Matrix (pEEM) spectroscopy

Published version at: **DOI:** 10.1002/bit.27483

Characterization of Lysozyme PEGylation products using polarized Excitation Emission Matrix (pEEM) spectroscopy.

Ana Luiza de Faria e Silva,[1] Saioa Elcoroaristizabal,[1] and Alan G. Ryder.[1]*

[1]Nanoscale BioPhotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Galway, H91 CF50, Ireland.

* To whom all correspondence should be addressed.

Tel: +353 91 49 2943; **Email:** alan.ryder@nuigalway.ie (A.G.R)

Grant number: 14/IA/2282, Advanced Analytics for Biological Therapeutic Manufacture, Science Foundation Ireland (SFI) with co-funding under the European Regional Development Fund.

Running Title: Lysozyme PEGylation analysis by polarized emission.

Citation: Characterization of Lysozyme PEGylation products using polarized Excitation-Emission Matrix (pEEM) spectroscopy. A.-L. de Faria e Silva, S. Elcoroaristizabal, and A.G. Ryder. Biotechnology and Bioengineering, in press, (**2020**) **DOI:** 10.1002/bit.27483

This document has most of the proof changes implemented, but the definitive source is the published version.

Abstract

The growing use of therapeutic proteins requires accurate analytical techniques for measuring biophysical and structural changes during manufacturing. This is particularly true for PEGylation of proteins, because characterization of PEGylation reactions and products can often be difficult because of the relatively small impact on protein structure, the lack of an accessible PEG chromophore, and the heterogeneous final product mixtures. Intrinsic fluorescence spectroscopy is one potential solution because of its relatively high sensitivity to small changes in protein structure and its suitability for online or at-line measurements.

Here we use PEGylation of lysozyme as a model system to determine the efficacy of polarized Excitation Emission Matrix (pEEM) spectroscopy as a rapid tool for characterizing the structural variability of the lysozyme starting materials and PEGylated products with PEG to protein ratio (PPR). Dynamic Light Scattering (DLS) showed that as PPR increased from 0 to 2.8, the hydrodynamic radius increased from ~2.2 to 4.8 nm. pEEM measurements provided several sources of information: Rayleigh scatter to identify size changes and aggregate/particle formation, and fluorescence emission to assessed chemical and structural change. PEGylation induced sufficient physicochemical changes in lysozyme which produced changes in the pEEM spectra largely due to variations in hydrophobic environment of tryptophan residues close to a PEG attachment site. These significant spectral changes when modelled using conventional multivariate analysis methods were able to easily discriminate the raw product solutions according to degree of PEGylation and were also able to predict PPR with reasonable accuracy (RMSEC ~10%, relative error of prediction (REP) <20%) considering the reference Size Exclusion Chromatography method error of ~7.2%. Variable selection of the pEEM data, suggest that equivalent predictions could be obtained with faster and simpler 2D spectra making the method a more viable online measurement method.

Keywords: Protein, Lysozyme, PEGylation, Fluorescence, Polarized.

Introduction

Some protein-based therapies have short serum half-lives caused by high kidney clearance rates and/or are unstable in the presence of certain endogenous agents (*i.e.* proteases) and these effects can affect therapeutic efficacy (Goodman, Gilman, Brunton, Lazo, & Parker, 2006). One solution is conjugation of polyethylene glycol (PEG) which can significantly improve stability and serum half-life (Kontermann & ebrary Inc., 2012; Pisal, Kosloski, & Balu-Iyer, 2010). PEG is a non-toxic, non-immunogenic, non-antigenic, highly water soluble, and FDA-approved polymer (Veronese & Pasut, 2005) and conjugation via amino group modification is a common method for attaching small molecules to proteins (Jevsevar, Kunstelj, & Porekar, 2010).

Protein conjugation often involves chemically stressful conditions such as the modification of specific amino acids (Luo et al., 2011) and physical stressors such as agitation. These can trigger changes in folding/unfolding routes, induce aggregation, and or fragmentation, leading to loss of activity and immunogenicity issues (Ratanji, Derrick, Dearman, & Kimber, 2014), and/or reduction in process yields (Ross & Wolfe, 2016; Sharma, 2007). This, combined with the increased structural complexity of conjugates, requires the use of more sensitive quality control (QC) tools to better identify and assess biophysical and structural changes during manufacturing. Lysozyme is an enzyme found in body fluids, that has an important antimicrobial function, and is widely used in the food and pharmaceutical industry (Wu et al., 2019). It is a relatively small (~14kDa, 129 amino acids) and stable enzyme, which is widely used in research as a model system for studying protein structure, stability, function and chemical modification. Here PEGylation of lysozyme is used as a model system for developing analytical methods based on intrinsic fluorescence emission.

Biopharmaceuticals are usually characterized using a variety of standard chromatographic or spectroscopic methods, however, PEG-protein conjugate analysis is more demanding because of the lack of a suitable chromophore in PEG for spectroscopic analysis, which limits the use of absorbance based methods (Cheng, Chuang, Chen, & Roffler, 2012). Furthermore the PEG ligand is commonly a mixture of different molecular weight polymers, and this coupled with the presence of multiple attachment sites in proteins, usually leads to the formation of a population of differently conjugated species after PEGylation (Povoski, Davis, Colcher, & Martin, 2013; Veronese & Pasut, 2005). A range of analytical methods can be used to assess the integrity, purity, size, the degree of PEGylation, total protein concentration, high molecular weight species (HMWS) content, and lot-to-lot variability of different proteins and/or their PEGylated forms. Size exclusion chromatography (SEC) (Pai et al., 2011), sodium dodecyl sulfate polyacrylamide gel electrophoresis (Y. Gokarn et al., 2015; Hinds & Kim, 2002), mass spectrometry (Y. R. Gokarn, McLean, & Laue, 2012; Thiagarajan, Semple, James, Cheung, & Shameem, 2016), circular dichroism (Freitas & Abrahao-Neto, 2010; Y. R. Gokarn et al., 2012; Pai et al., 2011), dynamic light scattering (Y. R. Gokarn et al., 2012), and Fourier-transform infrared spectroscopy (FT-IR) (Deygen & Kudryashova, 2016; Grosshans et al., 2018) are the most common techniques employed. However, no single technique can provide all the protein quality information required for intermediate or final product analysis. SEC is commonly used for protein analysis and quantification of PEGylation degree but has several limitations. First, protein contact with buffer components (e.g. salts/organic solvents) and column surface can induce changes in structure and/or increase secondary interactions generating misleading results (Hughes et al., 2009). Also, the long column equilibration times and run times (mainly when it is necessary to use bigger columns to increase resolution) preclude its use as an online, inexpensive monitoring tool for protein analysis. Furthermore, SEC may not always be able to measure weakly bound, soluble aggregates, or insoluble aggregates the presence of both which could influence conjugation reactions (Berkowitz, 2006; Liu, Andya, & Shire, 2006). *In-situ* PEGylation monitoring is particularly difficult because PEG lacks a suitable chromophore or fluorophore for diagnostic purposes, the inherent protein structural complexity, and the high protein concentrations involved.

Here we investigate the efficacy of polarized, intrinsic protein fluorescence (IPF) measurements for the characterization of lysozyme PEGylation products. IPF spectroscopy results in minimal structural perturbation compared to the use of extrinsic fluorophores (Ghisaidoobe & Chung, 2014; Ohadi, Legge, & Budman, 2015), very sensitive to structural changes (Raynal, Lenormand, Baron, Hoos, & England, 2014) and can be implemented in-line or on-line, which makes it a potential PAT tool. There are three fluorescent amino acids, phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp), with Trp being the dominant emitter (Lakowicz, 2006). Most proteins have multiple fluorophores, many in close proximity (within 10Å) which produces extensive Förster Resonance Energy Transfer (FRET) which affects all emission parameters. For lysozyme, there are six tryptophan (Trp) and three tyrosine (Tyr) fluorophores which contribute to IPF (Arai, Glabe, & Luecke, 2012; Girardi, Holdom, Davies, Sutton, & Beavil, 2009). FRET is also very sensitive to structural changes and provides potentially very diagnostic measurements for assessing protein structure and stability (Wiberg, Sterner-Molin, & Jacobsson, 2004). Furthermore, Trp emission is very sensitive to the local physicochemical environment and emission changes can be used to monitor conformational change, denaturation, and binding (Ghisaidoobe & Chung, 2014). To extract more information from *in-situ* measurements, one can use multidimensional measurements like Excitation Emission Matrix (EEM) and Anisotropy Resolved Multidimensional Emission Spectroscopy, ARMES, (Casamayou-Boucau & Ryder, 2017; Groza, 2016; Groza, Li, & Ryder, 2015). Here, we first investigate the use of pEEM spectroscopy to measure protein structural changes and other physicochemical changes caused by PEGylation. We then evaluate the efficacy of pEEM for reaction monitoring and predicting the degree of lysozyme PEGylation.

Materials and methods

Reagents and Materials: Lysozyme (LZ) from chicken egg white (L6876), 5kDa Methoxypolyethylene glycol p-nitrophenyl carbonate (mPEG-pNPC), L-lysine (K), Amicon Ultracentrifuge Filters and phosphate buffer (PB) reagents were purchased from Sigma-Aldrich and used as received. All solutions were prepared with high purity water (HPW) from Honeywell (Chromatography grade) and then membrane filtered (0.2 µm PES Captiva filters, Merck Millipore) prior to use. Lysozyme-PEG conjugates were prepared by mixing the protein with varying PEG:protein molar ratios (1, 2, 4, 6, 12, 24). mPEG-pNPC was the PEGylating agent selected because of the slower reaction rate compared to other agents which enabled us to easily vary the degree of PEGylation (Veronese & Pasut, 2005). The reaction takes place between the amino reactive groups in the protein (6 α-amino from lysine residues and 1 N-terminal NH_2) yielding a stable urethane linkage. To generate products with varying PEGylation profiles the PEG:protein mixtures were incubated in 1×1cm pathlength quartz cuvettes (Lightpath Optical, UK) with gentle stirring at 20 or 30°C in the fluorimeter for 2, 24, or 72 hours. All reactions used a single starting material source lots of lysozyme and PEGylation agent. Reactions were stopped by adding excess L-lysine and the final mixture was purified using Amicon Ultracentrifuge Filters (10kDa MW cut-off).

PEGylation parameters: The reaction conditions used here were expected to result in the attachment of between one and six PEG residues to each lysozyme molecule, with the PEGs being attached following the order of lysine reactivity: Lys33>Lys97>Lys116. This results in a heterogenous population of PEGylated species in the final reaction mixture, and thus the PEG-LZ PEGylation profile is a critical quality attribute.

We used two parameters from SEC measurements to define the degree of PEGylation: PEG-LZ$_{\%}$ and PPR. The total PEG-LZ$_{\%}$ (Equation1) parameter was defined as the sum of the % area under the curve (AUC) of the four resolved PEGylated species peaks. The individual species concentration was defined as a percentage of the area in the SEC data, e.g. %2PEG =(AUC_{peak3}/AUC_{total})×100). A PEG to protein ratio, PPR, (Equation 2) was also defined which should be a more useful univariate value as it took into account the actual PEG contribution of each species resolvable by SEC and thus provided an estimate of the total number of PEG molecules attached to the protein sample.

$$\boldsymbol{Total\ PEG\text{-}LZ_{\%}} = \left(\frac{(\mathbf{AUC_{Peak2} + AUC_{Peak3} + AUC_{Peak4} + AUC_{Peak5}})}{\mathbf{AUC_{Total}}}\right) \times \mathbf{100} \qquad \textbf{Equation 1}$$

$$\boldsymbol{PPR} = ((\%\ \mathbf{1\ PEG - LZ \times 1}) + (\%\ \mathbf{2\ PEG - LZ \times 2}) + (\%\ \mathbf{3\ PEG - LZ \times 3}) + (\%\ \mathbf{Multi\ PEG - LZ \times 4}))/\mathbf{100}$$

Equation 2

Data collection: All spectra were collected from non-degassed solutions held in quartz cuvettes. UV-visible spectra (200–600 nm, 2 nm spectral resolution) were measured using an Agilent Cary 60 spectrometer with phosphate buffer pH=8.0 as reference. pEEM spectra were collected at 20°C using a Cary Eclipse fluorimeter (Agilent Technologies) fitted with broadband wire grid polarizers (Casamayou-Boucau & Ryder, 2017) and a temperature-controlled multi-cell holder. Spectra were collected using an excitation/emission range of $\lambda_{ex/em}$

= 240–320/260–450 nm, with a step size of 2 nm in each case, 10 nm excitation/emission slit widths, and a scan rate of 1200 nm min^{-1}. Data were acquired using four polarization configurations: HH (horizontal-horizontal), HV (horizontal-vertical), VH (vertical-horizontal), and VV (vertical-vertical) and spectra were corrected for instrument response. EEM_{VH} data were G factor ($G=EEM_{HV}/EEM_{HH}$) corrected (Ameloot, vandeVen, Acuna, & Valeur, 2013), and these corrected spectra were hereafter designated the perpendicularly polarized ($EEM_{\perp}=EEM_{VH}\times G$), whereas the EEM_{VV} was the parallel polarized ($EEM_{\parallel}$). Spectra were corrected for instrumental response using a correction factor calculated using BAM (Federal Institute for Standards and Materials, Germany) calibration standards (Resch-Genger et al., 2005).

For SEC, DLS, and SDS-PAGE, analysis, aliquots of starting materials and purified reaction products stored at –70°C were defrosted overnight at 4°C before testing and each aliquot had the same number of freeze-thaw cycles. SEC was undertaken using an Agilent 1260 HPLC with a diode array detector. Samples were filtered using 0.20 µm PES filters (Agilent) prior to injection (10µL) at 30ºC onto a 300×7.8 mm mAb PAC-SEC1 column (5 µm particle size, ThermoFisher) using a 50mM Sodium Phosphate + 300 mM NaCl buffer mobile phase (pH 6.8) and a 0.76 mL/min flow rate. All samples were injected in triplicate. SDS-PAGE used the method proposed by Laemmli (Laemmli, 1970): 7.5 µL of sample was mixed with 2.5 µL of 4× Laemmli loading buffer (BioRad), heated at 90 °C for 5 min and then loaded into the precast 4–15% Mini-PROTEAN® TGX™ protein gels. 10 µL of pre-stained SDS-PAGE standards (Bio-Rad) were used as the molecular marker. The gels were run (using a Mini-PROTEAN Tetra Cell, BioRad) in a 1× solution of Tris/Glycine/SDS SDS running buffer at room temperature using a constant voltage (200V) for approximately 30 min (or until the dye front reached the end the gel). Gels were fixed by heating (microwave) for 50 seconds with a solution of ethanol/acetic acid and stained with BaI_2 (Kurfurst, 1992) and/or Comassie blue solution. DLS data were collected at 20 °C, after filtration (0.20µm PES filter), using a Malvern Zetasizer Nano ZS (173° detection angle). Each sample was measured 5 times (each measurement was an average of 10 runs of 10 s. duration).

Data analysis and Chemometrics: For pEEM multivariate data analysis, Raman scatter was minimized using blank subtraction (Ameloot et al., 2013) (Phosphate Buffer pH 8.0) and Rayleigh scattering (RS) correction was implemented by replacing the Rayleigh scatter area with missing values (Bahram, Bro, Stedmon, & Afkhami, 2006; Zepp, Sheldon, & Moran, 2004), which despite removing part of the emission spectrum, allowed for complete RS band removal minimizing artefact introduction. pEEM data also provides information related to the presence of particles (Zhang, Zhao, Li, & Tong, 2001) from the Rayleigh scatter signal. Here, the scatter signal was extracted from the $EEM_{\parallel}$ data as a separate bilinear component using PARAFAC modelling (Casamayou-Boucau & Ryder, 2017; Rinnan, Booksh, & Bro, 2005) and used for qualitative assessment of protein aggregation. This scatter component was used to calculate univariate scattering values, for either the whole scatter volume or at several discrete excitation and emission wavelengths.

Absorbance and pEEM spectra were smoothed using a Savitzky–Golay filter (with a second-order polynomial with a 15-point window size) and then normalized to maximum value (Savitzky & Golay, 1964). Robust Principal Component Analysis (ROBPCA) (Hubert, Rousseeuw, & Vanden Branden, 2005) and Partial Least Squares (PLS) regression (Supplemental Information, SI) were used to compare the ability of the different spectroscopic methods to deliver two objectives: first, to discriminate starting materials and products using ROBPCA and second, to determine if unfolded PLS (U-PLS) based modelling could be used to quantify the degree of PEGylation.

Results & Discussion:

The first change observed during the reaction was an increasing yellowing of the reaction mixture caused by the release of a p-nitro phenol (pNP) group during conjugation, which immediately caused fluorescence quenching *(vide infra)*. In order to remove the small molecule reagents, the reaction mixture was purified using Amicon centrifuge filters (10 kDa MW cut-off), and the retained sample was recovered using the original buffer (Phosphate pH=8.0) to an approximate concentration of 1.0 g/L (yield=75%). The main change (compared to the unquenched starting material) observed in the absorption spectra after purification was a 4% decrease in A_{280} (lysozyme = 0.99±0.01 and PEG-LZ = 0.95±0.02 g/L), which agreed with the observed 4% decrease in fluorescence intensity. This was consistent for all products and is probably caused by some protein loss during the purification and reconstitution steps. The purification strategy adopted here allowed for removal of the pNP quencher (Figure S1, SI) but did not remove any free PEG residues. However, a control experiment showed that free PEG did not cause any significant change in Pegylated product emission properties even at the highest concentration (24 molar excess) used for these reactions (Figure S2, SI).

3.1 Lysozyme and Pegylated products characterization by conventional techniques: SEC was used to evaluate the degree of PEGylation as all reactions produced mixtures of pegylated species. The reaction product mixture chromatograms (Figure 2A) showed five peaks, which we assigned as follows: peak1 (non-PEGylated lysozyme, Rt=14.6min), Peak 2 (1PEG–LZ, Rt=12.8min), Peak 3(2PEG–LZ, Rt=11.8min), Peak 4 (3PEG–LZ, Rt=11.1min) and Peak 5 (Multi-PEG–LZ, Rt=10.8min). The relative areas of peaks 3-5 (Table 1) increased with higher PEGylating agent concentration and longer reaction times. There was also a shift in the chromatograms of some products, which occurred after implementing a column recovery procedure. This probably changed the column packing, changing retention times, however, this did not seem to affect peak areas used for product characterisation. Here, we define the PEGylation profile as being the distribution of different pegylated species as obtained by SEC measurements. SDS-PAGE (Figure 2B, lanes 10-13) indicated that more than four species were formed for some of the reactions which used higher excesses of the PEG reagent (12 and 24 molar excess) but unfortunately the SEC conditions used did not allow for resolution of species with more than three attached PEG. There were some small measurement variations (between injections) and the average chromatograms of three injections were therefore used for calculating the AUC of each peak. DLS showed a small increase in hydrodynamic radius (Table 1) consistent with PEG attachment (Morgenstern, Baumann, Brunner, & Hubbuch, 2017; Pfister & Morbidelli, 2014). There were also changes in polydispersity related to very low levels of larger particles (Figure 2), but this is largely unreliable information due to the low levels involved.

The normalised absorbance spectra (Figure 2E) showed no shifts or spectral changes due to PEGylation and thus cannot be used to discriminate products from starting material. ROBPCA of the normalized absorbance spectra (Figure 2F) used two principal components (PC), explaining 47.54% (PC1) and 27.27% (PC2) of the total variance. PC1 clearly separated starting materials from R4-6 based on small differences in the 250–270 nm region and in an increased absorbance at 350 nm. PC2 had a significative contribution to the total variance but overall it was not possible to identify any trend related to the degree of PEGylation in the purified products. This means that absorbance-based measurements are not suitable for reaction monitoring nor for Pegylated products characterization. The aggregation index (UV-AI) did not provide any reliable information about HMWS formation, presumably because of very low absorbance at 350 nm (de Faria e Silva, Elcoroaristizabal, & Ryder, 2020). The primary use of absorption spectra was to quickly evaluate sample preparation reproducibility in terms of protein concentration. The variability of both the starting materials (RSD) and the purified products (RSD= <2%, and <3% respectively) was low, indicating that sample preparation and handling was well controlled.

3.2 Reaction Monitoring by fluorescence: For some reactions, $EEM_{||}$ spectra were collected hourly over a 24-hour period (Figure S4/S5, SI). The first significant issue noticed was the dramatic decrease (~6×) in fluorescence intensity due to quenching, first by the nitrophenol group of the PEGylating agent itself, and second by the para nitrophenol leaving group. Nitro containing compounds are well known quenchers, (Lakowicz, 2006), and one expects that the smaller leaving group will be a more effective quencher here. Despite the low and noisy signal caused by the strong fluorescence quenching after addition of the PEG regent it was possible to see an increasing blue-shift in emission peak and quenching during the reaction. Fluorescence emission is determined by the excitation wavelength, with λ_{ex} = 280 nm exciting both Tyr and Trp, whereas at λ_{ex}> 290 nm, mostly Trp is excited (Lakowicz, 2006). Among the six lysozyme tryptophan residues, Trp 62 and Trp108 are considered to be the major contributors to native state intrinsic fluorescence (Imoto, Forster, Rupley, & Tanaka, 1972; Takahashi et al., 1988). As most lysozyme emission originates from these more externally located Trp, the spectral shift (from 355 to ~338 nm) is mostly due to quenching of these Trp by the nitro containing leaving group, with potentially a small contribution due to an increasing hydrophobic character in the environment around these Trp residues as more PEG is attached. It is probable that the quenching effect is dominant because of large decrease in emission intensity observed (and by recovery of emission intensity for the purified products). Thus, any EEM based reaction monitoring will be mostly based on this secondary, quenching effect rather than a change directly attributable to structural change.

3.3 Characterization of PEGylation products by pEEM: Once the products had been purified, the fluorescence intensity recovered as the small molecule quenchers were removed, as seen in the pEEM measurements (I_{MAX}, Table 1). Normalising the data to remove the effect of intensity variation, the EEM difference spectra showed significant, but relatively small spectral changes. The I_{350}/I_{330} (λ_{ex} = 296 nm) ratio (Figure S3, SI) showed a small increase with increasing degree of PEGylation which could be explained as being due to a progressive drop in the intensity of the directly excited, solvent exposed Trp emission concomitant with an increase in the blue shifted Trp emission. This suggested an increased relative emission from Trp in more hydrophobic environments. We can discount quenching based effects here because all the nitro containing species have been removed in the purification, as confirmed by the control experiment with p-nitrophenol.

The structure of PEG residues attached to proteins in solution has been shown by SANS and NMR to be essentially separate from the protein core, and not wrapped around the protein (Cattani, Vogeley, & Crowley, 2015; Pai et al., 2011). If this is the case here it would suggest that there should be a small effect on Trp emission spectra. This is what was observed with maximum changes of ~4% being observed in the difference spectra. As more PEG residues are added we saw a progressive emission change, which was dependant on the proximity of the attachment site with Trp62 and 108. Analysis of the distances involved (calculated using UCSF Chimera v1.14) showed that, of the three potential attachment sites (K33>K97>K116), K97 and K116 were significantly closer (by ~9 and 3 nm on average) to these Trp compared to K33 (**Figure 1**C). This suggested that the biggest emission changes might arise after attachment of the second PEG at one of these sites. This correlated with the PLS regression models for the individual species (*vide infra*) which showed that the models for the 2 and 3 PEG-LZ had the best linear correlations (Figure S7, SI).

For the control reactions (R28-30, Figure 3A), the differences in the Trp and Tyr spectral regions indicated structural changes caused by solution stirring, sample handling, and the purification processes. In addition there was an overall increase in light scatter (i.e. RS volume extracted from $EEM_{\parallel}$) observed for most reactions, which indicated HMWS formation (Masuda, Higashitani, & Yoshida, 2007), probably of weakly bound species. Independent DLS measurements (at 0, 6 and 24h respectively) showed an increase in sample polydispersity (PdI=0.49 to 0.81) with stirring but no significant change in hydrodynamic radius (R_h=2.08 to 2.07nm). In terms of size changes due to PEGylation, the PPR ratio (Table 1) correlated linearly with the hydrodynamic radius values from DLS measurements (r^2= 0.95) but the correlation with Rayleigh scatter (RS) from the purified product was less good (r^2 = 0.72). The primary reason for this poorer correlation is the reproducibility associated with RS measurements, particularly due to shot noise. RS data should be averaged over several measurements to reduce measurement variance which here were >10% in some cases and was much more significant than for the emission component.

The major anomalous reactions (see also PCA analysis below) were for R1-6 which were significantly different compared to all other reactions (Figure 3B/C). This starting material variance can be explained by the fact that R1-3 were the first reactions carried out and probably the least controlled samples in terms of sample handling. The starting materials used for R4-6 had also been used for a control experiment with 24 hr stirring, and were thus stirred for much longer than the other reactions. So, despite the fact that a single lot of material was used for all reactions, sample handling and mechanical shear induced by stirring can cause significant structural changes in lysozyme which is manifested as spectral changes in emission. A separate unfolding experiment using guanidine indicated that the R1-R6 samples had some spectral similarities to the unfolded protein, (Figure S6, SI) which supports the hypothesis that the extended stirring/different sample handling of the R1-6 samples was the root cause. This further demonstrates the utility of fluorescence EEM measurements for quality assurance testing of protein starting material variance (de Faria e Silva et al., 2020).

3.4 Discriminating variance in starting materials and PEGylated products using pEEM: ROBPCA of normalized pEEM data easily discriminated starting materials from products, and also products with different degrees of PEGylation (Figure 4). Three PCs were required to account for 81% ($EEM_{\parallel}$), 75% ($EEM_{\perp}$), and 85% (EEM_T) of total variance. The various measurements produced somewhat different trends, which is due to the different levels of light scatter, noise, and fluorescence signal in each. The EEM_T data was the least noisy, whereas the $EEM_{\perp}$ data should have the purest emission but is noisy due to low measurement intensities. The $EEM_{\|}$ model seemed to separate better samples with low level of conjugation, whereas $EEM_{\perp}$ better discriminated samples with higher degree of PEGylation (R19–21), and this may be due to the changing nature of the size and photophysical changes taking place. Scores plots also showed that starting material (R1-6) used in the first reactions were significantly different to those used for all other reactions. Their products were also discriminated because of spectral changes, which presumably are linked to structure/composition differences. DLS measurements (Table 1) yielded the same size for all of the starting material samples which indicates that the size changes involved (if any) were too small to be discriminated by DLS. Thus the variability being observed in pEEM is probably related to more subtle changes like partial unfolding where there is a minimal change in overall size.

The loadings plots showed the areas of highest spectral variance in the pEEM data (Figure 4D/E). When all samples were considered (n=66), the main spectral feature (ROBPC1) responsible for the separation of starting materials and products was the change in the contribution between the Tyr/Trp area (λ_{ex} ~280 nm) and Trp only area (λ_{ex} >302 nm) caused by reaction induced structural changes leading to variations in IFE and FRET rates. ROBPC2 seemed to be more related to differences between the various products and represent Trp emission changes caused by an increasing hydrophobic character in the environment around Trp 62 and Trp108: a positive contribution from more exposed Trp (λ_{em}>346) to the lower PEGylated species and a negative contribution for the higher PEGylated species. When we consider only the purified products (thus not including the large starting material variation), the model (Figure 5) better discriminated the structural changes induced by PEGylation and ROBPC1 (Figure 5C/D), seemed to be related to the emission of a more solvent exposed Trp population (λ_{ex}~280). The $EEM_{\|}$ and $EEM_{\perp}$ (Figure 5) models had similar discrimination trends (inverted PC2 orientation in the scores plots) with the $EEM_{\perp}$ model showing slightly better discrimination. Comparison of the PC1 loadings showed significant spectral changes, with the $EEM_{\|}$ having weaker, blue shifted emission when excited at 300 nm, but more red shifted emission for λ_{ex} <300 nm compared to $EEM_{\perp}$ (Figure 5D). Overall, the ROBPCA study indicated that there was sufficient spectral change induced by PEGylation to consider developing a predictive model for PEGylation quantification.

3.5 Predicting degree of PEGylation using pEEM: We first had to evaluate the different combinations of measurements (UV-Vis and pEEM) and PEGylation degree parameters (Total PEG-$LZ_{\%}$, PPR, and individual PEG-LZ species concentration) to determine which combination was best for PEGylation status determination. Despite a reasonably good linear correlation between fluorescence and individual PEG-LZ content of some species (e.g. 2 and 3 PEG-LZ, Figure S7 and Table S1, SI), most models had high relative errors of prediction (>30%) which is unsurprising as the fluorescence signal is a combination of all species contained in the sample. The significantly better correlations for the 2PEG species content, does tend to validate our hypothesis for explaining fluorescence emission changes.

There was a poor correlation between absorbance spectra and any of the parameters evaluated here (as shown by, model performance values in Figures S8 and Table S2,SI), and thus absorbance spectroscopy is not a good candidate for PEGylation analysis as anticipated. The pEEM based models were much better, with REP values that were ~30% lower, for the prediction of both PEG-$LZ_{\%}$ and PPR (Table S2, SI) and several pEEM models were further optimised. Optimization involved comparing models using non pre-processed data (NoPrep), normalized to the maximum (Norm), and mean centered (MC). The results (Table S2, SI) indicated that both $EEM_{\parallel}$ and $EEM_{\perp}$ data yielded statistically similar prediction results, and that pre-processing did not have much of an impact on model accuracy. The error in the SEC based reference method was estimated at ~7.2% for PPR and 4.4% for total-PEG (SI) and the PPR error is comparable to the RMSEC values (~10%) obtained for pEEM based models and the higher %REP errors. It should be noted that despite the higher error obtained with pEEM (REP $EEM_{\parallel}$=17%), the EJCRB (Figure S8, SI) shows that all the models were accurate. Further, improvements in prediction are not feasible using chemometric approaches on this data, as this will lead to over fitting and non-robust solutions. The key to improving predictive model capability, lies in the need to improve signal to noise ratio (SNR), and potentially the use of a better parameter for assessing PEGylation.

As high SNR pEEM measurements might not always be feasible for reaction monitoring, we next evaluated how to reduce the data acquisition time by investigating the use of variable selection (Mehmood, Liland, Snipen, & Saebo, 2012) to remove any non-essential information from the pEEM, and see if it was also possible to substitute simpler 2D measurements for pEEM. iPLS was better than Variable Importance in Projection (VIP) for variable selection as it both facilitated a substantial decrease in the data collection time (from 410s to 30s) and selected variables that are compatible with simpler 2D measurements (Table 2) without compromising model accuracy. Elliptical Joint Confidence Region (EJCR) testing of the different models and data for the degree of PEGylation (Figure S8C/D, SI) showed that for all the models built using variable selection, the ideal point (0,1) lay inside the ellipse. In assessing model quality, the $EEM_{\perp}$ (iPLS) model was best for PEG-LZ% whereas for the PPR parameter the iPLS models were significantly worse.

Overall, the PPR value was slightly better correlated with the emission properties, but more poorly predicted (REP and LOD) than PEG-$LZ_{\%}$, which was probably a consequence of the way the parameter was calculated. However, the differences are all within the estimated error in REP, which is optimistically 1-2%. For predicting PPR, $EEM_{\parallel}$ measurements with iPLS seemed to be the best option based on statistical parameters, but with this data, the differences were not statistically significant ($p>0.05$ based on a randomization test). In terms of data collection time, $EEM_{\parallel}$ may offer an advantage compared to $EEM_{\perp}$ as it can be implemented as a single measurement. In terms of the selected variables (Figure 6), VIP selected variables over the full emission range thus necessitating EEM measurements which reduces the potential for online use. However, iPLS selected variables that are compatible with faster and simple 2D measurements, which are more suitable for online measurements. Thus, one could collect a selection of 3 or 6 2D spectra (λ_{ex}=244, 288, 300nm and 244, 248, 258, 272, 284, 306nm for PEG-$LZ_{\%}$ and PPR respectively) instead of a full EEM and obtain similar results.

Conclusions

We have demonstrated for the first time the feasibility of using pEEM measurements with standard chemometric data analysis methods for assessing physicochemical variances in lysozyme with different degrees of PEGylation. pEEM because of its better sensitivity to changes in protein emission provides an alternative route for characterizing starting material and PEGylated product mixture variance which is faster than chromatography and more informative than simpler spectroscopic techniques like absorbance spectroscopy. The fluorescence emission component of the signal detects the physicochemical changes induced by multi PEG residue attachment. ROBPCA of pEEM data indicated that changes were due mostly to Trp/Tyr emission fluctuations caused by changes in Trp environment and/or varying FRET rates induced by structural changes with PEGylation, suggesting thus that one could build predictive models for PEGylation degree using pEEM.

The Rayleigh scatter data (Table 1) from $EEM_{\parallel}$ measurements provides size change information for the product solutions and this moderately correlated (r^2=0.72) with the PPR and hydrodynamic radius from DLS (r^2=0.65). Unfortunately, single measurement RS data were not sufficiently accurate because of measurement variation. The data needs to be made more reproducible by both averaging multiple scans per sample and implementing an explicit intensity calibration step. In contrast, the fluorescence data used here for analysis is more reproducible because normalised data was used to eliminate the lamp intensity fluctuation effects. Thus, pEEM measurements by being both sensitive to protein structure and size changes offers a unique and potentially rapid, non-destructive method for product characterisation and quantifying the degree of PEGylation.

For PAT use, the poorer SNR of pEEM measurements compared to unpolarized EEM is the primary limitation, particularly here with the induced fluorescence quenching from the nitro leaving group. However, this could be addressed by using a slower scan speed and multiple acquisitions, but with a concomitant trade off in terms of acquisition time. It would be better to move from a scanning-based system to the use of CCD based spectrometers (Quatela et al., 2018), which would reduce measurement time significant and thus facilitate high SNR data acquisition. Furthermore, we have shown by the use of variable selection that the full EEM space may not be required and by reducing the measurement to those excitation wavelengths containing the essential spectral information, a substantial decrease in the acquisition time is possible. This would facilitate spectral averaging thus improving emission SNR and the reproducibility of the RS data. This would enable the use of pEEM measurements as a viable PAT solution for in- or on-line process monitoring of protein or peptide conjugation reactions like PEGylation or Antibody Drug Conjugate manufacture.

Supplemental information available

Supporting information is available providing additional data and details on the chemometric analysis.

Acknowledgments

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grand number (14/IA/2282, *Advanced Analytics for Biological Therapeutic Manufacture*, to AGR). We also thank Agilent Technologies (Mulgrave Victoria, Australia) for the loan of a fluorescence spectrometer. The authors declare that there are no conflicts of interest.

References

Ameloot, M., vandeVen, M., Acuna, A. U., & Valeur, B. (2013). Fluorescence anisotropy measurements in solution: Methods and reference materials (IUPAC Technical Report). *Pure and Applied Chemistry, 85*(3), 589-608. doi:10.1351/pac-rep-11-11-12

Arai, H., Glabe, C., & Luecke, H. (2012). Crystal structure of a conformation-dependent rabbit IgG Fab specific for amyloid prefibrillar oligomers. *Biochimica Et Biophysica Acta-General Subjects, 1820*(12), 1908-1914. doi:10.1016/j.bbagen.2012.08.016

Bahram, M., Bro, R., Stedmon, C., & Afkhami, A. (2006). Handling of Rayleigh and Raman scatter for PARAFAC modeling of fluorescence data using interpolation. *Journal of Chemometrics, 20*(3-4), 99-105. doi:10.1002/cem.978

Berkowitz, S. A. (2006). Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. *Aaps Journal, 8*(3), E590-E605. doi:DOI 10.1208/aapsj080368

Casamayou-Boucau, Y., & Ryder, A. G. (2017). Extended wavelength anisotropy resolved multidimensional emission spectroscopy (ARMES) measurements: better filters, validation standards, and Rayleigh scatter removal methods. *Methods Appl Fluoresc, 5*(3), 037001. doi:10.1088/2050-6120/aa7763

Cattani, G., Vogeley, L., & Crowley, P. B. (2015). Structure of a PEGylated protein reveals a highly porous double-helical assembly. *Nature Chemistry, 7*(10), 823-828. doi:10.1038/nchem.2342

Cheng, T. L., Chuang, K. H., Chen, B. M., & Roffler, S. R. (2012). Analytical Measurement of PEGylated Molecules. *Bioconjugate Chemistry, 23*(5), 881-899. doi:10.1021/bc200478w

de Faria e Silva, A. L., Elcoroaristizabal, S., & Ryder, A. G. (2020). Multi-attribute quality screening of immunoglobulin G using polarized Excitation Emission Matrix spectroscopy. *Analytica Chimica Acta, 1101*, 99-110. doi:10.1016/j.aca.2019.12.020

Deygen, I. M., & Kudryashova, E. V. (2016). New versatile approach for analysis of PEG content in conjugates and complexes with biomacromolecules based on FTIR spectroscopy. *Colloids and Surfaces B-Biointerfaces, 141*, 36-43. doi:10.1016/j.colsurfb.2016.01.030

Freitas, D. D., & Abrahao-Neto, J. (2010). Biochemical and biophysical characterization of lysozyme modified by PEGylation. *International Journal of Pharmaceutics, 392*(1-2), 111-117. doi:10.1016/j.ijpharm.2010.03.036

Ghisaidoobe, A. B. T., & Chung, S. J. (2014). Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Forster Resonance Energy Transfer Techniques. *International Journal of Molecular Sciences, 15*(12), 22518-22538. doi:10.3390/ijms151222518

Girardi, E., Holdom, M. D., Davies, A. M., Sutton, B. J., & Beavil, A. J. (2009). The crystal structure of rabbit IgG-Fc. *Biochemical Journal, 417*, 77-83. doi:10.1042/Bj20081355

Gokarn, Y., Agarwal, S., Arthur, K., Bepperling, A., Day, E. S., Filoti, D., . . . Zarraga, I. E. (2015). Biophysical Techniques for Characterizing the Higher Order Structure and Interactions of Monoclonal Antibodies. *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization, Vol 2: Biopharmaceutical Characterization: The Nistmab Case Study, 1201*, 285-327. doi:BOOK_DOI 10.1021/bk-2015-1201

Gokarn, Y. R., McLean, M., & Laue, T. M. (2012). Effect of PEGylation on Protein Hydrodynamics. *Molecular Pharmaceutics, 9*(4), 762-773. doi:10.1021/mp200470c

Goodman, L. S., Gilman, A., Brunton, L. L., Lazo, J. S., & Parker, K. L. (2006). *Goodman & Gilman's the pharmacological basis of therapeutics* (11th ed.). New York: McGraw-Hill.

Grosshans, S., Rudt, M., Sanden, A., Brestrich, N., Morgenstern, J., Heissler, S., & Hubbuch, J. (2018). In-line Fourier-transform infrared spectroscopy as a versatile process analytical technology for preparative protein chromatography. *Journal of Chromatography A, 1547*, 37-44. doi:10.1016/j.chroma.2018.03.005

Groza, R. C. (2016). *Anisotropy Resolved Multi‑dimensional Emission Spectroscopy (ARMES): A new tool for the quantitative and structural analysis of proteins. PhD thesis.* (Ph.D.), National University of Ireland Galway, Galway, Ireland.
Groza, R. C., Li, B., & Ryder, A. G. (2015). Anisotropy resolved multidimensional emission spectroscopy (ARMES): A new tool for protein analysis. *Analytica Chimica Acta, 886*, 133-142. doi:10.1016/j.aca.2015.06.011
Hinds, K. D., & Kim, S. W. (2002). Effects of PEG conjugation on insulin properties. *Advanced Drug Delivery Reviews, 54*(4), 505-530. doi:10.1016/S0169-409x(02)00025-X
Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics, 47*(1), 64-79. doi:10.1098/004017004000000563
Hughes, H., Morgan, C., Brunyak, E., Barranco, K., Cohen, E., Edmunds, T., & Lee, K. (2009). A Multi-Tiered Analytical Approach For the Analysis and Quantitation of High-Molecular-Weight Aggregates in a Recombinant Therapeutic Glycoprotein. *Aaps Journal, 11*(2), 335-341. doi:10.1208/s12248-009-9108-1
Imoto, T., Forster, L. S., Rupley, J. A., & Tanaka, F. (1972). Fluorescence of lysozyme: emissions from tryptophan residues 62 and 108 and energy migration. *Proc Natl Acad Sci U S A, 69*(5), 1151-1155. doi:10.1073/pnas.69.5.1151
Jevsevar, S., Kunstelj, M., & Porekar, V. G. (2010). PEGylation of therapeutic proteins. *Biotechnology Journal, 5*(1), 113-128. doi:10.1002/biot.200900218
Kontermann, R., & ebrary Inc. (2012). *Therapeutic proteins strategies to modulate their plasma half-lives*. In (pp. xviii, 354 p.).
Kurfurst, M. M. (1992). Detection and Molecular-Weight Determination of Polyethylene Glycol-Modified Hirudin by Staining after Sodium Dodecyl-Sulfate Polyacrylamide-Gel Electrophoresis. *Analytical Biochemistry, 200*(2), 244-248. doi:Doi 10.1016/0003-2697(92)90460-O
Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680-685. doi:10.1038/227680a0
Lakowicz, J. R. (2006). *Principles of fluorescence spectroscopy* (3rd ed.). New York: Springer.
Liu, J., Andya, J. D., & Shire, S. J. (2006). A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. *Aaps Journal, 8*(3), E580-E589. doi:DOI 10.1208/aapsj080367
Luo, Q. Z., Joubert, M. K., Stevenson, R., Ketchem, R. R., Narhi, L. O., & Wypych, J. (2011). Chemical Modifications in Therapeutic Protein Aggregates Generated under Different Stress Conditions. *Journal of Biological Chemistry, 286*(28), 25134-25144. doi:10.1074/jbc.M110.160440
Masuda, H., Higashitani, K., & Yoshida, H. (2007). *Powder technology : fundamentals of particles, powder beds, and particle generation*. Boca Raton: CRC Press/Taylor & Francis.
Mehmood, T., Liland, K. H., Snipen, L., & Saebo, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62-69. doi:10.1016/j.chemolab.2012.07.010
Morgenstern, J., Baumann, P., Brunner, C., & Hubbuch, J. (2017). Effect of PEG molecular weight and PEGylation degree on the physical stability of PEGylated lysozyme. *International Journal of Pharmaceutics, 519*(1-2), 408-417. doi:10.1016/j.ijpharm.2017.01.040
Ohadi, K., Legge, R. L., & Budman, H. M. (2015). Intrinsic fluorescence-based at situ soft sensor for monitoring monoclonal antibody aggregation. *Biotechnology Progress, 31*(5), 1423-1432. doi:10.1002/btpr.2140
Pai, S. S., Hammouda, B., Hong, K. L., Pozzo, D. C., Przybycien, T. M., & Tilton, R. D. (2011). The Conformation of the Poly(ethylene glycol) Chain in Mono-PEGylated Lysozyme and Mono-PEGylated Human Growth Hormone. *Bioconjugate Chemistry, 22*(11), 2317-2323. doi:10.1021/bc2003583
Pfister, D., & Morbidelli, M. (2014). Process for protein PEGylation. *Journal of Controlled Release, 180*, 134-149. doi:10.1016/j.jconrel.2014.02.002

Pisal, D. S., Kosloski, M. P., & Balu-Iyer, S. V. (2010). Delivery of Therapeutic Proteins. *Journal of Pharmaceutical Sciences, 99*(6), 2557-2575. doi:10.1002/jps.22054

Povoski, S. P., Davis, P. D., Colcher, D., & Martin, E. W. (2013). Single molecular weight discrete PEG compounds: emerging roles in molecular diagnostics, imaging and therapeutics. *Expert Review of Molecular Diagnostics, 13*(4), 315-319. doi:10.1586/Erm.13.19

Quatela, A., Gilmore, A. M., Gall, K. E. S., Sandros, M., Csatorday, K., Siemiarczuk, A., . . . Camenen, L. (2018). A-TEEM (TM), a new molecular fingerprinting technique: simultaneous absorbance-transmission and fluorescence excitation-emission matrix method. *Methods and Applications in Fluorescence, 6*(2), 7. doi:10.1088/2050-6120/aaa818

Ratanji, K. D., Derrick, J. P., Dearman, R. J., & Kimber, I. (2014). Immunogenicity of therapeutic proteins: Influence of aggregation. *Journal of Immunotoxicology, 11*(2), 99-109. doi:10.3109/1547691x.2013.821564

Raynal, B., Lenormand, P., Baron, B., Hoos, S., & England, P. (2014). Quality assessment and optimization of purified protein samples: why and how? *Microbial Cell Factories, 13*, 10. doi:10.1186/s12934-014-0180-6

Resch-Genger, U., Pfeifer, D., Monte, C., Pilz, W., Hoffmann, A., Spieles, M., . . . Nording, P. (2005). Traceability in fluorometry: Part II. Spectral fluorescence standards. *Journal of Fluorescence, 15*(3), 315-336. doi:10.1007/s10895-005-2629-9

Rinnan, A., Booksh, K. S., & Bro, R. (2005). First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. *Analytica Chimica Acta, 537*(1-2), 349-358. doi:10.1016/j.aca.2005.01.044

Ross, P. L., & Wolfe, J. L. (2016). Physical and Chemical Stability of Antibody Drug Conjugates: Current Status. *Journal of Pharmaceutical Sciences, 105*(2), 391-397. doi:10.1016/j.xphs.2015.11.037

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry, 36*(8), 1627-1639. doi:10.1021/ac60214a047

Sharma, B. (2007). Immunogenicity of therapeutic proteins. Part 1: Impact of product handling. *Biotechnology Advances, 25*(3), 310-317. doi:10.1016/j.biotechadv.2007.01.005

Takahashi, I., Takahashi, K., Ichimura, M., Morimoto, M., Asano, K., Kawamoto, I., . . . Nakano, H. (1988). Duocarmycin-a, a New Antitumor Antibiotic from Streptomyces. *Journal of Antibiotics, 41*(12), 1915-1917.

Thiagarajan, G., Semple, A., James, J. K., Cheung, J. K., & Shameem, M. (2016). A comparison of biophysical characterization techniques in predicting monoclonal antibody stability. *MAbs, 8*(6), 1088-1097. doi:10.1080/19420862.2016.1189048

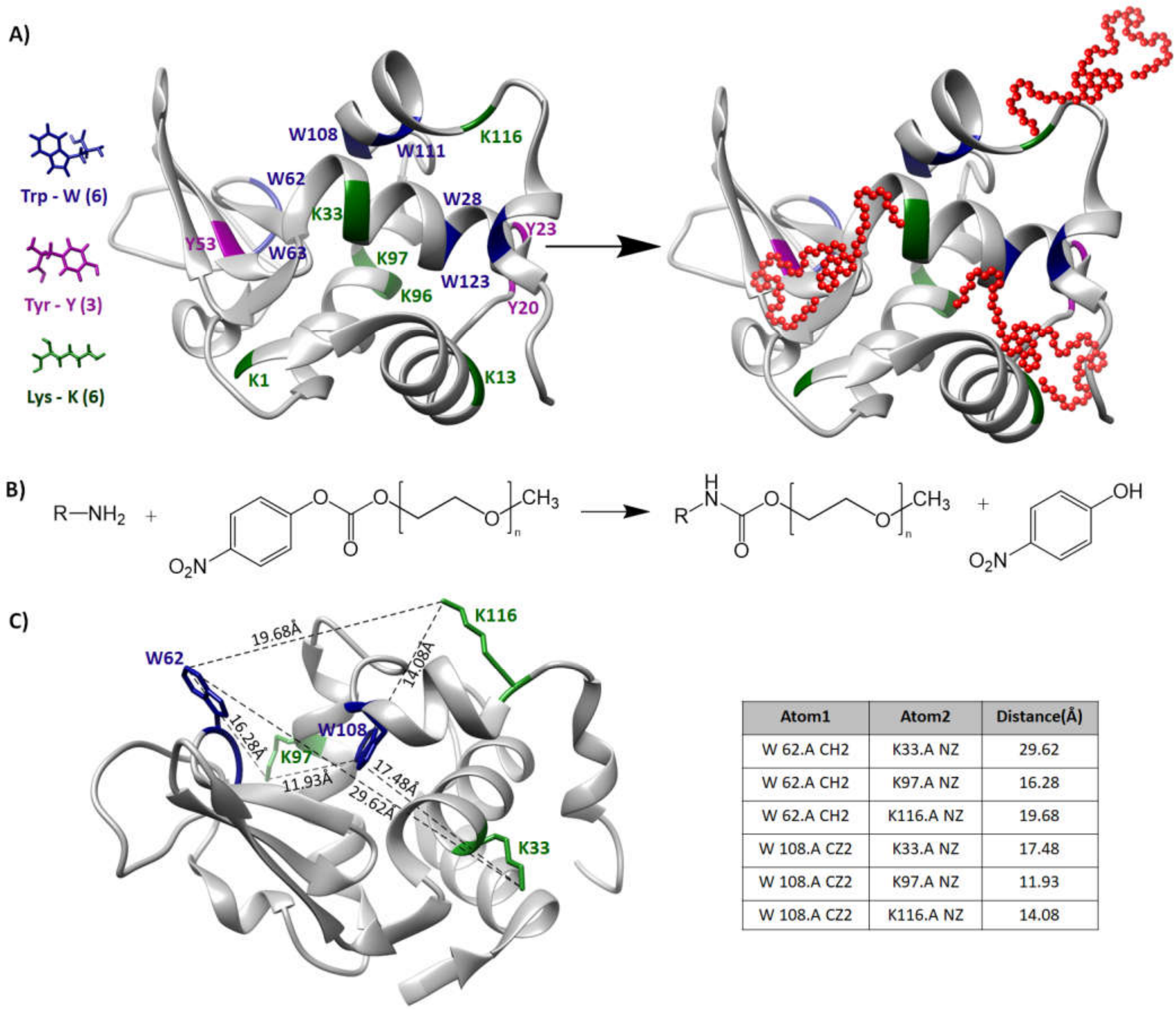
Veronese, F. M., & Pasut, G. (2005). PEGylation, successful approach to drug delivery. *Drug Discovery Today, 10*(21), 1451-1458. doi:Doi 10.1016/S1359-6446(05)03575-0

Wiberg, K., Sterner-Molin, A., & Jacobsson, S. P. (2004). Simultaneous determination of albumin and immunoglobulin G with fluorescence spectroscopy and multivariate calibration. *Talanta, 62*(3), 567-574. doi:10.1016/j.talanta.2003.08.024

Wu, T. T., Jiang, Q. Q., Wu, D., Hu, Y. Q., Chen, S. G., Ding, T., . . . Chen, J. C. (2019). What is new in lysozyme research and its application in food industry? A review. *Food Chemistry, 274*, 698-709. doi:10.1016/j.foodchem.2018.09.017

Zepp, R. G., Sheldon, W. M., & Moran, M. A. (2004). Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices. *Marine Chemistry, 89*(1-4), 15-36. doi:10.1016/j.marchem.2004.02.006

Zhang, S. Z., Zhao, F. L., Li, K. A., & Tong, S. Y. (2001). Determination of glycogen by Rayleigh light scattering. *Analytica Chimica Acta, 431*(1), 133-139. doi:10.1016/s0003-2670(00)01241-1



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

Figure 1: ***(A)*** Crystal structure of hen egg-white lysozyme (PDB entry: 1DPX Chimera software) showing location of Trp and Tyr fluorophores) and the lysine sites for conjugation. ***(B)*** Reaction scheme for the conjugation of p-nitrophenyl carbonate PEG to lysine. ***(C)*** Distance between residues Trp 62 and 108 and Lys 33,97,116 calculated using structure analysis from Chimera.

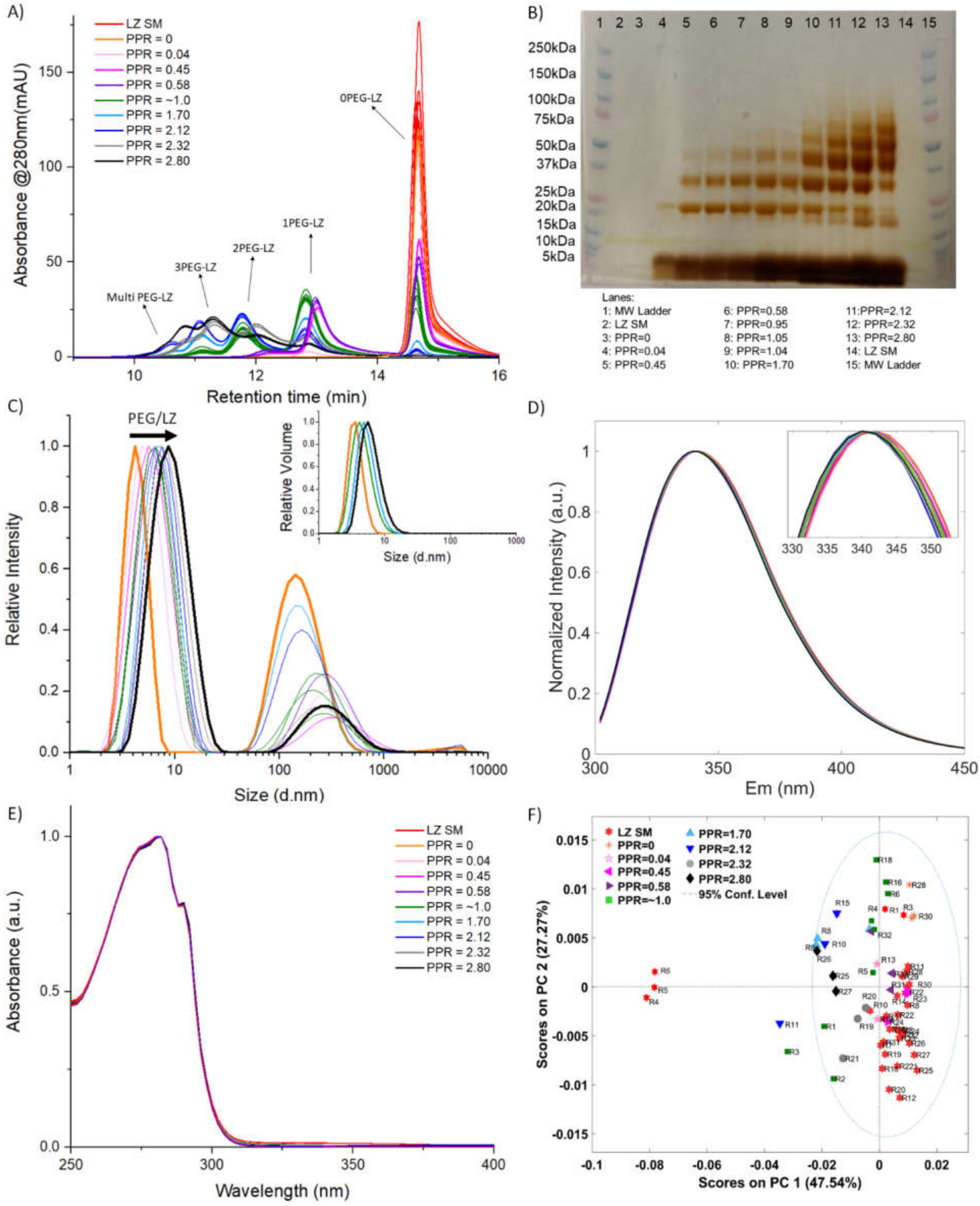


Figure 2: Analysis of LZ starting material and purified reaction products: (A) Unprocessed SEC chromatograms; (B) SDS-PAGE with iodine staining; (C) Size distribution by intensity and volume from DLS measurements; (D) 2D Fluorescence spectra using 280 nm excitation; (E). Normalized absorbance spectra of

all starting materials and purified PEG-LZ products; (F) ROBPCA scores plot from the UV data showing a lack of sensitivity for discriminating PEGylated products.

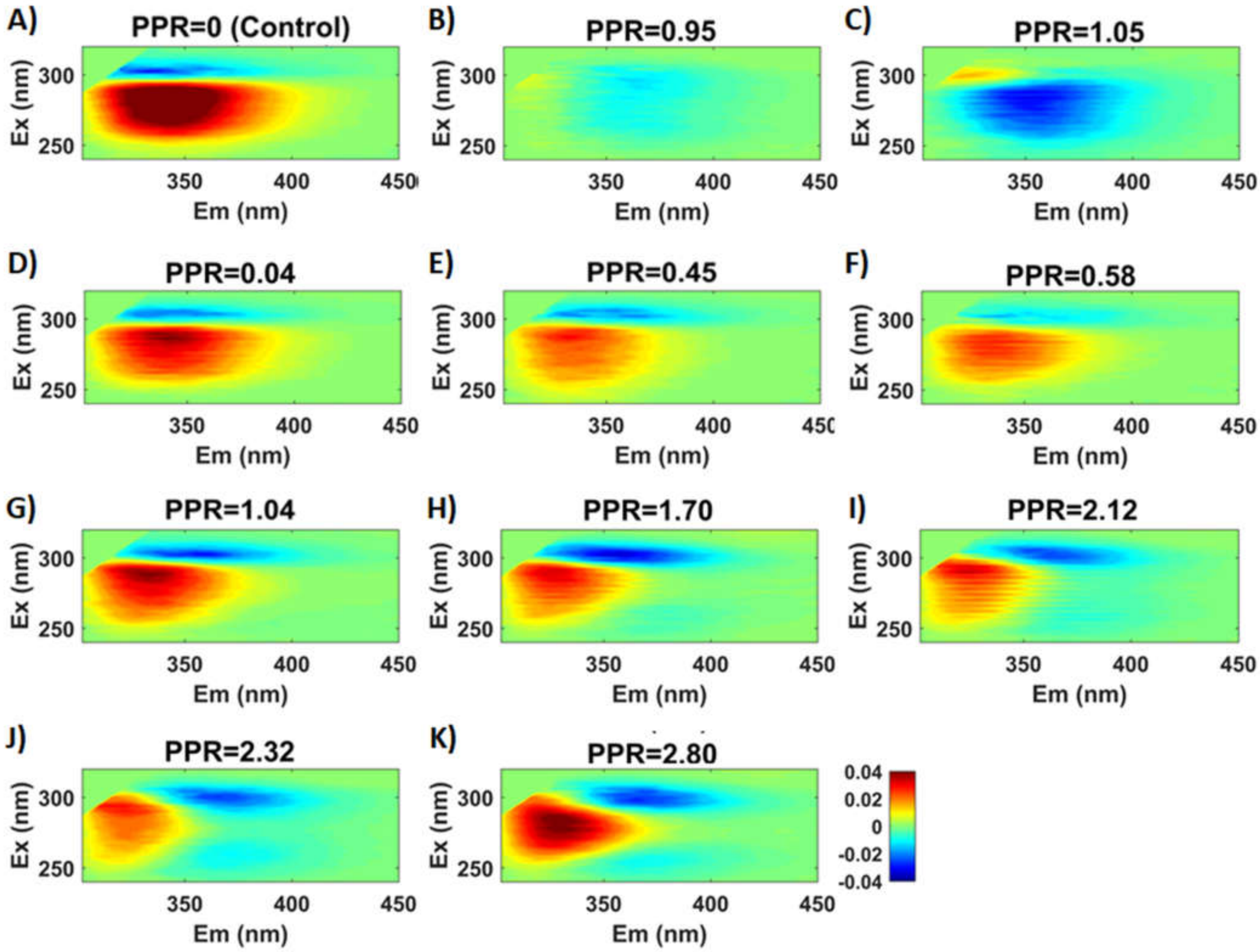


Figure 3: Difference spectra showing the emission changes between the purified product (PROD) and the lysozyme starting material (SM) with varying (0.04-2.80) PEG to protein ratio (PPR). Difference spectra were calculated using normalized EEM$_{\|}$ data as: $EEM_{Dif.} = EEM_{Mean\ PROD} - EEM_{Mean\ SM}$. The PEG to protein ratio was calculated as described in Equation 2.

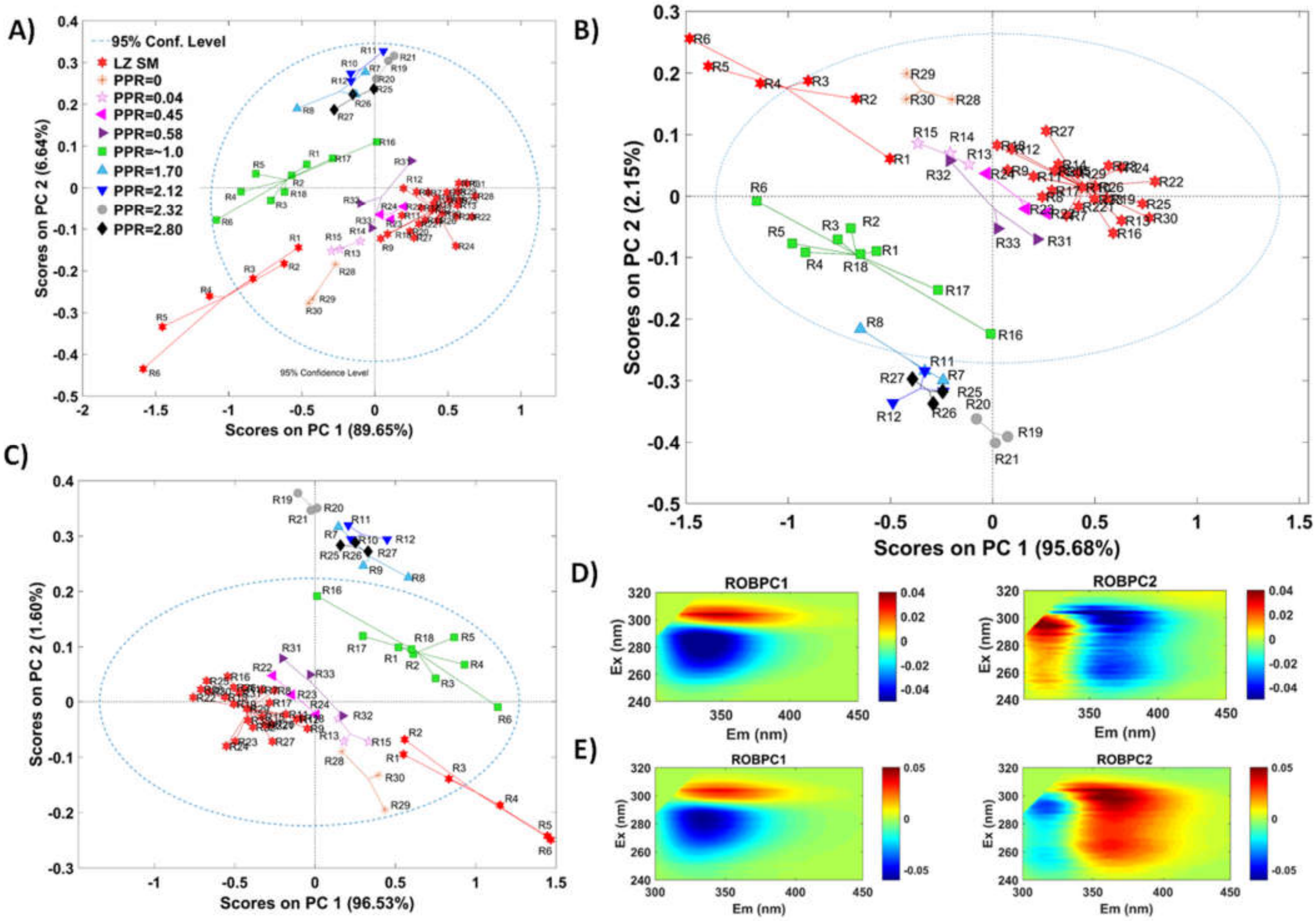


Figure 4: ROBPCA scores plots obtained using normalized: (A) EEM‖; (B) EEM⊥; and (C) EEM_T data. Refolded ROBPC1 and 2 loadings from the (D) EEM⊥ and (E) EEM‖ normalized ROBPCA models.

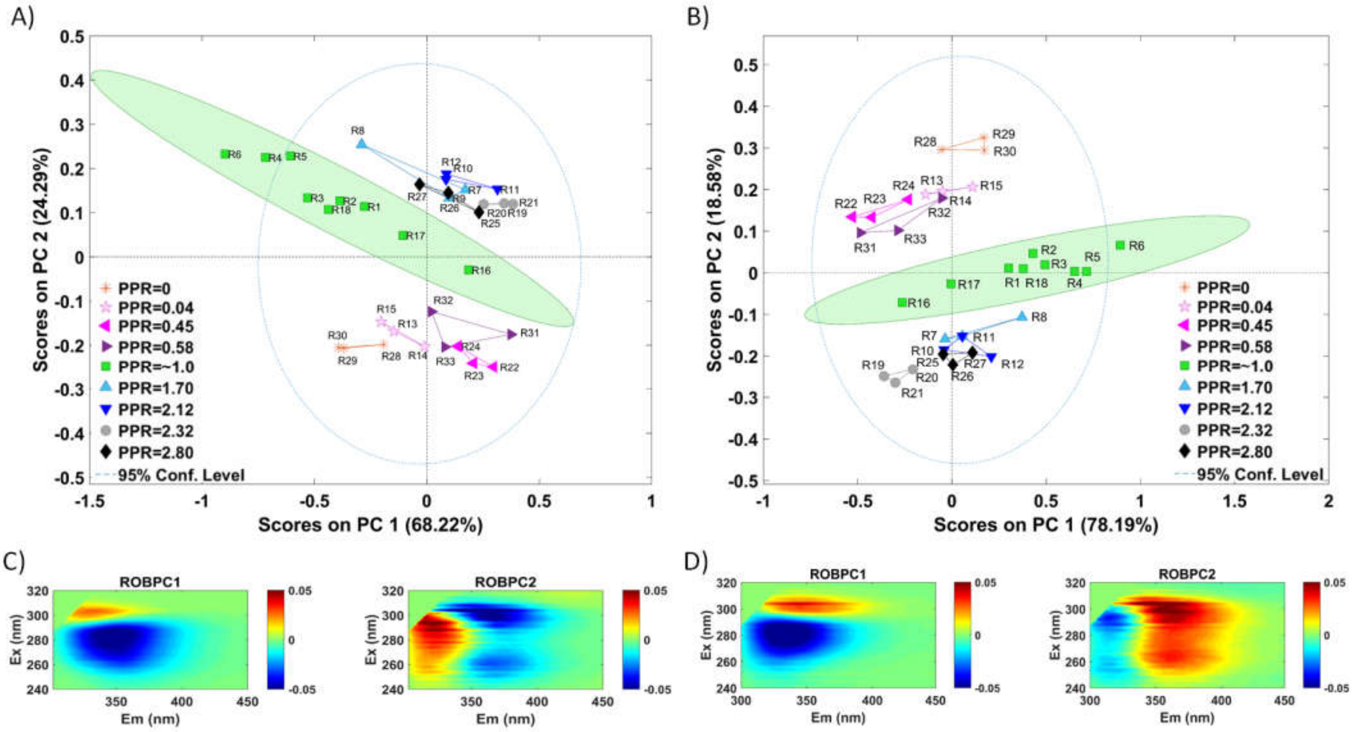


Figure 5: ROBPCA scores and loadings plots of the purified products from all reactions (n=33) using normalized (A/C) $EEM_{\parallel}$ and (B/D) $EEM_{\perp}$ spectra.

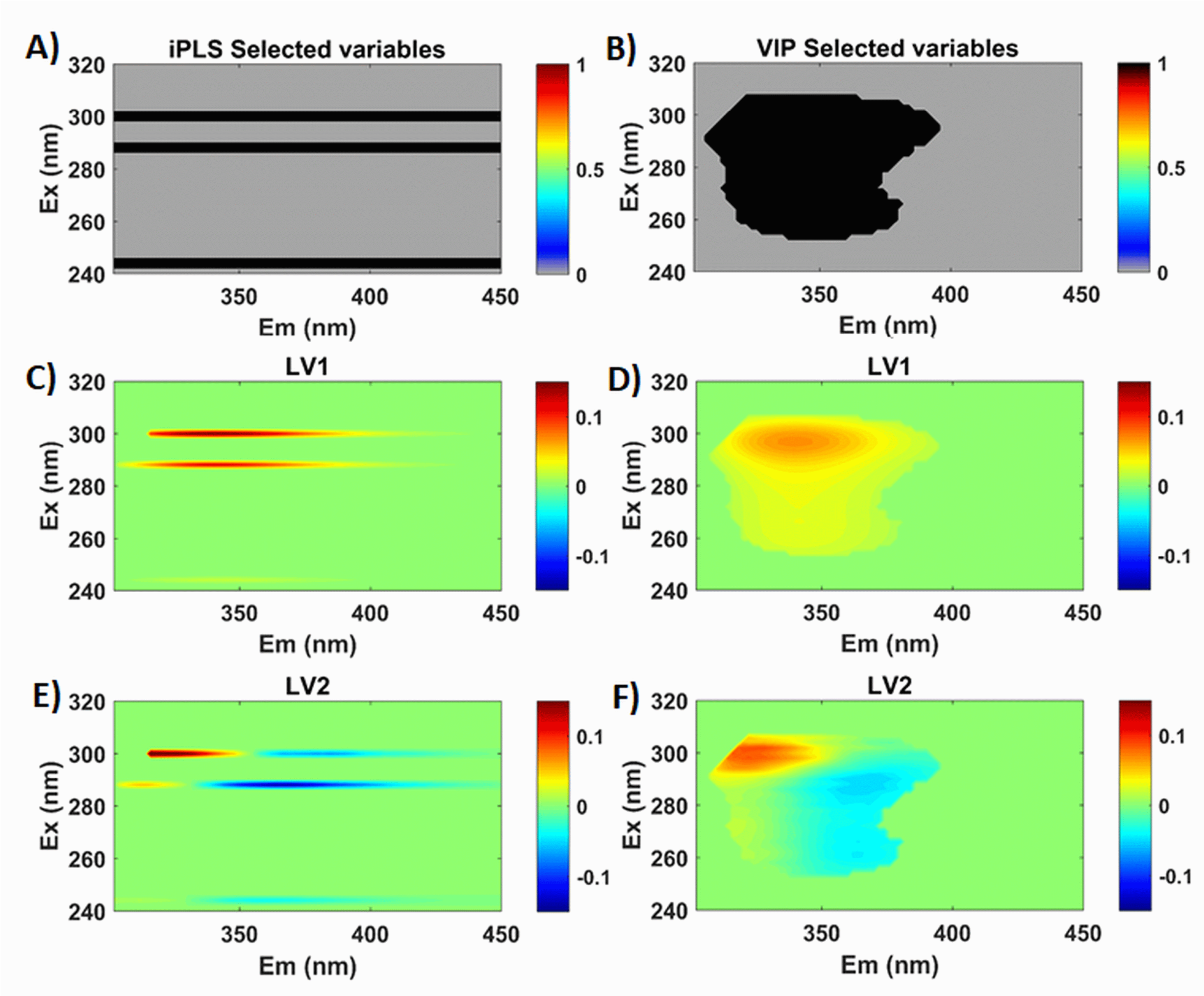


Figure 6: (**A)** iPLS and (B) VIP selected spectral variables (highlighted in black) from EEM$_{\|}$ data used for the prediction of: PEG-LZ$_{\%}$. (Second row) First (LV1) and (Third row) second (LV2) latent variables loadings (LVs) for PEG-LZ$_{\%}$ prediction using (C/D) iPLS and (D/F) VIP selected variables..

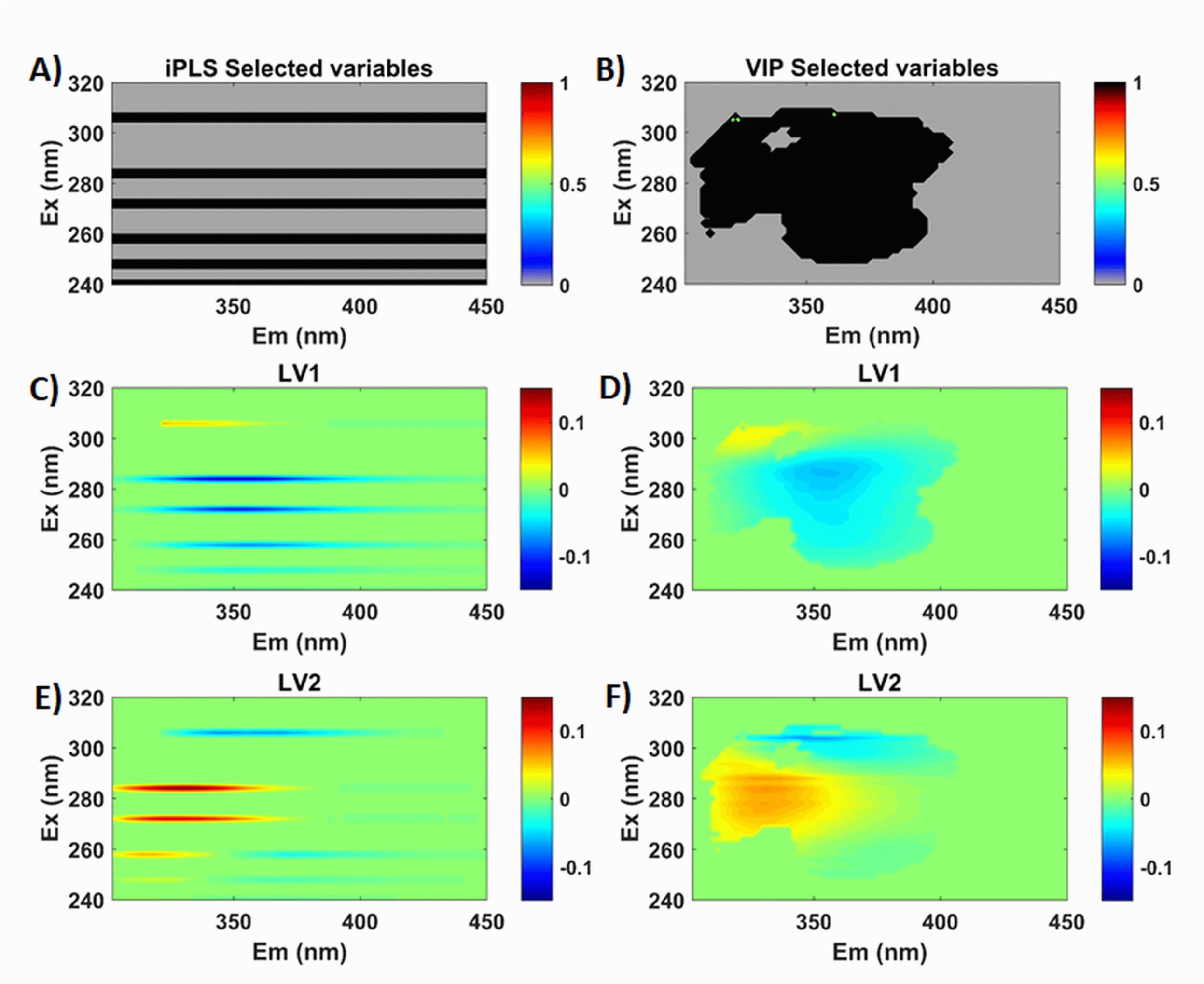


Figure 7: (A) iPLS and (B) VIP selected spectral variables (highlighted in black) from EEM$_{\parallel}$ data used for PPR prediction. (Second row) First (LV1) and (Third row) second (LV2) latent variables loadings (LVs) for PPR prediction using (C/D) iPLS and (D/F) VIP selected variables.

Published version at: **DOI:** 10.1002/bit.27483

Table 1: SEC PEGylation profile, concentration, fluorescence intensity maxima (I_{MAX}), UV aggregation index (UV-AI), Total Rayleigh Scatter volume (RS), and hydrodynamic radius (R_h) for starting materials (SM) and PEG-LZ products. Values are the mean ± SD of three replicates for each reaction condition.

Reaction		SEC							UV-Vis		EEM$_\parallel$		EEM$_\perp$	EEM$_T$	DLS
	Parameters	%LZ SM	% 1PEG-LZ	% 2PEG-LZ	% 3PEG-LZ	% Multi PEG-LZ	Total PEG-LZ %	PPR	Conc. (g/L)	UV-AI (%)	I_{MAX} (a.u.)	RS (a.u.)	I_{MAX} (a.u.)	I_{MAX} (a.u.)	R_h (nm)
R28-30 0 PEG/LZ 20°C 24h Control	**SM (#55,57,59)**	100	-	-	-	-	-	-	1.00± 0.00	0.07± 0.02	871± 6	11406 ±436	629± 12	2121± 29	2.00 ±0.06
	Purif. Product (#56,58,60)	100	-	-	-	-	-	-	0.93± 0.00	0.12± 0.01	856± 5	6024 ±158	629± 8	2111± 18	2.19 ±0.91
R13-15 1.0 PEG/LZ 20°C 2h	**SM (#25,27,29)**	100	-	-	-	-	-	-	1.00± 0.00	0.07± 0.01	879± 11	7197 ±236	637± 5	2145± 22	2.11 ±0.03-
	Purif. Product (#26,28,30)	96.0± 0.1	3.9± 0.1	0.1± 0.0	-	-	4.07± 0.13	0.04± 0.00	0.95± 0.01	0.47± 0.20	859± 21	6797 ±1494	631± 12	2117± 41	2.75± 0.16
R22-24 2.0 PEG/LZ 20°C 24h	**SM (#43,45,47)**	100	-	-	-	-	-	-	1.00± 0.00	0.08± 0.03	869± 7	7272 ±59	636± 4	2135± 15	2.07 ±0.04
	Purif. Product (#44,46,48)	57.0± 2.2	40.6± 1.6	2.4± 0.6	-	-	43.00± 1.90	0.45± 0.02	0.97± 0.01	0.17± 0.03	857± 10	7383 ±315	619± 9	2091± 22	3.00± 0.72

R31-33 4.0 PEG/LYSOZYME 20°C 24h	**SM (#61,63,65)**	100	-	-	-	-	-	-	1.00± 0.00	0.20± 0.09	878± 18	8414 ±175	645± 1	2157± 15	2.03 ±0.01
	Purif. Product (#62,64,66)	48.1± 1.3	46.3± 0.4	5.6± 1.4	-	-	51.90± 1.23	0.58± 0.02	0.96± 0.01	0.12± 0.04	852± 16	9712 ±1421	618± 12	2086± 37	3.27± 0.08
R1-3 6.0 PEG/LZ 20°C 24h	**SM (#2,4,6)**	100	-	-	-	-	-	-	0.99± 0.00	0.08± 0.03	913± 12	6984 ±404	667± 26	2244± 26	2.01 ±0.04
	Purif. Product (#2,4,6)	33.0± 3.5	43.6± 0.4	19.5± 2.2	3.5± 0.7	0.4± 0.1	66.58± 3.99	0.95± 0.88	0.99± 0.01	0.55± 0.24	897± 11	11942 ±1402	646± 12	2178± 33	3.33± 0.12
R4-6 6.0 PEG/LZ 20°C 24h	**SM (#7,9,11)**	100	-	-	-	-	-	-	1.00± 0.00	2.35± 0.03	891± 10	11453 ±1699	665± 12	2211± 39	2.01 ±0.04
	Purif. Product (#8,10,12)	28.4± 5.4	43.6± 0.4	22.7± 3.8	4.7± 1.5	0.6± 0.3	71.15± 6.18	1.05±0. 16	0.95± 0.01	0.29± 0.09	889± 22	10354 ±1359	651± 13	2181± 49	3.25± 0.71
R16-18 6.0 PEG/LZ 20°C 24h	**SM (#31,33,35)**	100	-	-	-	-	-	-	0.98± 0.00	0.13± 0.03	878± 14	15056 ±156	635± 10	2139± 41	2.08 ±0.03
	Purif. Product (#32,34,36)	28.3± 1.2	44.9± 0.3	22.2± 0.9	4.1± 0.4	0.5± 0.1	72.22± 1.17	1.04±0. 03	0.94± 0.02	0.25± 0.03	853± 12	9497 ±739	616± 8	2084± 26	3.45± 0.12
R7-9	**SM (#13,15,17)**	100	-	-	-	-	-	-	0.99± 0.00	0.11±0. 08	885± 5	11797 ±211	645± 7	2170± 24	2.05 ±0.01

12.0 PEG/LZ 20°C 24h	**Purif. Product (#14,16,18)**	9.4±0.3	33.7±1.0	37.2±0.5	16.2±1.1	3.4±0.5	90.72±0.52	1.70±0.03	0.93±0.01	0.34±0.10	854±2	16481±1902	612±5	2076±6	3.77±0.08
R10-12 12.0 PEG/LZ 30°C 24h	**SM (#19,21,23)**	100	-	-	-	-	-	-	0.99±0.00	0.16±0.13	879±19	10143±258	639±12	2146±33	2.07±0.02
	Purif. Product (#20,22,24)	4.7±0.5	22.5±1.9	37.4±1.2	26.5±1.8	9.0±1.4	95.32±0.64	2.12±0.08	0.93±0.06	0.35±0.06	825±13	14910±1143	605±19	2030±47	4.08±0.14
R19-21 24.0 PEG/LZ 20°C 24h	**SM (#37,39,41)**	100	-	-	-	-	-	-	0.99±0.00	0.06±0.01	875±5	6835±117	640±8	2149±23	2.09±0.01
	Purif. Product (#38,40,42)	1.0±0.0	21.2±1.1	35.1±0.6	32.5±1.1	10.9±0.9	96.05±5.51	2.32±0.04	0.96±0.01	0.24±0.06	857±9	13092±465	607±6	2066±20	4.30±0.25
R25-27 24.0 PEG/LZ 20°C 72h	**SM (#49,51,53)**	100	-	-	-	-	-	-	0.99±0.00	0.09±0.03	864±19	7583±59	632±11	2126±39	2.03±0.07
	Purif. Product (#50,52,54)	-	11.1±1.0	23.0±0.6	40.7±0.3	25.1±0.8	99.97±0.06	2.80±0.01	0.93±0.01	0.31±0.07	854±13	15003±615	601±15	2052±44	4.83±0.11

Published version at: **DOI:** 10.1002/bit.27483

Table 2: Statistical parameters of PLS models performance using selected datasets NoPrep-EEM$_{\perp}$, NoPrep-EEM$_{\parallel}$, for predicting PEG-LZ$_{\%}$ and Norm-EEM$_{\perp}$ and NormMC-EEM$_{\parallel}$ for PPR. Sensitivity (slope of the calibration line) was 1 in all cases. For all models, 28 samples were used for calibration model and 8 for validation.

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

† Limit of detection (LOD): ratio between 3.3 times the standard error of the intercept and slope of the calibration line.