

<https://helda.helsinki.fi>

---

## Analysis of Biologics Molecular Descriptors towards Predictive Modelling for Protein Drug Development Using Time-Gated Raman Spectroscopy

Itkonen, Jaakko

Multidisciplinary Digital Publishing Institute

2022-08-05

---

Itkonen, J.; Ghemtio, L.; Pellegrino, D.; Jokela (née Heinonen), P.J.; Xhaard, H.; Casteleijn, M.G. Analysis of Biologics Molecular Descriptors towards Predictive Modelling for Protein Drug Development Using Time-Gated Raman Spectroscopy. *Pharmaceutics* 2022, 14, 1639.

---

<http://hdl.handle.net/10138/346894>

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

## Article

# Analysis of Biologics Molecular Descriptors towards Predictive Modelling for Protein Drug Development Using Time-Gated Raman Spectroscopy

Jaakko Itkonen <sup>1,†</sup> , Leo Ghemtio <sup>1,†</sup> , Daniela Pellegrino <sup>1</sup>, Pia J. Jokela (née Heinonen) <sup>1,2</sup>, Henri Xhaard <sup>3</sup> and Marco G. Casteleijn <sup>4,\*</sup> 

<sup>1</sup> Drug Research Program, Division of Pharmaceutical Biosciences, Faculty of Pharmacy, University of Helsinki, 00100 Helsinki, Finland

<sup>2</sup> Orion Pharma, 02101 Espoo, Finland

<sup>3</sup> Drug Research Program, Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, 00100 Helsinki, Finland

<sup>4</sup> VTT Technical Research Centre Finland, 02150 Espoo, Finland

\* Correspondence: marco.casteleijn@vtt.fi; Tel.: +358-40-120-7303

† These authors contributed equally to this work.

**Abstract:** Pharmaceutical proteins, compared to small molecular weight drugs, are relatively fragile molecules, thus necessitating monitoring protein unfolding and aggregation during production and post-marketing. Currently, many analytical techniques take offline measurements, which cannot directly assess protein folding during production and unfolding during processing and storage. In addition, several orthogonal techniques are needed during production and market surveillance. In this study, we introduce the use of time-gated Raman spectroscopy to identify molecular descriptors of protein unfolding. Raman spectroscopy can measure the unfolding of proteins in-line and in real-time without labels. Using K-means clustering and PCA analysis, we could correlate local unfolding events with traditional analytical methods. This is the first step toward predictive modeling of unfolding events of proteins during production and storage.

**Keywords:** pharmaceutical proteins; Raman spectroscopy; PCA; K-means clustering; in-line measurement; biologics; protein unfolding; CD; DLS; tryptophan fluorescence



**Citation:** Itkonen, J.; Ghemtio, L.; Pellegrino, D.; Jokela (née Heinonen), P.J.; Xhaard, H.; Casteleijn, M.G. Analysis of Biologics Molecular Descriptors towards Predictive Modelling for Protein Drug Development Using Time-Gated Raman Spectroscopy. *Pharmaceutics* **2022**, *14*, 1639. <https://doi.org/10.3390/pharmaceutics14081639>

Academic Editors: David Barlow and Tihomir Tomašič

Received: 18 February 2022

Accepted: 3 August 2022

Published: 5 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Biotechnological drugs and their development started in the early 1980s and they play an increasingly important role in the treatment of many diseases, such as anemia, cystic fibrosis, cancer, and neurological diseases [1–3]. Protein drugs are produced by the use of recombinant DNA technologies in expression hosts and may have post-translational modifications [4,5]. Protein-based drugs have different characteristics compared to non-biologics; they have a higher specificity resulting in greater efficacy and reduced adverse effects [6]. On the other hand, proteins are relatively fragile molecules and prone to unfolding and forming aggregates [7,8], which are a major problem in terms of efficacy, limited solubility, and increased viscosity but may also represent the main cause of immunological responses [9,10]. Their presence, nature, and amounts are thus often considered critical quality attributes [11].

To assess the incidence of unfolding problems, sensitive analytical techniques are necessary to monitor and quantify protein aggregate levels [12]. Most techniques, however, require offline measurements of (intermediate) product(s), such as:

- (i) The detection and characterization of (sub)visible particles (e.g., visual inspection, optical microscopy, light obscuration, flow imaging, fluorescence microscopy, conductivity-based particle counter, laser diffraction, dynamic light scattering (DLS), nanoparticle tracking analysis, MALLS, turbidimetry, and nephelometry);

- (ii) the use of separation techniques for the detection and characterization of aggregates, i.e., (denaturing/reducing) size exclusion chromatography, SDS/Native PAGE, capillary-SDS electrophoresis, and AF4 [13];
- (iii) other techniques, e.g., electron/atomic force microscopy, mass spectrometry, macroion mobility spectrometry, and AUC [14]. Label-free methods for evaluating protein folding states, such as infrared spectroscopy, Raman spectroscopy, UV/VIS absorption spectroscopy, fluorescence spectroscopy, and circular dichroism spectroscopy, are relevant to mention as they are utilized as in-line analytical techniques due to their non-invasive nature. A recent review outlines the importance of Raman spectroscopy to biopharmaceuticals in greater detail [15]. However, the sensitivity, robustness, and the ability of these label-free techniques for quantification need to be improved for protein applications.

One of the first reports to evaluate protein production within living cells was using fluorescent dyes combined with fluorescence spectroscopy to detect antibody aggregates in CHO cell lysates [16]. We recently evaluated the production of CNTF in living *E. coli* cells without dyes or tags using time-gated Raman spectroscopy [17]. However, the resulting spectra are complex and multivariate. Consequently, the raw data produced can be difficult to interpret.

In recent years, multivariate data analysis and preprocessing methods have considerably increased the ability to identify relevant information contained in Raman-, and other electromagnetic-spectra, for a better qualitative and quantitative analysis of biological samples [18–20]. Principal component analysis (PCA) and data K-means clustering are well-established techniques and allow the identification of the spectral features with the highest degree of variability [21–23]. K-means clustering allows the spectra to be grouped based on spectral similarity and, therefore, identify similar features and distributions. The clustering uses information contained in the individual spectra, and the results are reported as dendrograms to show the classes available. PCA is a powerful approach widely used to discriminate different Raman spectra using scores plots and enables to derive information regarding the basis of the spectral variability. These methods are suitable for handling large multidimensional data sets and exploring the complete spectral information.

Protein aggregation can, in part, be induced due to changes in secondary protein structures. Here, we evaluate protein unfolding events *in vitro* in a buffered, controlled environment with fluorescence spectroscopy, circular dichroism spectroscopy, and time-gated Raman spectroscopy to identify molecular descriptors of protein unfolding within time-gated Raman spectra. This first step *in vitro* is needed before future evaluation of protein production using Raman spectroscopy within cells has meaning. Protein aggregation was determined with DLS. The proteins taken into consideration were divided based on the characteristics of their secondary structure:  $\alpha$ -helix,  $\beta$ -sheet, and  $\alpha/\beta$ -mix.

## 2. Materials and Methods

### 2.1. Chemical, Reagents, and Protein Samples

Calcium chloride, 2-(N-morpholino)-ethanesulfonic acid (MES), sodium chloride, glucose, ethylene diaminetetraacetic acid (EDTA), sodium azide, silver nanoparticles—40 nm particle size (AgNP; #730807) and potassium dihydrogen phosphate were obtained from Merck Sigma-Aldrich (Darmstadt, Germany). Potassium chloride was obtained from Honeywell Riedel de Haën (Seelz, Germany) and disodium hydrogen phosphate was obtained from Fisher Scientific (Hampton, VA, USA). Buffer A (20 mM phosphate buffer saline (pH 7.6), buffer B (20 mM MES 150 mM NaCl (pH 7.4), and the silver nanoparticle solution were prepared as described before [17].

In this study, several proteins were evaluated (Table 1). BSA (Bovine Serum Albumin (7% solution; SRM 927e)) was obtained from NIST (Gaithersburg, MD, USA);  $F_{ab}$ ,  $F_{(ab')_2}$ , IgG glycosylated (Immunoglobulin G), pepsin, ovalbumin (OVA), and ScTIM (triosephosphate isomerase) were obtained from Merck Sigma-Aldrich (Darmstadt, Germany); IgG non-glycosylated was obtained from AntibodyGenie (Dublin, IRL). CNTF (ciliary neurotrophic

factor) was prepared as earlier described [24,25] using buffer B; *LmTIM*<sub>E65Q</sub> was prepared as described earlier [26] and diluted with buffer A.

**Table 1.** Overview of the proteins used in this study.

Protein ( $\alpha$ , $\beta$ , $\alpha/\beta$ )	DLS	Tryptophan Fluorescence	CD	Time-Gated <sup>a</sup>
BSA ( $\alpha$ )	yes	yes	yes	yes
CNTF ( $\alpha$ )	no <sup>b</sup>	yes	yes	no <sup>c</sup>
F <sub>ab</sub> ( $\beta$ )	yes	yes	yes	no
F <sub>(ab')<sub>2</sub></sub> ( $\beta$ )	yes	no	no	no
IgG <sub>glycosylated</sub> ( $\beta$ )	yes	yes	yes	yes
IgG <sub>non-glycosylated</sub> ( $\beta$ )	yes	no	yes	yes <sup>d</sup>
Pepsin A (EC 3.4.23.1) ( $\alpha/\beta$ )	yes	no	yes	yes
Ovalbumin ( $\alpha/\beta$ )	yes	yes	yes	yes
ScTIM (EC 5.3.1.1) ( $\alpha/\beta$ )	yes	yes	yes	no
<i>LmTIM</i> <sub>E65Q</sub> (EC 5.3.1.1) ( $\alpha/\beta$ )	no	yes	yes	yes

<sup>a</sup> Samples as prepared for tryptophan fluorescence; <sup>b</sup> the DLS evaluation of *hCNTF* was recently published in Itkonen et al., 2020 [27]; <sup>c</sup> the time-gated evaluation of *hCNTF* protein aggregates was recently published in Kögler et al., 2020 [17]; <sup>d</sup> Sample was heated with a custom-built heat unit (supplementary data Figure S1).

All protein powders, except the CNTF and *LmTIM*<sub>E65Q</sub> samples, were dissolved and diluted in buffer A prior to analysis with CD spectroscopy, DLS, tryptophan fluorescence, and time-gated Raman spectroscopy. Molecular descriptors derived from each technique are listed in Table 2.

**Table 2.** Molecular descriptors per analytical technique.

Technique	Parameter 1	Parameter 2	Parameter 3
DLS	Z-average	Hydrodynamic diameter	Polydispersity index
Tryptophan fluorescence	Fluorescence intensity (internal) quenching	Red/blue shift	Tryptophan oxidation (peak at 515 nm)
CD	Melting temperature (°C)	Van't Hoff enthalpy (kJ/mol)	-
Time-gated Raman spectroscopy	Raman spectra similarity clustering according to temperature	Relevant time-gated Raman peaks <sup>a</sup>	-

<sup>a</sup> According to PCA analysis.

## 2.2. Dynamic Light Scattering

The dynamic light scattering (DLS) was performed using a Zetasizer APS (Malvern Panalytic, UK) with a 96-well plate autosampler at a 90° fixed angle and using an internal heat-controlled measuring cuvette. All protein samples (60  $\mu$ L; final concentration 0.2 mg/mL) were filtered with a 0.45  $\mu$ m syringe filter and diluted in buffer A and kept on ice prior to measurements. The pepsin solution was milky prior to dilution. Thus, a 1 mg/mL pepsin solution was spin-filtered prior to dilution with a 0.22  $\mu$ m filter (Merck-Millipore) at 12,000  $\times$  *g* for 4 min to remove additional impurities. To assess the aggregation behavior of protein samples upon heating, Rh measurements with DLS were carried out during thermal ramping (2–80 °C). Samples were heated from 2 to 80 °C with a 1 °C step size to determine the point of aggregation. The measured data were collected and analyzed with the Zetasizer Software (NANO,  $\mu$ V, APS) v6.02 (Malvern Panalytic; Malvern, UK). Three parameters were evaluated: the Z-average of the hydrodynamic diameter (automatic evaluation; *n* = 3), the lowest value of the hydrodynamic diameter (manual evaluation; *n* = 1), and the polydispersity index (PDI).

## 2.3. Tryptophan Fluorescence

Protein samples were diluted in buffer A to 0.2 mg/mL and each sample (100  $\mu$ L in triplicate) was heated for 10 min in a water bath (VWR, PA, USA) at 25, 30, 37, 40, 45, 50, 52, 55, 57, 60, 65, 67, 70, 75, and 85 °C and placed on ice prior to measurement. The *LmTIM*<sub>E65Q</sub> was subjected to higher temperatures due to its known higher thermostability (25, 50, 55,

60, 65, 70, 75, 78, 80, 83, 86, 90, 95, and 99 °C. The samples were analyzed in white closed-bottom microtiter plates (Hamilton, Reno (NV), USA) with a Thermo Scientific Varioskan LUX (ThermoFisher Scientific, Waltham (MA), USA). IgG, BSA, OVA, and F<sub>ab</sub> samples were excited at 295 nm and the emission spectrum were recorded between 314–550 nm, while LmTIM<sub>E65Q</sub> was excited at 280 nm and the emission was recorded at 290–550 nm.

#### 2.4. Circular Dichroism (CD) Spectroscopy

Samples were diluted to 0.1 mg/mL with deionized, sterile water from 1 mg/mL solutions. A Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK) was used to collect CD-spectroscopy data between 22 °C and 90 at 280 nm using a 0.1 cm path-length quartz cuvette. Data were collected every 1 nm utilizing 1 s as the integration time. Each measurement was performed in triplicate with baseline correction. Pro-Data Viewer software SX v2.5.0 (Applied Photophysics, Leatherhead, UK) was used to analyze the spectra. The melting temperature was determined via thermal unfolding of the protein sample between 190 and 260 nm with a 2 °C step size at 1 °C/min ramp rate with ±0.2 °C tolerance and subsequently analyzed with the Global3 software package v3.1 (Applied Photophysics, Leatherhead, UK).

#### 2.5. Time-Gated Raman Spectroscopy

The time-resolved measurements were performed as described earlier [17,28,29], with minor alterations. The time-gated Raman measurements (time-gated) were performed with a reduced laser power of approximately 20 mW (checked with an Ophir Nova II laser power meter, Ophir Optronics Solutions Inc., Jerusalem, Israel) at the sample to avoid photobleaching. Data acquisition software and setup control were carried out by Timegated instruments software (Timegate Instruments Oy, Oulu, Finland). Protein samples were prepared as described in Section 2.3, except for the IgG non-glycosylated sample. All protein measurements, except for IgG non-glycosylated, were performed at ambient temperature and humidity, and spectra were the sum of 11 repeats at an acquisition time of 14 min. Two separated detector ranges (1900–900 and 1400–400 cm<sup>-1</sup>) were measured in such a way. Protein samples (50 µL) were measured in a custom-made aluminum round-bottom well [17], positioned on top of a 3D-printed plastic holder within the Timegated Instruments Sample-Cube using the Timegated a common BWTek sampling probe.

The analysis of the IgG non-glycosylated sample was performed using a 3D-printed heat-ramping prototype (Figure S1) attached to a Julabo's SL-26 heat circulating water bath. Between the 3D-printed plastic, its aluminum cover, and the time-gated sampling area, 420 µL of distilled water was covered with a 0.13–0.16 mm thick cover glass (Paul Marienfeld GmbH & Co., Lauda-Königshofen, Germany), sealed with pivot grease (Eppendorf, Hamburg, Germany) to ensure optimal heat transfer. The sample (40 µL) was applied in an aluminum crucible (40 µL ME-51119870; Mettler Toledo, Switzerland) and sealed with 0.13–0.16 mm thick cover glass (Paul Marienfeld GmbH & Co., K, Lauda-Königshofen, Germany) and pivot grease (Eppendorf, Hamburg, Germany) and then placed on the Time-gated sampling area. Heat ramping was performed according to Table S2 (supplementary data). The time-gated spectra were the sum of 6 repeats (1811–555 cm<sup>-1</sup>) at an acquisition time of 6.05 min per temperature. The temperature in the crucible in the time-gated sampling area was checked with a FLIR TG165 imagining infrared thermometer prior to Time-gated sampling to ensure the correct temperature at the Time-gated sampling site. The thermometer had an error of ±1.5%, or a minimum of 1.5 °C, and the emittance was set to 0.25 according to the manufacturer's recommendations and was in line with a roughened aluminum surface [30].

#### 2.6. Data Preprocessing

The implementation of multivariate data analysis methods requires pretreatment of the raw data. The preprocessing helps to eliminate unwanted signals and to enhance the discrimination of structural features [20].

All spectral data (Table 2) was collected as text comma-separated values files to facilitate the manipulation and provide a graphical representation of the data. Data analysis was performed using R (R-Studio, Boston (MA), USA). Before statistical analysis, the data sets have been corrected for baseline and vector normalized to facilitate comparison.

### 2.7. Data Analysis

K-means clustering analysis is an unsupervised learning algorithm widely used for spectral image analysis [18]. In summary, it partitions the observations into clusters, with the cluster centroid representing the whole cluster. The pre-processed spectra are grouped according to their spectral similarity, forming clusters for particular temperature points, each characterizing regions of the image with similar molecular properties. The dissemination of similarity can be visualized over the sample image or as a dendrogram showing the hierarchical relationship between classes. Along with the spectra, additional parameters like the number of clusters ( $k$ ) and the initial cluster centers are calculated. The centroids are set by shuffling the dataset and randomly selecting  $K$  points, and then each point of the dataset is associated with the nearest centroid. The process is repeated until there is no change in the centroids. Eventually,  $k$  clusters with the most similar spectra present themselves and the centroids are determined by taking the average of all data points that belong to each group.

PCA is a multivariate analysis broadly used to reduce the dimensionality of large data sets [18]. It is mainly used to represent a multivariate data table as a smaller set of variables to identify trends, jumps, clusters, and outliers. PCA identifies a new coordinate system in the  $K$ -dimensional space that maximizes variation in the data space. This reduction help discover relationships between observations and variables and among the variables. Important information is extracted from the data and expressed as a set of indices called principal components. The importance of each PC is identified by ordering the Eigenvalues in descending order, corresponding to the descending order of variance, and denotes their importance to the dataset. The PCs contribute less in decreasing order; the first PCs contain the most information. The loading of a PC provides information on the source of the variability inside the spectra, derived from variations in the molecular components recorded from different experiments. The pre-processed dataset was evaluated by PCA, with the covariance matrix to represent the dataset by eigenvectors accounting for most of the variance and identifying the spectra' similarities. Usually, the first two or three PCs represent the highest variance present in the data sets [19].

K-means clustering and PCA were performed on the preprocessed time-gated Raman spectral data of each protein (graphs are shown in the supplementary data). The dendrograms (panel 1 in the supplementary data) show the different classes available, grouped by temperature. The cluster means (panels 2 in the supplementary data) depict the variance and the contribution of the first components. The Scree plots (panels 3 in the supplementary data) corresponding to each PC obtained by PCA have peaks that can be attributed to the protein constituents and show the region of the Raman spectra where the main differences occur (panels 4 in the supplementary data). Their respective negative and positive loadings contribute substantially to the differentiation of the protein structure. This enables one to derive information regarding the basis of the spectral variability. PCA provides thus insight into the source of the spectral variability and, therefore, the differentiation of the protein's structural components.

## 3. Results

### 3.1. Alpha Helical Proteins

CNTF is a small  $\alpha$ -helical, dimeric protein of 22.8 kDa and comprises 4  $\alpha$ -helical bundles per monomer [24]. The aggregation and unfolding of CNTF, studied by DLS and CD (Figure S2), is described in a recent publication [27], and starts at 38 °C, with an estimated  $T_m$  of 53 °C (Tables 3–5). Our CD data is in line with this earlier observation and displays a typical band at 220 nm (Figure S2), which is due to the peptide  $n \rightarrow \pi^*$

transitions and is indicative of  $\alpha$ -helical structures [31]. In the tryptophan fluorescence spectra, upon unfolding, CNTF exhibits a redshift of approximately 10 nm, indicating the movement of the polar groups from a hydrophobic environment to a hydrophilic environment. In addition, the fluorescence spectra showed a lowering trend in the intensities as the temperature increased.

BSA comprises 3  $\alpha$ -helical domains with a molecular weight of 66.5 kDa [32]. The  $\alpha$ -helical structure is evident from the CD spectra, where the negative band at 208 nm, present due to the exciton splitting of the lowest peptide  $\pi \rightarrow \pi^*$  transitions, is more prominent due to  $3_{10}$   $\alpha$ -helix structures (Figure S2) [31]. BSA aggregation, as evaluated with DLS, starts at  $\sim 58$  °C, where we observed an initial increase in intensity (data not shown) and size (Figure S3). After a short plateau, the particle size of the aggregates increased swiftly, which is even observable in the sample polydispersity (Figure S3).

From the tryptophan fluorescence spectra (Figure S4), we observed a redshift of approximately 10 nm, indicating the movement of the polar groups from a hydrophobic environment to a hydrophilic environment, even though the intensities of the BSA spectra were relatively low at 25–30 °C and with intensities highest at 45–50 °C, to then come down to the same level of intensity slowly as 25 °C at temperatures above 70 °C.

The time-gated Raman spectra of BSA showed a clear drop in intensity between 55 and 57 °C (Figure S5), especially around  $1650\text{ cm}^{-1}$ . In addition, the spectra grouped according to their similarity shown in Figure S6A,B panel 1, showed that the spectra recorded at 25–55 °C differed from the spectra recorded at 57–85 °C. This difference correlates with the start of aggregation seen in DLS. The two groups were well discriminated in the cluster means, as shown in panel 2 (Figure S6A,B). PC1 and PC2 contributed to most of the explained variance and allowed the discrimination between the two groups. Raman peaks were identified that contribute to the PC scores. The positive and negative correlation of PC1 and PC2 is depicted in Figure S6A,B (panel 4), where zero is the dashed line. At Raman peaks, where the spectral differences between data exist (i.e., the correlation of PC1 and PC2 is in the opposite direction), the corresponding physical changes in protein bonds were relevant changes due to thermal unfolding. The significant differences in the secondary structure of BSA due to the increase in temperature are summarized in Table S3.

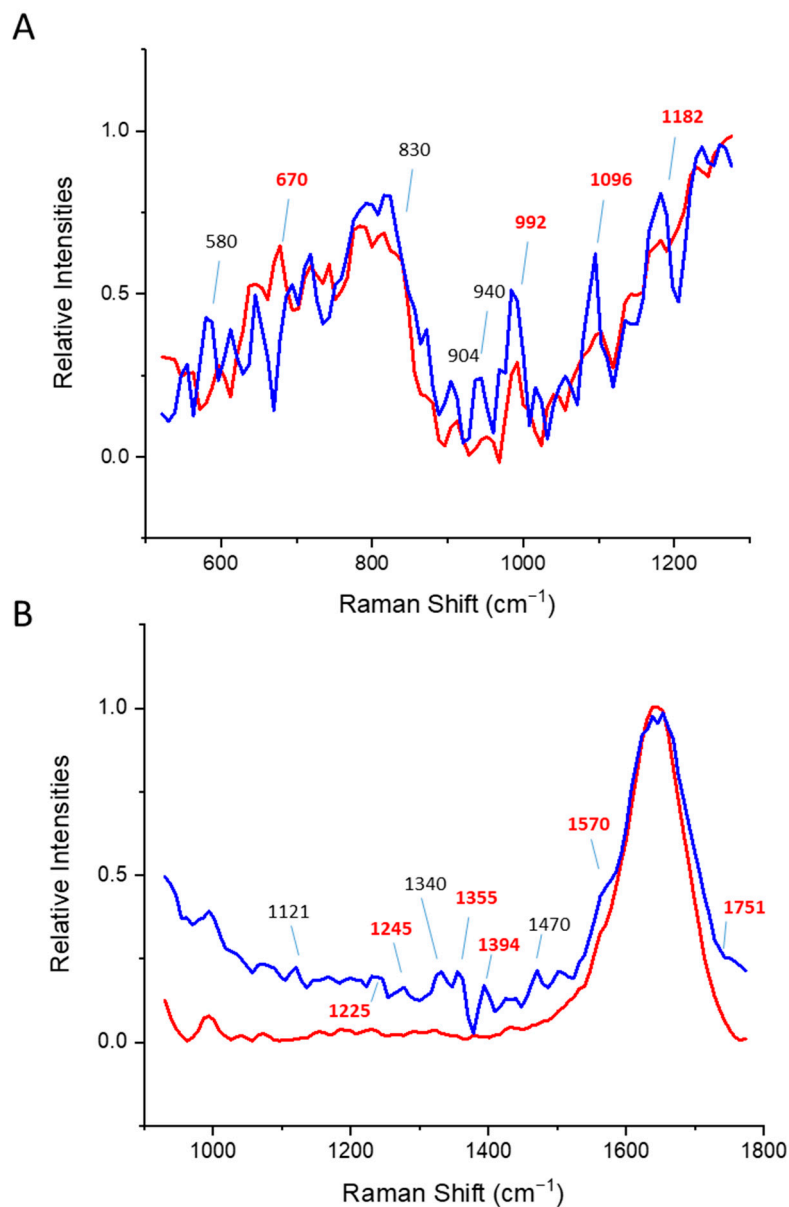
### 3.2. Beta Sheet Proteins

One example of a  $\beta$ -sheet protein is pepsin A, an archetypal aspartic proteinase belonging to the class of endopeptidases. The aspartic proteinases display a predominant  $\beta$ -fold with only a few short helical segments [33], though it has been classified by Rygula et al. (2013) [34] as  $\beta$ -sheet protein. This was evident from the CD-spectra (Figure S7), as the spectrum was a mix of mainly random coil and  $\beta$ -sheets. At the same time, the protein appeared to be disordered at pH 7.0. Pepsin started aggregating at this pH at about 64 °C (Figure S8B). The main observation from the tryptophan fluorescence spectra at different temperatures (Figure S9) was that the intensities were very low for all spectra, and as such, no conclusions can be drawn.

The time-gated Raman spectra of pepsin showed no clear drop in intensity due to the increase in temperature (Figure S10). As such, the spectra grouped according to their similarity at different temperature points shown in Figure S13A (panel 1) did not show significant differences. This corresponded to the observations in the CD-spectra that the protein, in this particular low-quality sample, was already partly unfolded at room temperature. PC1 and PC2 contributed to only 50% of the explained variance. Raman peaks were identified that contributed to the PC scores. However, the large changes in the relative intensities observed in the grouping in Figure S13B panel 1 show that the spectra at clusters 25, 30, 40, 45, 75, and 85 °C and at clusters 37 and 50–70 °C show similarities. These observations did not correspond with the aggregation temperature of 64 °C, indicating further that the pepsin protein solution used in this study was of poor quality.

The positive and negative correlation of PC1 and PC2 is depicted in Figure S13A,B (panel 4), where zero is the dashed line. At Raman peaks where the spectral differences

between data exist (i.e., the correlation of PC1 and PC2 is in the opposite direction), the corresponding physical changes in protein bonds are relevant changes due to thermal unfolding. Changes in the secondary structure following the increase in temperature can be observed in Figure 1, Figure S10, and Figure S13A,B, and summarized in Table S4.



**Figure 1.** Averaged ( $N = 11$ ) and normalized time-gated spectra of pepsin at 25 °C (blue) and 65 °C (red). (A,B) correspond to the individual, non-processed, and non-normalized Time-gated spectra as presented in the left and right panes in Figure S10. (i.e., raw data). Raman values of significance, as presented in Table S4 (supplementary data), are depicted in bold and red.

When directly comparing the relative time-gated Raman spectra of pepsin at ambient temperature and 65 °C (Figure 1), we observed changes in the secondary structure upon aggregation. It is also evident that at ambient temperature, pepsin was already partly unfolded, indicated by the peak at 1245 cm<sup>-1</sup>, typical of random coil structures. The fermi doublet ratio of  $\sim 1$  showed that the buried tryptophans were more exposed to the hydrophilic matrix. The ratio of the tyrosine peaks at 850/830 cm<sup>-1</sup> = 0.65 at 25 °C to 0.43 at 65 °C showing tyrosines act as a strong H-bond donor with little change between these temperatures.



Another class of  $\beta$ -sheet proteins are antibodies or their fragments [35,36], which during the last decades, have proven themselves as a highly effective and specific class of biological drugs if they maintain a high thermostability and low aggregation propensity [37]. Their structures are very well characterized and proteolytic cleavage to remove the Fc tail results in either  $F_{ab}$  or  $F_{(ab')_2}$  fragments [35]. In addition, IgG antibodies are naturally modified by the decoration of glycan sugars [38]; however, deglycosylated IgG has its place as a therapeutic as well [39]. The  $\beta$ -sheet secondary structure was well characterized by the CD spectra of both the  $F_{ab}$  fragment and the full IgGs (Figure S7). We derived the clear right twisted anti-parallel  $\beta$ -sheet formation [40] as expected in the  $F_{ab}$  fragment and the non-glycosylated IgG spectra [41] as they consist of seven  $\beta$ -strands with four strands forming one  $\beta$ -sheet and three strands forming a second sheet. However, glycosylated IgG appeared to be more relaxed or exhibited left twisted anti-parallel  $\beta$ -sheets (Figure S7). The stability of IgG, or its derived fragments, appeared similar based on their melting temperatures (Table 3); however, individual  $F_{ab}$  fragments appeared to aggregate at slightly lower temperatures (Figure S8 and Table 5). Tryptophan fluorescence spectra comparing the  $F_{ab}$  fragments and the full IgG appeared very similar (Figure S9).

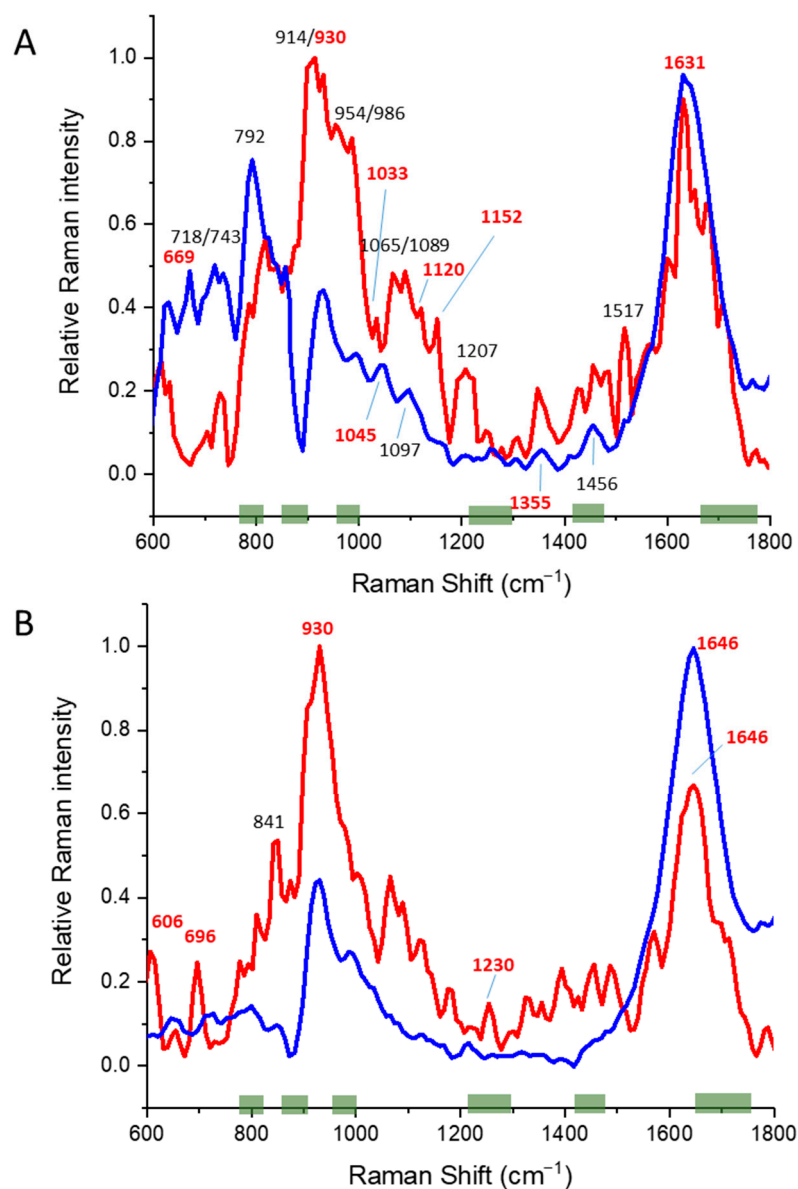
However, when overlaying the time-gated Raman data of glycosylated and non-glycosylated IgGs, there were clear differences at 20 °C (Figure 2A). Since aggregation had its onset at ~64 °C (Table 4), we also compared time-gated Raman spectra at 65 °C (Figure 2B), where the changes differed in response to the rise in temperature. We observed a clear shift in the amide I peak from 1631 to 1646  $\text{cm}^{-1}$  due to the loss of  $\beta$ -sheet structures [42] and a reduction of the 1097  $\text{cm}^{-1}$  peak. Yet, at ambient temperatures, both spectra differed greatly. In non-glycosylated IgG time-gated spectra, we observed changes at 954/986  $\text{cm}^{-1}$  (likely the protein backbone) and the 1207  $\text{cm}^{-1}$  peak (cysteine or the  $\nu\text{SO}_4$  peak) [43]. To our surprise, we observed the very characteristic carotenoid peaks at 1152 and 1517  $\text{cm}^{-1}$  due to C–C and conjugated C=C bond stretches [44], on which we can only speculate them to be a remnant of the production process.

In the glycosylated IgG spectra (Figure 2A,B), we observed the reduction of the C–H (def) peak at 1456  $\text{cm}^{-1}$ , Trp C $\alpha$ -H (def) peak at 1455  $\text{cm}^{-1}$ , and changes in 1045  $\text{cm}^{-1}$ , 1097  $\text{cm}^{-1}$ , all likely marker bands of aromatic side chains affected upon heating. The relative Raman intensities between 600–900  $\text{cm}^{-1}$  appeared to drop dramatically; however, this could also be an artifact due to the detector shift between lower and higher wavenumbers (Raman shift) in this particular measurement. Despite these differences, the peak at 606  $\text{cm}^{-1}$ , likely the CCC deformation in-plane vibration mode of the phenylalanine ring, in the non-glycosylated spectra did not change due to heating [45]. Of interest is the ratio of the Raman peak intensity seen in the tyrosine doublet Raman bands near 850 and 830  $\text{cm}^{-1}$  of 0.89 at 20 °C and 1.14 at 65 °C. This shift indicates that upon heating, the 10–12 tyrosines present in the IgG were more exposed to the solvent, albeit the shift is small.

In the non-glycosylated IgG sample, we observed the disappearance of the carotenoid peaks, the likely protein back-bone peaks at 954/986  $\text{cm}^{-1}$ , the 1207  $\text{cm}^{-1}$  peak, and the C–N peak at 1120  $\text{cm}^{-1}$ . Overall, we saw fewer changes in the overall spectra compared to glycosylated IgG. The  $\text{Int}_{850}/\text{Int}_{830}$  ratio of tyrosine shift from 0.89 at 20 °C to 1.53 at 65 °C indicates a higher exposure of tyrosines to the solvent upon heating.

The time-gated Raman spectra of glycosylated IgG did not show a clear drop in intensity in the amide I peak (1650–1660  $\text{cm}^{-1}$ ) due to the increase in temperature (Figure S11). The intensity rose first and then dropped at the higher temperatures. As with pepsin, the spectra grouped according to their similarity at different temperature points shown in Figure S14A (panel 1) do not show any differences. However, there is no indication in the CD-spectra that the protein was partly unfolded (Figure S7). The one high intensity at 50 °C at 880  $\text{cm}^{-1}$  indicated a specific large change in the tryptophan environment. In addition, the grouping in Figure S14B panel 1 shows that the spectra at clusters 25, 30, 40, 37, 40, 45, 60, 70, and 85 °C and at clusters 50, 52, 55, 57, 65, and 75 °C show similarities. Yet, PC1 and PC2 contributed around 50% of the explained variance, so it does not allow the discrimination between the two groups. These observations did not directly correlate with

the aggregation temperature of 64 °C and CD melting temperatures ( $T_m$ ) of 65.5 and 72.1 °C. Upon closer inspection in Figure S14B panel 2, we observed the second cluster was grouped due to changes around a Raman shift of 1460–1480  $\text{cm}^{-1}$  (C–H and aliphatic side chains) at these temperatures. PC1 and PC2 contribute around 60% of the explained variance. Significant changes in the IgG secondary structure due to the increase in temperature can be observed in Figures S14 and S15, summarized in Table S5.



**Figure 2.** Averaged and normalized Time-gated spectra of IgG (glycosylated; blue;  $N = 11$ ) and IgG (non-glycosylated; red;  $N = 6$ ) at (A) 25 °C, (B) and at 65 °C. Individual, non-processed, and non-normalized Time-gated spectra (i.e., raw data) are shown in Figures S11 and S12. Regions of significance in respect of the glycosylation status of proteins, as determined by Brewser et al. (2011) [46] are shown in green. Raman values of significance, as presented in Table S5, are depicted in bold and red.

Summarizing, for  $\beta$ -sheet proteins, an interesting observation is the relevance change of C=O stretching peak between 1760–1840  $\text{cm}^{-1}$ , most likely in the carbonyl groups upon thermal unfolding. Overall, the intensity decrease in the amide I peak is likely due to the increased interaction of amino acids with water, thus indicating the unfolding of the  $\beta$ -sheet proteins. In addition, significant changes observed both in IgG type structures

and pepsin are changes due to tryptophan in the fingerprint region. The latter observation correlates with the major changes observed in the tryptophan fluorescence spectroscopy.

### 3.3. Alpha/Beta Proteins

Ovalbumin is the main protein found in egg white (~55% of the total protein) and consists of 385 amino acids, with a relative molecular mass of 42.7 kDa. Ovalbumin contains several post-translational modifications, including N-terminal acetylation (G1), phosphorylation (S68, S344), and glycosylation (N292). Ovalbumin's internal signal sequence (residues 21–47) is not cleaved off but remains as part of the mature protein. Ovalbumin displays sequence and three-dimensional homology to the serpin superfamily, but unlike most serpins, it is not a serine protease inhibitor.

The secondary structure, also supported by the CD-spectrum (Figure S16), was comprised of  $\alpha$ -helices (12) and  $\beta$ -sheets (15). The  $\beta$ -sheets form the core of the protein, while the  $\alpha$ -helices form the outside of the protein, especially in its dimeric form [47].

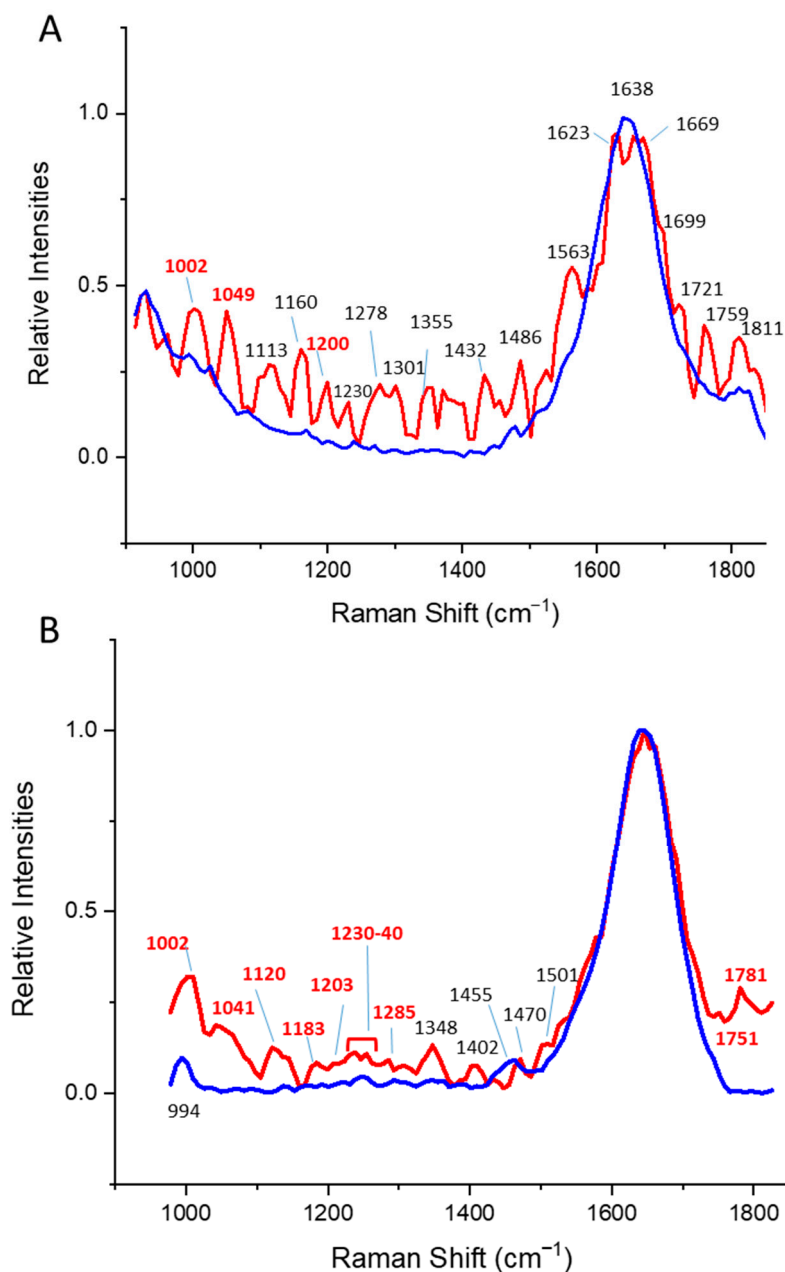
Ovalbumin aggregation could not be evaluated with DLS in this study as the samples seemed contaminated with larger molecules or were already partly aggregated prior to heating in the DLS-cuvette (Figure S17A). Earlier reports indicated aggregation to start at ~71 °C [48]. According to the tryptophan fluorescence spectra (Figure S18), the intensities of the spectra were highest between 37–40 °C and 50 °C, with a lower intensity at the highest temperatures (75–85 °C). These results indicate that neighboring amino acids initially reduced quenching, while the temperature-induced changes in the secondary structures then increased quenching. In addition, we observed a blue shift, and since W149 and W268 are buried at room temperature and bound to charged groups, water molecules may create a blue shift in this environment [49]. W185 is buried and stabilized by hydrophobic groups [47].

The time-gated Raman spectra (Figure S19) show a dramatic drop in intensity of the amide I peak between 40–50 °C, indicating denaturation [50], with some recovery at 60 °C. In addition, the spectra grouped according to their similarity at different temperatures shown in Figure S21A,B panel 1, show that the spectra from 25–40 °C were more similar than the spectra from 55–85 °C. Combined with the tryptophan fluorescence data, it appears that our sample started denaturing at lower temperatures than earlier reported [48].

Significant changes in the secondary structure of ovalbumin following the increase in temperature were observed in the time-gated Raman spectra (Figures 3 and S19) and are summarized in Table S6. Upon heating, we observed the reduction of  $\alpha$ -helical structures, as indicated by the amide I peak splitting and the rise of the tryptophan indole ring peak at 1561  $\text{cm}^{-1}$  (Figure 3A). Furthermore, due to the overall reduction of the peak intensity of the amide I peak, the peak details seen in the normalized data appear enhanced in the heated sample. PCA analysis of the time-gated Raman spectra in Figure S21A,B, panel 4, identified the relevant changes due to a positive correlation in the PC1 components of Raman peaks of phenylalanine (at 1015–1020  $\text{cm}^{-1}$ ), the amide II (at 1200  $\text{cm}^{-1}$ ) and C=O stretching bond (at 1750–1860  $\text{cm}^{-1}$ ). Further modeling of the amide I peak during thermal unfolding could give a better insight into the unfolding of ovalbumin.

Triosephosphate isomerase (TIM) comprises the classical  $\alpha/\beta$  barrel [51] and the wild-type exists as a dimeric protein. The *Leishmania mexicana* (*Lm*) mutant E65Q (*Lm*TIM<sub>E65Q</sub>) is a thermostable variant of the wild-type protein [26]. This mixed  $\alpha$ -helix/ $\beta$ -sheet structure is clearly observed in the CD-spectrum (Figure S16), and the *Lm*TIM<sub>E65Q</sub> variant appears to be better folded than *Saccharomyces cerevisiae* (*Sc*) TIM under the same conditions. *Sc*TIM aggregation evaluated with DLS in this study (Figure S17B) indicates aggregation to start at ~58 °C, which was in correlation with the CD-melting curve (Table 3). In the tryptophan fluorescence spectra (Figure S18), the intensities of the spectra were highest between 37–40 °C, with a lower intensity at the highest temperatures (75–85 °C). These results indicate that neighboring amino acids initially reduced quenching, while the temperature-induced changes in the secondary structures then increased quenching. In addition, we observe for both TIMs a redshift above 300  $\text{nm}^{-1}$ , indicating exposure of the tryptophans

to the solvent. In *Sc*TIM, both W89 and W156 are buried and stabilized by hydrophobic groups [52], while in the *Lm*TIM<sub>E65Q</sub> W11, W160, and W194 are buried and stabilized by hydrophobic groups, while the buried W91 is bound to a polar group [53]. Both the W167 in *Sc*TIM and W170 in *Lm*TIM<sub>E65Q</sub> reside in the hinge of the catalytic loop and thus are exposed to the solvent [54].



**Figure 3.** Averaged ( $N = 11$ ) and normalized Time-gated spectra of (A) ovalbumin at 25 °C (blue) and 50 °C (red), and (B) *Lm*TIM<sub>E65Q</sub> at 25 °C (blue) and 86 °C (red) correspond to the individual, non-processed, and non-normalized time-gated spectra as presented in Figures S19 and S20 (i.e., raw data). Raman values of significance, as presented in Tables S6 and S7, are depicted in bold and red.

The time-gated Raman spectra show a large reduction in relative intensity between 83 and 86 °C for *Lm*TIM<sub>E65Q</sub> (Figure S20), which is in accordance with the CD melting curve and tryptophan fluorescence measurement (Tables S9 and S11). In addition, the spectra grouped according to their similarity shown in Figure S22A panel 1 show that the spectra from 25–86 °C were more similar than the spectra from 90–99 °C. However, spectra

grouped in Figure S22B, differ, as we observed earlier in the IgG spectra, due to changes around Raman Shift of 1460–1480  $\text{cm}^{-1}$  (C–H and aliphatic side chains) at 25, 30, 40, 45, 75, and 85 °C. PC1 and PC2 contribute to most of the explained variance and allows the discrimination between the two groups.

Significant changes in the secondary structure due to the increase in temperature in  $LmTIM_{E65Q}$  were observed in time-gated Raman spectra (Figures S20 and S22; summarized in Table S7). Upon heating, we observed the changes in the secondary structure of the  $\alpha/\beta$  barrel, as indicated by the significance of changes in the amide II peaks at 1285  $\text{cm}^{-1}$  ( $\alpha$  helix/  $\beta$ -sheet) and 1230–1240  $\text{cm}^{-1}$  ( $\beta$ -sheet) (Figure 3A). Furthermore, due to the reduction of the peak intensity of the amide I peak, details in other regions of the data were enhanced. The unfolding of this variant of  $LmTIM$  is well understood [53]. At neutral pH, the active dimer unfolds into partially unfolded monomers, which are prone to aggregation. Hence, the relatively small changes in the amide I peak, compared to ovalbumin, are due to remaining partially folded monomers. Unlike in the ovalbumin spectra, we observed significant changes in the peptide bonds in  $LmTIM_{E65Q}$ .

When comparing common features of unfolding in ovalbumin and TIM, we observed that the K-clustering profiles correlated well with the tryptophan fluorescence. In both cases, when the red shift occurs and the intensities drop, we also observed a reduction in the amide I peaks in the time-gate Raman spectra. The relevant changes derived from the PCA analysis observed for both proteins (Figures S21 and S22, and Table 6) are due to the phenylalanine peak 980–1020  $\text{cm}^{-1}$ , the amide II peak (1200–1205  $\text{cm}^{-1}$ ), and the C=O stretching bond between 1820–1860  $\text{cm}^{-1}$ .

#### 4. Discussion

Pharmaceutical proteins have proven to be very important in the field of medicines and vaccines [1]. The practice of pharmacovigilance has gained significant momentum since 1963, following the thalidomide tragedy [55]. As such, the evaluation of proteins during biotechnological production, downstream processing, and storage are crucial. Since protein function is linked directly to their three-dimensional shape or structural fold, the evaluation of changes in their secondary structure during the above-mentioned processes sheds important insights into their stability, which is directly linked to safety and efficacy [56].

Earlier, we addressed the notion that there is a need to directly monitor the intermediate products during protein production within the living cells [17]. Before we can utilize Raman spectroscopy for this task, we first need to understand how protein Raman spectra change in the function of unfolding and aggregation. In this study, we induced changes in the secondary structure of several proteins in in-vitro conditions via thermal ramping. Then we used time-gated Raman spectroscopy coupled with PCA analysis of the Raman spectra to evaluate relevant molecular descriptors toward predictive modeling of unfolding and aggregation of pharmaceutical proteins. Furthermore, we created a novel 3D-printed heat-exchange Raman sample holder for expensive samples. We evaluated several proteins to find common molecular descriptors within time-gated Raman spectra regarding the thermal unfolding of proteins and compare the changes in Raman spectra with established spectroscopic methods often used to evaluate changes in the secondary structure of proteins and the formation of aggregates. Finally, we could identify different changes due to thermal unfolding between the three protein classes ( $\alpha$ ,  $\beta$ ,  $\alpha/\beta$ ). For each protein, a deeper analysis using NMR unfolding studies and additional unfolding studies using different unfolding methods would be very insightful for each specific protein; however, this was not the aim of this study.

Overall, CD measurements are in accordance with earlier reports (references are listed in Table 3), taking into account differences in concentrations or buffer. The DLS results, as summarized in Table 4, show insight into the start of aggregation and cannot be compared to CD melting curves; however, similar trends can be observed upon heating the different proteins. The tryptophan-fluorescence measurements, summarized in Table 5, showed how the local environment of tryptophan changes during unfolding and aggregation. Changes

in intensities are due to quenching events [57], while red and blue shifts are due to changes in specific interactions with tryptophans within the protein [49]. We observe all these effects, but not tryptophan oxidation [58].

**Table 3.** Summarized CD results.

Protein ( $\alpha$ , $\beta$ , $\alpha/\beta$ )	T <sub>m</sub> (°C)	Van't Hoff Enthalpy (kJ/mol)	Literature Value T <sub>m</sub> (°C)
BSA ( $\alpha$ )	61.2 ± 0.1 <sup>a</sup>	256.6 ± 8.4	63 [59]
CNTF ( $\alpha$ )	55.0 ± 1.9 <sup>b</sup>	360.0 ± 34.3	53 [27]
F <sub>ab</sub> ( $\beta$ )	73.9 ± 0.3	437.5 ± 17.4	61–70 [60,61]
IgG <sub>glycosylated</sub> ( $\beta$ ) <sup>c</sup>	65.5 ± 0.2	336.6 ± 7.9	60–68
IgG <sub>non-glycosylated</sub> ( $\beta$ )	72.1 ± 0.4	203.9 ± 10.6	71–77 [60,62]
Pepsin ( $\beta$ )	71.5 ± 0.2	265.2 ± 12.6	62–66 [63]
Ovalbumin ( $\alpha/\beta$ )	49.9 ± 0.2 <sup>d</sup>	352.1 ± 17.0	52 <sup>e</sup> [64]
ScTIM ( $\alpha/\beta$ )	72.3 ± 0.1	181.3 ± 2.2	71–76 [65]
ScTIM ( $\alpha/\beta$ )	55.3 ± 1.7	360.0 ± 13.4	~58 [66]
LmTIM <sub>E65Q</sub> ( $\alpha/\beta$ )	81.0 ± 0.3 <sup>a</sup>	172.9 ± 12.6	83 [26]

<sup>a</sup> The protein was not totally unfolded at 92 °C; <sup>b</sup> Tighter  $\alpha$ -helical folding occurred between 30–40 °C; <sup>c</sup> Additional unfolding at 35 °C; <sup>d</sup> Protein might be aggregated to disordered at pH 7 during analysis; <sup>e</sup> At pH = 8.0.

**Table 4.** Summarized DLS results.

Protein ( $\alpha$ , $\beta$ , $\alpha/\beta$ )	Hydrodynamic Diameter at 20 °C [nm]	T <sub>aggregation</sub> [°C]	Literature Value T <sub>aggregation</sub> [°C] <sup>c</sup>
BSA ( $\alpha$ )	8.0	58	~62 [67]
CNTF ( $\alpha$ )	ND	ND	38 <sup>a</sup> [27]
F <sub>ab</sub> ( $\beta$ )	205.8	60	-
F <sub>(ab')2</sub> ( $\beta$ )	11.2	63	-
IgG <sub>glycosylated</sub> ( $\beta$ )	12.3	64	55–80 [68]
IgG <sub>non-glycosylated</sub> ( $\beta$ )	ND	ND	-
Pepsin ( $\beta$ )	50.9	64	-
Ovalbumin ( $\alpha/\beta$ )	28.8	- <sup>b</sup>	71 [48]
ScTIM ( $\alpha/\beta$ )	68.4	58	-
LmTIM <sub>E65Q</sub> ( $\alpha/\beta$ )	ND	ND	-

<sup>a</sup> Hydrodynamic radius at 2 °C is 2.42 ± 0.50 nm and 2.95 ± 0.22 nm (two different buffers were used [27]);

<sup>b</sup> Could not be determined (see Figure S12B); <sup>c</sup> To best of our knowledge.

**Table 5.** Summarized tryptophan fluorescence results.

Protein ( $\alpha$ , $\beta$ , $\alpha/\beta$ )	Maximum Fluorescence Intensity at Temperature [°C]	Red/Blue Shift <sup>a</sup>	Tryptophan Oxidation (Peak at 515)
BSA ( $\alpha$ )	45	Red (10 nm)	No
CNTF ( $\alpha$ )	30	Red (12 nm)	No
F <sub>ab</sub> ( $\beta$ )	85	Red (4 nm)	No
F <sub>(ab')2</sub> ( $\beta$ )	ND <sup>b</sup>	ND	ND
IgG <sub>glycosylated</sub> ( $\beta$ )	85	Red (5 nm)	No
IgG <sub>non-glycosylated</sub> ( $\beta$ )	ND	ND	ND
Pepsin ( $\beta$ )	37	No	No
Ovalbumin ( $\alpha/\beta$ )	37	Blue (5 nm)	No
Ovalbumin ( $\alpha/\beta$ )	50	Blue (5 nm)	No
ScTIM ( $\alpha/\beta$ )	65	Red (10 nm)	No
LmTIM <sub>E65Q</sub> ( $\alpha/\beta$ )	86	Red (6 nm)	No

<sup>a</sup> Comparing the maximum of the peaks at lowest temperatures to the highest; <sup>b</sup> ND = Not Determined.

**Table 6.** Summarized results of changes in the time-gated spectra identified with K-means clustering and PCA analysis.

Protein Structural Class	Most Relevant Changes [cm <sup>-1</sup> ]	Bond Type	Relevant Temperature Change Correlates with
α	880–900	Trp	Trp-fluorescence
	910	Ser <sup>a</sup>	DLS
	940	N-Ca-C	CD
	950	N-Ca-C	
	970	Ser/His <sup>a</sup>	
	1180	Val/Arg/other amino acids <sup>a</sup>	
	1280	Amide II	
	1310	Phe, Tyr, Trp <sup>a</sup>	
β	1350–1390	Trp <sup>a</sup>	Trp-fluorescence
	1695–1760	Amide I/carbonyl stretch	
α/β	980–1020	Phe	Trp-fluorescence
	1200–1205	Amide II	DLS
	1820–1860	C=O	CD

<sup>a</sup> According to De Gelder et al. (2007) [69].

Changes in the time-gated spectra are related to changes in the secondary structure of the protein. While the intensity reduction of the peptide-bond (C–N) peak would clearly indicate protein degradation and similar changes of the phenylalanine peak at ~1000 cm<sup>-1</sup> would indicate changes in the protein concentration, other changes are more subtle and a reflection of secondary structure changes. Table 6 summarizes shared regions per protein structural class, though only one α-helical protein was evaluated.

One drawback of the TimeGated<sup>TM</sup> device used in the study was a baseline artifact due to switching the measuring window from a lower to a higher range. We chose this set-up to maximize the number of datapoints and thus lower the deviation. We intentionally do not correct for the baseline in this study as we were aiming for a method that can be utilized by only rapidly evaluating the raw data. However, if changes in both regions were significant, the PCA analysis isolated changes as significant. The difference in intensity between the low and the higher wavenumber scans of the same sample is due to the higher intensity from the Amide I peak during the second scan. In the non-glycosylated sample, we opted for larger steps to cover the same spectrum in one scan, but here we have fewer repetitions per scan and thus introduce a larger standard deviation.

When we compare time-gated Raman data of (partly) unfolded CNTF [17] with the thermal unfolding of BSA, we observed β-sheet formation due to the rise of a peak at 1332 cm<sup>-1</sup> and changes in the amide III peak of tyrosine. The most significant changes in BSA are in the backbone, in the secondary C–N peak at 1130 cm<sup>-1</sup>, and in the peak at 1180 cm<sup>-1</sup> (Table 6). Unlike the other two protein classes, no common features can be observed other than tyrosine peak changes at 1240 cm<sup>-1</sup> (amide III (β-sheets); Table 6), thus shifting away from α-helical structures.

Finally, we combined all time-gated Raman spectra in the PCA analysis to utilize the noise filtering properties of the method on the lower-scanning region of the detector of the TimeGated<sup>TM</sup> device used in the study (500–1300 cm<sup>-1</sup>). The positive and negative correlation of PC1 compared to PC2 for all the Raman peaks combined to show three regions of particular interest (Figure S23). Changes in the peaks in, or close, to the amide III region (1180–1300 cm<sup>-1</sup>), tryptophan (880 cm<sup>-1</sup>), and phenylalanine (~1000 cm<sup>-1</sup>) were present in all proteins due to thermal unfolding.

Rather than dissecting each protein structure in detail to identify the structural changes during the unfolding and aggregation events as observed in the time-gated Raman spectra, we aimed to reduce the raw data to smaller components to identify significant changes pertaining to a protein structural class.

The data analysis used in this study is an unsupervised method; therefore, the resulting components do not necessarily reveal the features directly linked to the classification but

represent the sources of variation and the representative properties of the raw data. Even with the small data set presented here, relevant changes in secondary structures of proteins correlate with other, more traditional, label-free methods such as DLS, CD-spectroscopy, and tryptophan fluorescence. As such, identifying relevant changes is possible, but it will require a larger dataset to create a model predicting relevant structural changes based on changes in Raman spectra. Additionally, a higher resolution Raman dataset using more advanced detectors would be beneficial in reducing the noise-to-signal ratio and to improve the PCA analysis.

## 5. Conclusions

In combination with K-means clustering, PCA sheds further light on changes in the structural elements of the Raman spectroscopy spectra. Principal component analysis can be considered a noise filtering method. The relevant differences are captured in the first components, while the higher components contain noise only. The spectra can be reconstructed using only the first  $p$  components. The current study demonstrates the capabilities of time-gated Raman spectroscopy in characterizing structural changes of proteins under different experimental conditions without offline sampling and the addition of protein labels. Time-gated Raman spectroscopy gives valuable insights into secondary protein structures that correlate with observations with tryptophan fluorescence spectroscopy, dynamic light scattering, and circular dichroism spectroscopy. Additional variations to perturb proteins could be added as additional parameters to identify additional descriptors of protein unfolding. Raman signals can then be translated into high-level structural information of interest to derive statistical models from being used to predict the relative folding states of unknown samples compared to fully folded proteins. We intend to create additional data sets for building such models in the near future.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pharmaceutics14081639/s1>, Supplementary data file, which includes Table S1–S7 and Figures S1–S23. References [27,34,69–71] are cited in the Supplementary Materials.

**Author Contributions:** Conceptualization, M.G.C. and L.G.; methodology, J.I., L.G., D.P., P.J.J. and M.G.C.; data curation, L.G.; writing—original draft preparation, M.G.C.; writing—review and editing, J.I., P.J.J., M.G.C. and L.G.; visualization, M.G.C. and L.G.; supervision, M.G.C. and H.X.; project administration, M.G.C.; funding acquisition, M.G.C. and H.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Academy of Finland project 303884 and VTT technical research Centre of Finland. L.G. gratefully acknowledges the support of the Drug Discovery and Chemical Biology Network of Finland. J.I. has also been funded by grants from the Finnish Cultural Foundation, the Evald and Hilda Nissi Foundation, the Päivikki and Sakari Sohlberg Foundation, and the Paulo Foundation. D.P. received Erasmus+ program funding from the Italian Ministry of Education.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors acknowledge the use of Instruct-HiLIFE Crystallization unit (University of Helsinki, Biocenter Finland, and Instruct-FI). The use of the facilities and expertise of the Biocenter Oulu protein biophysical analysis core facility, a member of Biocenter Finland, is gratefully acknowledged. We kindly thank Rik Wierenga for providing us with the *LmTIM*<sub>E65Q</sub> protein. We also would like to thank Leena Pietilä for her kind laboratory assistance and Regina Casteleijn-Osorno of Aalto University (Finland) for comments that greatly improved the manuscript. We kindly like to thank M. Kögler of VTT (Finland) for his expert advice. We kindly acknowledge our funding sources and support from the CSC IT Center for Science Ltd. is thanked for organizing computational resources.



**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Casteleijn, M.G.; Richardson, D. Engineering Cells and Proteins—Creating pharmaceuticals. *Eur. Pharm. Rev.* **2014**, *2014*, *4*.
2. Leader, B.; Baca, Q.J.; Golan, D.E. Protein therapeutics: A summary and pharmacological classification. *Nat. Rev. Drug Discov.* **2008**, *7*, 21–39. [[CrossRef](#)] [[PubMed](#)]
3. Sauna, Z.E.; Lagassé, H.A.D.; Alexaki, A.; Simhadri, V.L.; Katagiri, N.H.; Jankowski, W.; Kimchi-Sarfaty, C. Recent advances in (therapeutic protein) drug development. *F1000Research* **2017**, *6*, 113. [[CrossRef](#)]
4. Quianzon, C.C.; Cheikh, I. History of insulin. *J. Community Hosp. Intern. Med. Perspect.* **2012**, *2*, 18701. [[CrossRef](#)]
5. Zheng, K.; Bantog, C.; Bayer, R. The impact of glycosylation on monoclonal antibody conformation and stability. *mAbs* **2011**, *3*, 568. [[CrossRef](#)] [[PubMed](#)]
6. Shi, S. Biologics: An Update and Challenge of Their Pharmacokinetics. *Curr. Drug Metab.* **2014**, *15*, 271–290. [[CrossRef](#)]
7. Frokjaer, S.; Otzen, D.E. Protein drug stability: A formulation challenge. *Nat. Rev. Drug Discov.* **2005**, *4*, 298–306. [[CrossRef](#)]
8. Mahler, H.C.; Friess, W.; Grauschopf, U.; Kiese, S. Protein aggregation: Pathways, induction factors and analysis. *J. Pharm. Sci.* **2009**, *98*, 2909–2934. [[CrossRef](#)]
9. Lipiäinen, T.; Peltoniemi, M.; Sarkhel, S.; Yrjönen, T.; Vuorela, H.; Urtti, A.; Juppo, A. Formulation and Stability of Cytokine Therapeutics. *J. Pharm. Sci.* **2015**, *104*, 307–326. [[CrossRef](#)]
10. Nejadnik, M.R.; Randolph, T.W.; Volkin, D.B.; Schöneich, C.; Carpenter, J.F.; Crommelin, D.J.A.; Jiskoot, W. Postproduction Handling and Administration of Protein Pharmaceuticals and Potential Instability Issues. *J. Pharm. Sci.* **2018**, *107*, 2013–2019. [[CrossRef](#)]
11. Rathore, A.S. Roadmap for implementation of quality by design (QbD) for biotechnology products. *Trends Biotechnol.* **2009**, *27*, 546–553. [[CrossRef](#)]
12. Den Engelsman, J.; Garidel, P.; Smulders, R.; Koll, H.; Smith, B.; Bassarab, S.; Seidl, A.; Hainzl, O.; Jiskoot, W. Strategies for the assessment of protein aggregates in pharmaceutical biotech product development. *Pharm. Res.* **2011**, *28*, 920–933. [[CrossRef](#)] [[PubMed](#)]
13. Hawe, A.; Romeijn, S.; Filipe, V.; Jiskoot, W. Asymmetrical flow field-flow fractionation method for the analysis of submicron protein aggregates. *J. Pharm. Sci.* **2012**, *101*, 4129–4139. [[CrossRef](#)] [[PubMed](#)]
14. Lebowitz, J.; Lewis, M.S.; Schuck, P. Modern analytical ultracentrifugation in protein science: A tutorial review. *Protein Sci. A Publ. Protein Soc.* **2002**, *11*, 2067. [[CrossRef](#)]
15. Esmonde-White, K.A.; Cuellar, M.; Lewis, I.R. The role of Raman spectroscopy in biopharmaceuticals from development to manufacturing. *Anal. Bioanal. Chem.* **2021**, *414*, 969–991. [[CrossRef](#)] [[PubMed](#)]
16. Oshinbolu, S.; Shah, R.; Finka, G.; Molloy, M.; Uden, M.; Bracewell, D.G. Evaluation of fluorescent dyes to measure protein aggregation within mammalian cell culture supernatants. *J. Chem. Technol. Biotechnol.* **2018**, *93*, 909–917. [[CrossRef](#)]
17. Kögler, M.; Itkonen, J.; Viitala, T.; Casteleijn, M.G. Assessment of recombinant protein production in *E. coli* with Time-Gated Surface Enhanced Raman Spectroscopy (TG-SERS). *Sci. Rep.* **2020**, *10*, 2472. [[CrossRef](#)] [[PubMed](#)]
18. Varmuza, K.; Filzmoser, P. *Introduction to Multivariate Statistical Analysis in Chemometrics*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2009.
19. Korenius, T.; Laurikkala, J.; Juhola, M. On principal component analysis, cosine and Euclidean measures in information retrieval. *Inf. Sci.* **2007**, *177*, 4893–4905. [[CrossRef](#)]
20. Gautam, R.; Vanga, S.; Ariese, F.; Umaphathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2015**, *2*, 8. [[CrossRef](#)]
21. Abdi, H.; Williams, L.J. Principal component analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459. [[CrossRef](#)]
22. Bonnier, F.; Byrne, H.J. Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems. *Analyst* **2012**, *137*, 322–332. [[CrossRef](#)] [[PubMed](#)]
23. Li, J.; Li, J.; Qin, J.; Zeng, H.; Wang, K.; Wang, D.; Wang, S. Confocal Raman microspectroscopic analysis on the time-dependent impact of DAPT, a  $\gamma$ -secretase inhibitor, to osteosarcoma cells. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *239*, 118372. [[CrossRef](#)] [[PubMed](#)]
24. Itkonen, J.M.; Urtti, A.; Bird, L.E.; Sarkhel, S. Codon optimization and factorial screening for enhanced soluble expression of human ciliary neurotrophic factor in *Escherichia coli*. *BMC Biotechnol.* **2014**, *14*, 92. [[CrossRef](#)] [[PubMed](#)]
25. Richardson, D.; Itkonen, J.; Nievas, J.; Urtti, A.; Casteleijn, M.G. Accelerated pharmaceutical protein development with integrated cell free expression, purification, and bioconjugation. *Sci. Rep.* **2018**, *8*, 11967. [[CrossRef](#)]
26. Williams, J.C.; Zeelen, J.P.; Neubauer, G.; Vriend, G.; Backmann, J.; Michels, P.A.; Lambeir, A.M.; Wierenga, R.K. Structural and mutagenesis studies of leishmania triosephosphate isomerase: A point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. *Protein Eng.* **1999**, *12*, 243–250. [[CrossRef](#)]
27. Itkonen, J.; Annala, A.; Tavakoli, S.; Arango-Gonzalez, B.; Ueffing, M.; Toropainen, E.; Ruponen, M.; Casteleijn, M.G.; Urtti, A. Characterization, Stability, and In Vivo Efficacy Studies of Recombinant Human CNTF and Its Permeation into the Neural Retina in Ex Vivo Organotypic Retinal Explant Culture Models. *Pharmaceutics* **2020**, *12*, 611. [[CrossRef](#)]

28. Kostamovaara, J.; Tenhunen, J.; Kögler, M.; Nissinen, I.; Nissinen, J.; Keränen, P. Fluorescence suppression in Raman spectroscopy using a time-gated CMOS SPAD. *Opt. Express* **2013**, *21*, 31632. [[CrossRef](#)]
29. Lipiäinen, T.; Pessi, J.; Movahedi, P.; Koivistoinen, J.; Kurki, L.; Tenhunen, M.; Yliruusi, J.; Juppo, A.M.; Heikkonen, J.; Pahikkala, T.; et al. Time-Gated Raman Spectroscopy for Quantitative Determination of Solid-State Forms of Fluorescent Pharmaceuticals. *Anal. Chem.* **2018**, *90*, 4832–4839. [[CrossRef](#)]
30. Wen, C.-D.; Mudawar, I. Emissivity characteristics of polished aluminum alloy surfaces and assessment of multispectral radiation thermometry (MRT) emissivity models. *Int. J. Heat Mass Transf.* **2005**, *48*, 1316–1329. [[CrossRef](#)]
31. Woody, R.W.; Tinoco, I. Optical rotation of oriented helices. III. Calculation of the rotatory dispersion and circular dichroism of the alpha-and 310-helix. *J. Chem. Phys.* **1967**, *46*, 4927–4945. [[CrossRef](#)]
32. Majorek, K.A.; Porebski, P.J.; Dayal, A.; Zimmerman, M.D.; Jablonska, K.; Stewart, A.J.; Chruszcz, M.; Minor, W. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Mol. Immunol.* **2012**, *52*, 174–182. [[CrossRef](#)] [[PubMed](#)]
33. Sielecki, A.R.; Fedorov, A.A.; Boodhoo, A.; Andreeva, N.S.; James, M.N.G. Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J. Mol. Biol.* **1990**, *214*, 143–170. [[CrossRef](#)]
34. Rygula, A.; Majzner, K.; Marzec, K.M.; Kaczor, A.; Pilarczyk, M.; Baranska, M. Raman spectroscopy of proteins: A review. *J. Raman Spectrosc.* **2013**, *44*, 1061–1076. [[CrossRef](#)]
35. Charles, A.; Janeway, J.; Travers, P.; Walport, M.; Shlomchik, M.J. *The Structure of A Typical Antibody Molecule*; Garland Science: New York, NY, USA, 2001.
36. Krapp, S.; Mimura, Y.; Jefferis, R.; Huber, R.; Sondermann, P. Structural analysis of human IgG-Fc glycoforms reveals a correlation between glycosylation and structural integrity. *J. Mol. Biol.* **2003**, *325*, 979–989. [[CrossRef](#)]
37. McConnell, A.; Zhang, X.; Macomber, J.; Chau, B.; Sheffer, J.; Rahmanian, S.; Hare, E.; Spasojevic, V.; Horlick, R.; King, D.; et al. A general approach to antibody thermostabilization. *mAbs* **2014**, *6*, 1274–1282. [[CrossRef](#)]
38. Cobb, B.A. The history of IgG glycosylation and where we are now. *Glycobiology* **2021**, *30*, 202–213. [[CrossRef](#)]
39. Crispin, M. Therapeutic potential of deglycosylated antibodies. *Glycobiology* **2013**, *30*, 202–213. [[CrossRef](#)]
40. Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3095–E3103. [[CrossRef](#)]
41. Crichton, R.R. 4—Structural and Molecular Biology for Chemists. In *Biological Inorganic Chemistry*; Crichton, R.R., Ed.; Elsevier: Amsterdam, The Netherlands, 2008; pp. 43–76.
42. Sadat, A.; Joye, I.J. Peak Fitting Applied to Fourier Transform Infrared and Raman Spectroscopic Analysis of Proteins. *Appl. Sci.* **2020**, *10*, 5918. [[CrossRef](#)]
43. Freire, P.T.C.; Barboza, F.M.; Lima, J.A.; Melo, F.E.A.; Filho, J.M. Raman Spectroscopy of Amino Acid Crystals. In *Raman Spectroscopy of Amino Acid Crystals*; IntechOpen: London, UK, 2017. [[CrossRef](#)]
44. Huang, Z.; McWilliams, A.; Lui, H.; McLean, D.I.; Lam, S.; Zeng, H. Near-infrared Raman spectroscopy for optical diagnosis of lung cancer. *Int. J. Cancer* **2003**, *107*, 1047–1052. [[CrossRef](#)]
45. Zhang, X.; Zhou, Q.; Huang, Y.; Li, Z.; Zhang, Z. Contrastive Analysis of the Raman Spectra of Polychlorinated Benzene: Hexachlorobenzene and Benzene. *Sensors* **2011**, *11*, 11510–11515. [[CrossRef](#)] [[PubMed](#)]
46. Brewster, V.L.; Ashton, L.; Goodacre, R. Monitoring the Glycosylation Status of Proteins Using Raman Spectroscopy. *Anal. Chem.* **2011**, *83*, 6074–6081. [[CrossRef](#)] [[PubMed](#)]
47. Stein, P.E.; Leslie, A.G.W.; Finch, J.T.; Carrell, R.W. Crystal structure of uncleaved ovalbumin at 1.95 Å resolution. *J. Mol. Biol.* **1991**, *221*, 941–959. [[CrossRef](#)]
48. Panalytical, M. Proteins Melting Point Characterization Using The Zetasizer Nano System. *Azo Nano* **2019**.
49. Vivian, J.T.; Callis, P.R. Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophys. J.* **2001**, *80*, 2093–2109. [[CrossRef](#)]
50. Friess, W.; Lee, G. Basic thermoanalytical studies of insoluble collagen matrices. *Biomaterials* **1996**, *17*, 2289–2294. [[CrossRef](#)]
51. Casteleijn, M.G. *Towards New Enzymes: Protein Engineering versus Bioinformatic Studies*; University of Oulu: Finland, UK, 2009.
52. Lolis, E.; Petsko, G.A. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5-Å resolution: Implications for catalysis. *Biochemistry* **1990**, *29*, 6619–6625. [[CrossRef](#)]
53. Lambeir, A.-M.; Backmann, J.; Ruiz-Sanz, J.; Filimonov, V.; Nielsen, J.E.; Kursula, I.; Norledge, B.V.; Wierenga, R.K. The ionization of a buried glutamic acid is thermodynamically linked to the stability of *Leishmania mexicana* triose phosphate isomerase. *Eur. J. Biochem.* **2000**, *267*, 2516–2524. [[CrossRef](#)]
54. Alahuhta, M.; Casteleijn, M.G.; Neubauer, P.; Wierenga, R.K. Structural studies show that the A178L mutation in the C-terminal hinge of the catalytic loop-6 of triosephosphate isomerase (TIM) induces a closed-like conformation in dimeric and monomeric TIM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2008**, *64*, 178–188. [[CrossRef](#)]
55. Caron, J.; Rochoy, M.; Gaboriau, L.; Gautier, S. The history of pharmacovigilance. *Therapies* **2016**, *71*, 129–134. [[CrossRef](#)]
56. Nicoud, L.; Owczarz, M.; Arosio, P.; Morbidelli, M. A multiscale view of therapeutic protein aggregation: A colloid science perspective. *Biotechnol. J.* **2015**, *10*, 367–378. [[CrossRef](#)] [[PubMed](#)]
57. Möller, M.; Denicola, A. Protein tryptophan accessibility studied by fluorescence quenching. *Biochem. Mol. Biol. Educ.* **2002**, *30*, 175–178. [[CrossRef](#)]
58. Luykx, D.M.; Casteleijn, M.G.; Jiskoot, W.; Westdijk, J.; Jongen, P.M. Physicochemical studies on the stability of influenza haemagglutinin in vaccine bulk material. *Eur. J. Pharm. Sci. Off. J. Eur. Fed. Pharm. Sci.* **2004**, *23*, 65–75. [[CrossRef](#)] [[PubMed](#)]

59. Jiang, B.; Jain, A.; Lu, Y.; Hoag, S.W. Probing Thermal Stability of Proteins with Temperature Scanning Viscometer. *Mol. Pharm.* **2019**, *16*, 3687–3693. [[CrossRef](#)]
60. Vermeer, A.W.P.; Norde, W. The Thermal Stability of Immunoglobulin: Unfolding and Aggregation of a Multi-Domain Protein. *Biophys. J.* **2000**, *78*, 394–404. [[CrossRef](#)]
61. Zav'yalov, V.; Tishchenko, V. Mechanisms of generation of antibody diversity as a cause for natural selection of homoiothermal animals in the process of evolution. *Scand. J. Immunol.* **1991**, *33*, 755–762. [[CrossRef](#)]
62. Vermeer, A.W.P.; Norde, W.; Amerongen, A.V. The Unfolding/Denaturation of Immunogammaglobulin of Isotype 2b and its Fab and Fc Fragments. *Biophys. J.* **2000**, *79*, 2150–2154. [[CrossRef](#)]
63. Jacobsen, F.; Stevenson, R.; Li, C.; Salimi-Moosavi, H.; Liu, L.; Wen, J.; Luo, Q.; Daris, K.; Buck, L.; Miller, S.; et al. Engineering an IgG Scaffold Lacking Effector Function with Optimized Developability. *J. Biol. Chem.* **2017**, *292*, 1865–1875. [[CrossRef](#)]
64. Kamatari, Y.O.; Dobson, C.M.; Konno, T. Structural dissection of alkaline-denatured pepsin. *Spectroscopy* **2004**, *18*, 227–236. [[CrossRef](#)]
65. Tani, F.; Shirai, N.; Nakanishi, Y.; Yasumoto, K.; Kitabatake, N. Role of the Carbohydrate Chain and Two Phosphate Moieties in the Heat-Induced Aggregation of Hen Ovalbumin. *Biosci. Biotechnol. Biochem.* **2004**, *68*, 2466–2476. [[CrossRef](#)]
66. Benítez-Cardoza, C.G.; Rojo-Domínguez, A.; Hernández-Arana, A. Temperature-Induced Denaturation and Renaturation of Triosephosphate Isomerase from *Saccharomyces cerevisiae*: Evidence of Dimerization Coupled to Refolding of the Thermally Unfolded Protein. *Biochemistry* **2001**, *40*, 9049–9058. [[CrossRef](#)] [[PubMed](#)]
67. Panalytical, M. Dynamic Light Scattering as a Method for Understanding the Colloidal Stability of Protein Therapeutics. *AZoM* **2019**.
68. Berner, C.; Menzen, T.; Winter, G.; Svilenov, H. Combining Unfolding Reversibility Studies and Molecular Dynamics Simulations to Select Aggregation-Resistant Antibodies. *Mol. Pharm.* **2021**, *18*, 2242–2253. [[CrossRef](#)] [[PubMed](#)]
69. De Gelder, J.; De Gussem, K.; Vandenabeele, P.; Moens, L. Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* **2007**, *38*, 1133–1147. [[CrossRef](#)]
70. Gremlich, H.-U.B.Y. *Infrared and Raman Spectroscopy of Biological Materials*; CRC Press: Boca Raton, FL, USA, 2021.
71. Cattell, R.B. The Scree Test for the Number Of Factors. *Multivar. Behav. Res.* **1966**, *1*, 245–276. [[CrossRef](#)]