Journal of Pharmaceutical and Biomedical Analysis 51 (2010) 39–45

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Implementation of an FTIR calibration curve for fast and objective determination of changes in protein secondary structure during formulation development

Sebastian Vonhoff[a], Jamie Condliffe[b], Heiko Schiffter[b,*]

[a] Division of Pharmaceutics, University of Erlangen-Nuremberg, Germany
[b] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Old Road Campus Research Building, Off Roosevelt Drive, Oxford OX3 7DQ, UK

ARTICLE INFO

Article history:
Received 1 May 2009
Received in revised form 22 July 2009
Accepted 27 July 2009
Available online 4 August 2009

Keywords:
FTIR
Secondary structure
Protein formulation
Circular dichroism
Multivariate data analysis
Calibration curve

ABSTRACT

The aim of this study was to develop a quick and objective method for the determination of changes in protein secondary structure by Fourier transform infrared spectroscopy (FTIR). Structural shifts from native regions (α-helix, intramolecular β-sheet) to aggregated strands (intermolecular β-sheet) were used to evaluate protein damage. FTIR spectra of 16 different proteins were recorded and quantified by peak fitting of the non-deconvolved and baseline corrected amide I bands. The resulting percentile secondary structures were correlated with the shape and intensity of the area normalized amide I bands using an interval partial least squares algorithm (iPLS). Structural elements were focused on the following regions: α-helix 1660–1650 cm^{-1}, intramolecular β-sheet 1695–1683 cm^{-1} and 1644–1620 cm^{-1}, intermolecular β-sheet 1620–1595 cm^{-1}. Three calibration curves were created from the data sets. Calculated α-helix content ranged from 0% to 79.59%, intramolecular β-sheet from 10.64% to 63.89% and intermolecular β-sheet from 0.23% to 9.70%. The linear relationship between actual values (as determined by peak fitting) and calculated values was evaluated by correlation coefficient and root mean square error of calibration while cross-validation was performed to detect possible outliers. Results were verified by including two proteins as validation standards and comparing the calculated values to peak fitting and X-ray data.

Structural changes of human serum albumin (HSA) due to elevated temperatures and the fibrillation of glucagon were quantified by calibration curve analysis. Performance and reliability of the iPLS algorithm were evaluated by comparing calculated secondary structure elements with results from peak fitting and circular dichroism. Different methods for the determination of secondary structure gave slightly different results but overall tendencies concurred. Additionally, formation of HSA aggregates could be linked to increasing β-sheet content by comparing SEC–HPLC and turbidity analysis with results from the FTIR calibration curves.

In summary, quantification of the α-helix to β-sheet transition by iPLS analysis proves to be a feasible and objective way for the determination of damage to protein secondary structure.

1. Introduction

A major challenge in the handling of protein pharmaceuticals is their poor in vitro stability in the liquid state [1]. Removal of water generally improves protein stability but can also result in damage to the secondary or tertiary structure as both are dependent on hydrophilic and hydrophobic interactions [2]. Even dried formulations that initially seemed stable can undergo changes during long term storage [3] with negative influence on their shelf-life. Addition of lyo- and cryoprotectants to the formulation has shown to preserve the native structure during the drying process and in the final product [4–6]. FTIR spectroscopy is a well established tool in the determination of protein secondary structure [7,8] and can provide valuable information in the development process of ideal formulations and process parameters.

Fig. 1 shows the typical FTIR spectrum of α-chymotrypsin and its amide bands containing information on secondary structure. The amide I band (1700–1600 cm^{-1}) is often used in literature as it is based on only few molecule vibrations and sufficient information for cross-referencing is available [5,9,10]. As the individual components within the amide bands overlap severely, interpretation of the spectra is rather difficult. Over the years, different approaches for evaluating FTIR protein spectra have been developed and each of them offers different advantages and disadvantages. Peak fitting, for example, gives detailed information about secondary structure,

* Corresponding author at: Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Old Road Campus Research Building, Off Roosevelt Drive, Oxford, OX3 7DQ, United Kingdom. Tel.: +44 01865 617711; fax: +44 01865 617701.
E-mail addresses: vonhoff@pharmtech.uni-erlangen.de (S. Vonhoff), jamie.condliffe@gmail.com (J. Condliffe), heiko.schiffter@eng.ox.ac.uk, heiko.schiffter@googlemail.com (H. Schiffter).

0731-7085/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2009.07.031

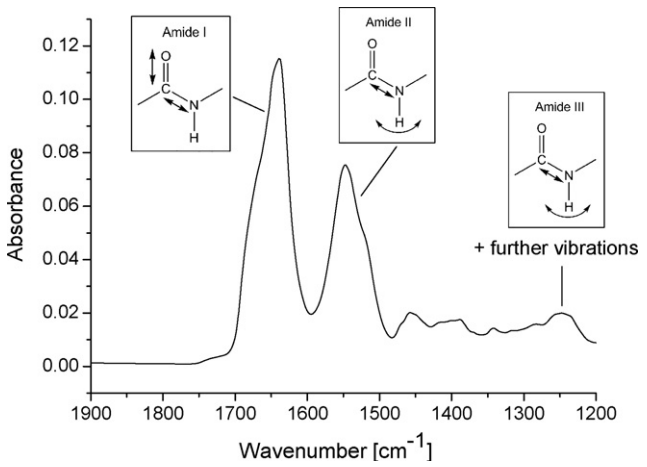

Fig. 1. Sample spectrum of α-chymotrypsin showing the three relevant regions for determination of protein secondary structure. Amide I (1700–1600 cm^{-1}), amide II (1600–1500 cm^{-1}) and amide III (1330–1220 cm^{-1}).

but suffers from a high degree of subjectivity due to Fourier self-deconvolution and band assignment [11]. Analysis by correlation coefficient of the second derivative spectra is relatively straightforward, but does not necessarily lead to correct results [12] and determination of area overlap of the raw spectra only gives information about the overall similarity of two spectra. An approach for quick and objective quantification of structural contents is multivariate data analysis [13]. Spectra are taken from a set of proteins with known secondary structure. During a calibration step, spectral absorbencies are correlated with the secondary structure contents by a matrix of absorptivity constants "*K*". The structural content of an unknown spectrum can then be quantified with help of *K*. The techniques most commonly used are singular value decomposition (SVD) and partial least squares (PLS).

SVD is a combination of inverse least squares and principal component analysis. Unlike quantification methods that are based directly on Beer's Law, SVD regresses the structural contents from variance spectra ("eigenvectors") and its loading factors ("scores"). This means that with SVD the absorbance matrix is decomposed during a first step into its eigenvectors and scores and in a second step an inverse least squares algorithm is applied to calculate the absorptivity constants.

PLS is a quantitative decomposition technique that is closely related to SVD. The main difference between both algorithms is that PLS decomposes both the spectral data and the structural contents giving two sets of vectors and scores. As the spectral information and the secondary structures are connected, the two sets of scores can be related to each other by regression and a calibration model can be constructed. Spectral decomposition and regression are performed in one step.

Goal of this study was to create a method suitable for quick and objective determination of damage to secondary structure during protein processing.

2. Materials and methods

2.1. Materials

All proteins were purchased from Sigma–Aldrich (Gillingham, Dorset, UK) as lyophilized powders or, in case of carboxypeptidase A, as suspension in toluene. Protein solutions were prepared in double-distilled and filtered (0.2 μm) water without any excipients. To achieve sufficient solubility glucagon solutions were prepared using diluted acetic acid (1% m/m) while Insulin was dissolved in 0.01 M HCl at pH 2. Carboxypeptidase A was dissolved by adding 1 M NaCl until the desired enzyme concentration was reached. Each sample was centrifuged at 10,000 rpm for 10 min and the supernatant was used for analysis in order to prevent clogging of the transmission cell by undissolved protein particles.

2.2. Turbidity analysis

Turbidity analysis of the heat treated HSA solutions was performed at 350 nm using a Lambda LS 25 spectrophotometer (PerkinElmer, MA, USA) connected to a waterbath at 25 °C. The transmittance baseline was set using a quartz cuvette containing double distilled and filtered (0.2 μm) water. Each measurement was repeated three times, and mean and relative standard deviation (rel. sdv.) were calculated.

2.3. SEC–HPLC

The amount of soluble protein aggregates was determined by size exclusion chromatography. The HPLC system consisted of a Prostar 210 solvent delivery systems with two pumps, a Prostar 500 column valve module, a Prostar 325 UV detector and a Model 410 autosampler (Varian Inc., Yarnton, UK). Size exclusion chromatography was performed on a 4.6 mm × 250 mm Zorbax GF 250 column (Agilent, West Lothian, UK) with a nominal particle size of 4.0–4.5 μm and a pore diameter of 150 Å. A 20 mM phosphate buffer at pH 6.8 with 250 mM NaCl was chosen as mobile phase. The flow rate was constant at 0.250 ml/min under isocratic conditions and the injection volume was 10 μl of a 20 mg/ml BSA solution. The column temperature was 30 °C and UV detection of the eluent was carried out at 280 nm. Peak identification and integration was performed using Galaxie chromatography software (Varian Inc, Yarnton, UK).

2.4. Circular dichroism

CD spectra were recorded on a Jasco J-710 spectropolarimeter (Jasco Deutschland GmbH, Germany) constantly purged with nitrogen. Before usage, calibration of the dichrometer was verified using d-10-camphersulfonic acid. HSA samples were prepared in the same way as for FTIR analysis and recorded at a scan speed of 0.20 nm/s using the Proteus transmission cell described under "FTIR spectroscopy instrumentation". As wavelengths below 185 resulted in poor signal-to-noise ratios, spectra ranged between 185 and 260 nm. Quantification was performed by the DICHROWEB server [14] using the SELCON3 analysis program [15] and SP175 protein reference set from 190 to 240 nm [16]. For comparison reasons with FTIR data the numbers for regular and distorted α-helix or β-sheet content were summed up. Validity of evaluation was determined by normalized root mean square deviation (NRMSD) between the measured and calculated data, which is given by

$$\mathrm{NMRSD} = \left[\frac{\sum_{N} (\theta_{exptl} - \theta_{calcd})^2}{\sum_{N} (\theta_{exptl})^2} \right]^{1/2} \quad (1)$$

where θ_{exptl} and θ_{calcd} are the experimental and calculated mean residue ellipticities and *N* is the number of data points used.

2.5. FTIR spectroscopy instrumentation

Protein spectra were recorded on an Omnic Nicolet MagnaIR 550 spectrometer (Thermo Fisher Scientific Inc., Germany) using a Proteus CaF_2 transmission cell, optical path length 6 μm (Thermo Fisher Scientific Inc., Germany). For each measurement a total of 42 scans was collected at slowest mirror speeds possible to minimize

spectral noise. The measurements were repeated three times and averaged to reduce baseline effects. Between measurements the cell was purged with water until no protein signal was detectable. Subtraction of the background signal was performed manually with a subtraction factor usually between 0.95 and 1.05 until a flat baseline between 1900 and 1740 cm^{-1} was achieved [17].

2.6. Data analysis

There are numerous examples available in the literature that describe the mathematical methods used in this paper [18–20]. Here we supply the basic formulae that describe the most important steps of the calculations. It is possible to relate known concentrations, *C* (defined as the occurrence of certain groups or structural types), to measured absorbencies, *A*, by a calibration step. This calibration can then be used to determine concentrations for newly measured spectra.

A and **C** can be related, in matrix form, using

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R} = \boldsymbol{C}^{\mathrm{T}}. \quad (2)$$

where **R** is the calibration matrix and T indicates the transpose of a matrix [21]. A solution which minimizes the squared error may be sought by differentiating Eq. (2), which, after a little manipulation, yields a solution for **R** which takes the form

$$\mathbf{R} = (\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}}. \quad (3)$$

This allows us to evaluate **R**, which may be used to calculate the concentrations of new spectra, **c**, from a measured set of absorbencies, **a**, such that

$$\mathbf{c} = \mathbf{R}^{\mathrm{T}}\mathbf{a}. \quad (4)$$

Unfortunately, the quantity $(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{A}$ from Eq. (3) is computationally inefficient to calculate. A number of algorithms exist which allow us to accurately approximate the values of $(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}\mathbf{A}$ and these have been used extensively to calculate structure concentrations in the past [13,22,23]. Two algorithms of interest are presented here in summary. The reader may wish to refer to the theoretical comparison of the two by Lorber et al. for a more detailed description [24].

Singular value decomposition is a firmly established technique which is used to improve the efficiency of matrix calculations. Here, **A** is decomposed into three matrices and rewritten as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}, \quad (5)$$

where **U** and **V** are orthogonal and **S** is diagonal, containing the singular values s_{ii}. Calculations are simplified further by removing vectors corresponding to small singular values from **U**, **S** and **V**, reducing their rank and leading to

$$\boldsymbol{A} \approx \boldsymbol{U}_1\boldsymbol{S}_1\boldsymbol{V}_1^{\mathrm{T}}. \quad (6)$$

This approximation to **A** can be substituted into Eq. (2), providing a new estimate for **R** which can be written

$$\boldsymbol{R}^{\mathrm{T}} = \boldsymbol{U}_1\boldsymbol{S}_1^{-1}\boldsymbol{V}_1^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}. \quad (7)$$

Partial least squares regression bears strong resemblance to singular value decomposition, but is often more robust, especially when a high degree of multicollinearity exists between elements of the **C** matrix. The matrices **A** and **C** are redefined as $\mathbf{A} = \mathbf{T}\mathbf{P}^{\mathrm{T}}$ and $\mathbf{C}^{\mathrm{T}} = \mathbf{U}\mathbf{Q}^{\mathrm{T}}$. These newly defined matrices are then decomposed in such a way that $\mathbf{U} = \mathbf{T}\mathbf{V}$ connects both decompositions. The partial least squares algorithm is employed throughout the work presented here. The software packages: PeakFit (Systat Software GmbH, Erkrath, Germany) and TQ Analyst (Thermo Fisher Scientific Inc., Germany) were used for all the calculations presented.

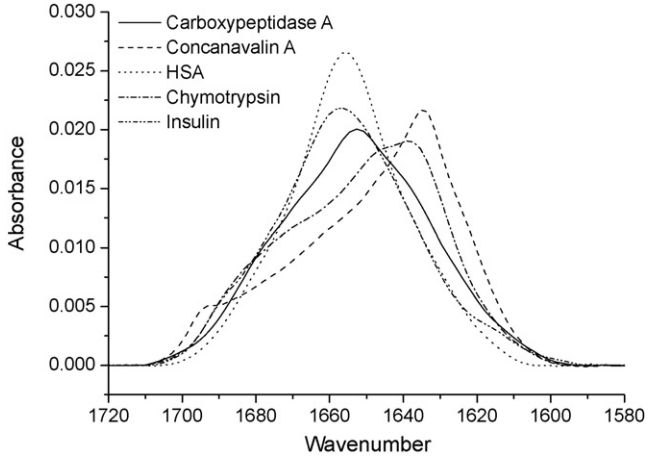

Fig. 2. Area normalized, baseline corrected and non-deconvolved amide I spectra of five proteins used for creating the calibration curves for α-helix, intramolecular β-sheet and intermolecular β-sheet.

3. Results and discussion

3.1. Calibration curve

Protein secondary structure was analyzed using the non-deconvolved, baseline corrected and area normalized amide I band of the spectrum (Fig. 2). Instead of assigning structures from X-ray measurements to the FTIR spectra during calibration [13,18], secondary structures for the content matrix **C** were determined by peak fitting as the crystal structure of proteins is not necessarily retained in aqueous solutions [25]. Furthermore, possible structural changes due to prior lyophilization by the manufacturer, pH-shifts or experimental conditions could be taken into account. As peak fitting is a rather subjective quantification method, results were compared with X-ray data by Levitt and Greer [26] to assure peak assignment and iteration procedures were performed reasonably. The first step during peak fitting consisted of qualitatively detecting the individual peaks adding up to the amide I band with help of 2nd derivative spectra. According to literature [8], peak positions were assigned to individual structural components. With the help of PeakFit, quantitative evaluation of the secondary structure was performed assuming a Gaussian peak shape. As no satisfactory peak fit could be obtained for α-chymotrypsin, X-ray data was used in this one case.

The same spectra that were used for peak fitting were then implemented as standards for the calibration curve. The shapes and intensities of the area normalized amide I bands were correlated with their percentile secondary structures using an interval partial least squares 1 algorithm [27]. During this step, critical shapes ("factors") within the standards were detected and pure component spectra were computed, each one representing only one specific secondary structure (Fig. 3). From this point on, the evaluation method was calibrated and unknown samples, as well as the standards themselves, could be quantified. As denaturation is mostly linked to a shift from native (α-helix or intramolecular β-sheet) to aggregated structures (intermolecular β-sheet) [5,28], conformations like β-turns and random coil were not included in the method.

If an excessive number of factors is used for describing a certain secondary structure, too much information is introduced into the model and overall quantification performance can be reduced [29]. To avoid effects due to overfitting, the structural contents of each standard were calculated by cross-validation using one, two, three or four factors describing the quantification models. Then, the number of factors giving the lowest predicted residual error sum

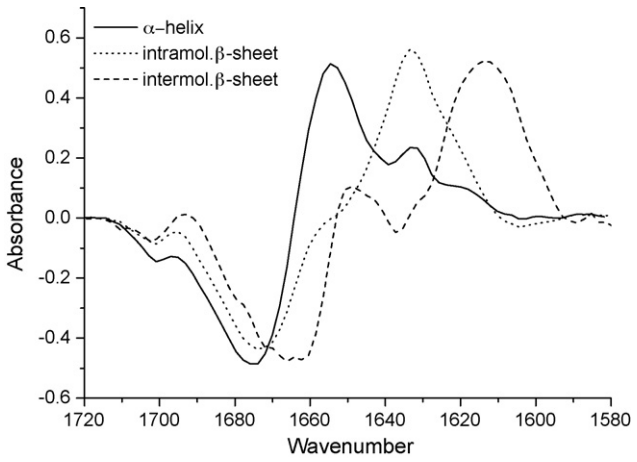

Fig. 3. Pure component spectra as derived from the standards. Within the defined regions, shapes that were determined as critical by iPLS were used to build up spectra representing one specific secondary structure. By combining all three spectra and weighting them individually, unknown samples can be quantified.

of squares ("PRESS") with $\sum(y_{i\,exp} - y_{i\,pred})^2$ was chosen for calibration, where $y_{i\,exp}$ equals the actual concentrations and $y_{i\,pred}$ the calculated concentrations.

Analysis was focused on the following areas to avoid regions containing noise or irrelevant data for the prediction of the individual secondary structures: α-helix 1660–1650 cm^{-1}, intramolecular β-sheet 1695–1683 cm^{-1} and 1644–1620 cm^{-1}, intermolecular β-sheet 1620–1595 cm^{-1}. A mean centering technique was applied to the data set before calibration.

To evaluate linear relationship, three calibration curves were created from the data sets, one for each type of secondary structure investigated (α-helix, intramolecular β-sheet and intermolecular β-sheet). Secondary structures calculated by multivariate data analysis were plotted against structural contents as determined by peak fitting ("actual") (Fig. 4).

As shown in Table 1, calculated α-helix content ranged from 0% (concanavalin A) to 79.59% (haemoglobin). Insulin showed highest deviation from actual α-helical values (56.14% actual vs. 51.23% calculated). Structural contents of α-chymotrypsin could not be specified by peak fitting due to problems during peak assignment and thus were replaced by values determined by X-ray analysis. Values calculated for intramolecular β-sheet ranged from 10.64% (haemoglobin) to 63.89% (concanavalin A). No calculated result was more than 3.69% away from its actual value (e.g. concanavalin A). The data set for intermolecular β-sheet only ranged from 0.23% (glucagon) to 9.70% (concanavalin A). Even though most protein standards were pre-lyophilized by the supplier, no extensive damage to secondary structure was induced which explains the low intermolecular β-sheet contents. As values for α-chymotrypsin were taken from X-ray analysis, intermolecular β-sheet could not be distinguished from intramolecular β-sheet and was ignored for calibration. As suggested by PRESS, two factors were used during calibration for α-helix and intramolecular β-sheet, while one factor was already sufficient for intermolecular β-sheet resulting in values of 159.80, 127.92 and 4.91, respectively. The low numbers for intermolecular β-sheet can be explained by its narrow percentile span (0.23–9.70%) compared to the other structural components.

The correlation coefficient r (Eq. (8)) was used to quantify the linear relationship between actual and calculated values:

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (8)$$

where n equals the number of standards and x_i and y_i the actual and calculated values. Results lay between 0.992 for α-helix and intramolecular β-sheet and 0.979 for intermolecular β-sheet.

The uncertainty of the calibration model was evaluated using the root mean square error (RMSE) of the calibration curve, which

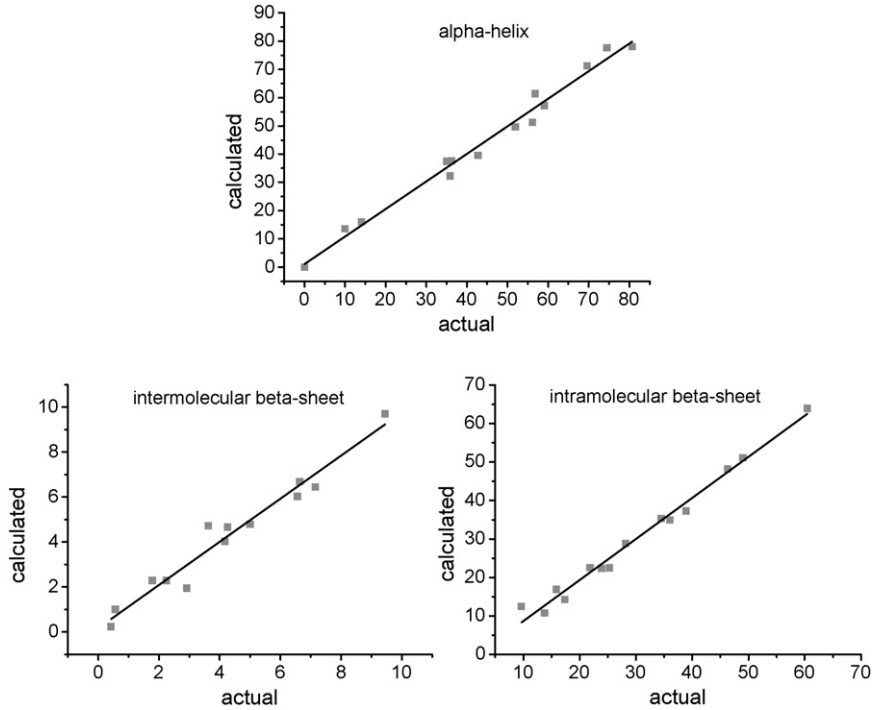

Fig. 4. Calibration curves for α-helix, intramolecular β-sheet and intermolecular β-sheet. Values determined by peak fitting ("actual") were correlated with calculated values by iPLS.

Table 1
Comparison of secondary structure by peak fitting (actual) with calculated values (calculated).

Protein	α-Helix (actual)	α-Helix (calculated)	Intramol. β-sheet (actual)	Intramol. β-sheet (calculated)	Intermol. β-sheet (actual)	Intermol. β-sheet (calculated)
Alkaline phosphatase	36.21%	37.68%	34.49%	35.15%	4.26%	4.66%
Bovine serum albumin	69.64%	71.19%	17.39%	13.95%	1.78%	2.28%
Carbonic anhydrase	14.05%	15.99%	46.29%	47.59%	7.15%	6.44%
Carboxypeptidase A	35.04%	37.45%	28.18%	27.74%	6.56%	6.03%
Catalase	35.86%	32.32%	38.91%	36.55%	6.64%	6.69%
Concanavalin A	0.00%	0.00%	60.45%	63.89%	9.45%	9.70%
α-Chymotrypsin	10.00%	13.61%	49.00%	52.47%	n/a	5.59%
β-Galactosidase	42.78%	39.51%	36.05%	35.47%	3.63%	4.72%
Glucagons	59.12%	57.15%	15.88%	17.05%	0.42%	0.23%
Haemoglobin	74.68%	79.59%	10.63%	10.64%	2.79%	2.09%
Human serum albumin	74.48%	77.62%	13.81%	10.76%	0.57%	1.00%
Insulin	56.14%	51.23%	23.92%	22.46%	5.00%	4.79%
Lactate dehydrogenase	56.81%	61.39%	25.32%	23.46%	2.92%	1.94%
Lysozyme	51.92%	49.57%	21.84%	21.84%	4.17%	4.03%
Myoglobin	80.71%	78.04%	9.68%	12.56%	2.25%	2.28%
Ribonuclease A	24.78%	23.39%	40.26%	42.09%	2.92%	3.65%

is given as

$$\mathrm{RMSE} = \sqrt{\frac{\sum (x_i - y_i)^2}{n}} \quad (9)$$

where n equals the number of standards and x_i and y_i the actual and calculated values. Values for α-helix, intramolecular β-sheet and intermolecular β-sheet were 2.95, 2.24 and 0.53, respectively.

Validity of results was evaluated by including two more standards for validation purposes: haemoglobin for highly helical and ribonuclease A for β-sheet rich proteins. Secondary structures were again determined by peak fitting and compared to the calculated values. However, validation standards are not included in the calibration set and do not affect its compilation. A performance index, calculated as the RMSE, for the validation standards is calculated giving information on how well unknown samples are quantified. Results were 0.986, 0.144 and 0.495 for α-helix, intramolecular β-sheet and intermolecular β-sheet, respectively indicating good quantification performance.

To minimize any bias by outliers, a cross-validation was performed. Each calibration standard was quantified as if it were a validation standard. This was accomplished by sequentially removing one standard from the calibration set, calibrating the method and using the new calibration model to quantify the excluded standard. The software repeated the process until all the calibration standards were quantified as validation standards. Correlation coefficients r for α-helix, intramolecular β-sheet and intermolecular β-sheet were 0.990, 0.982 and 0.972 while RMSE of cross-validation was 3.38, 3.02 and 0.615, respectively.

Comparing the calculated results with literature showed good agreement for most of the included proteins (Table 2) [26]. Only β-sheet content of myoglobin and haemoglobin were significantly higher than values determined by X-ray (by 12% and 10%, respectively). As both proteins exhibit a combination of very high α-helix and low β-sheet content, results could be biased by the strong absorbance between 1660 and 1650 cm^{-1} influencing the adjacent region actually assigned to β-sheet.

Table 2
Comparison of secondary structures as found by Levitt and Greer [30] with calculated data (this work).

Protein	α-Helix (X-ray)	α-Helix (FTIR)	β-Sheet (X-ray)	Intramol. β-sheet (FTIR)
Carbonic anhydrase	16%	16%	45%	48%
Carboxypeptidase A	39%	38%	30%	29%
Concanavalin A	3%	0%	60%	64%
α-Chymotrypsin	10%	14%	49%	51%
Haemoglobin	86%	80%	0%	10%
Insulin	53%	51%	15%	22%
Lysozyme	46%	50%	19%	22%
Myoglobin	85%	78%	0%	12%
Ribonuclease A	23%	23%	46%	42%

3.2. Applicability and verification of results

To evaluate the applicability and precision of the iPLS algorithm, changes in the secondary structure of glucagon and HSA were quantified using the calibration curves. Results were compared with secondary structure analysis performed by CD, peak fitting and, in case of HSA, size exclusion chromatography as well as turbidity analysis.

Denaturation temperature of HSA lies between 60 and 65 °C depending on the pH [30]. Measurement of a freshly dissolved solution indicated a pH of 7.05 at 25 °C and therefore a denaturation temperature of 65 °C was expected. HSA solutions with a concentration of 20 mg/ml were exposed to temperatures of 25, 40, 60, 70 and 80 °C for 30 min in a waterbath. After cooling down to room temperature, the samples were analyzed. As seen in Table 3, native structure of HSA underwent only minor changes until 60 °C. In comparison to the sample kept at 25 °C, helical structures were reduced by 5.71% while intra- and intermolecular β-sheets gained 3.2% and 0.84%, respectively. Any further increase of temperature led to a strong decrease of α-helix (down to 36.69%) and to an increase in intramolecular and intermolecular β-sheet (up to 29.52% and

Table 3
Changes in secondary structure of HSA due to elevated temperatures.

Temperature	Method	α-Helix	Intramol. β-sheet	Intermol. β-sheet
25 °C	iPLS	77.62%	10.76%	1.00%
	Peak Fit	76.23%	12.58%	1.18%
	CD	67.1%		3.5%
40 °C	iPLS	74.74%	12.50%	1.50%
	Peak Fit	75.51%	13.19%	1.14%
	CD	64.1%		4.5%
60 °C	iPLS	71.91%	13.96%	1.84%
	Peak Fit	72.59%	13.52%	2.03%
	CD	62.1%		5.2%
70 °C	iPLS	54.14%	22.05%	6.38%
	Peak Fit	59.83%	20.04%	6.25%
	CD	51.5%		12.5%
80 °C	iPLS	36.69%	29.52%	11.52%
	Peak Fit	45.61%	27.96%	11.36%
	CD	41.3%		22.4%

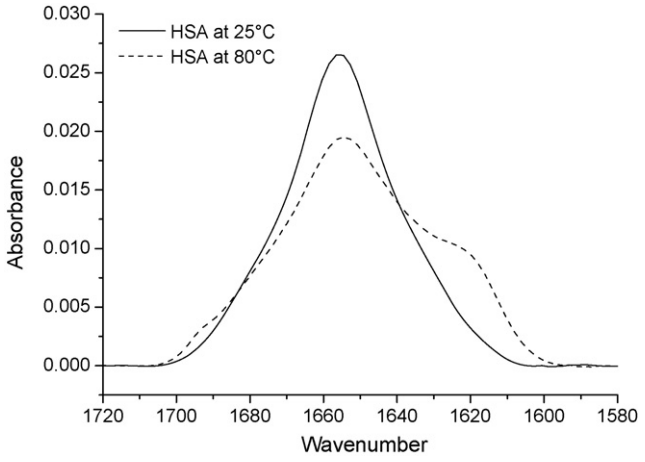

Fig. 5. Changes of the amide I band of human serum albumin after heat treatment. The spectrum underwent significant changes showing a decrease in height and spectral broadening. The shoulder at 1618 cm^{-1} correlates with intermolecular β-sheet due to newly formed aggregates.

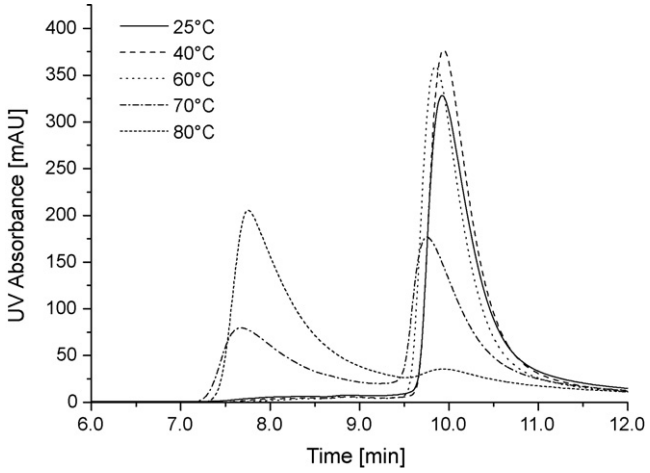

Fig. 6. SEC–HPLC diagram of heat treated HSA. Monomer content (detectable after 10 min) stays constant until 60 °C with no aggregate peak visible. For 70 and 80 °C noticeable aggregate formation is visible (signal after 7–8 min).

11.52%) supporting the DSC results by Barone [30]. Peak fitting was applied to verify results obtained by calibration curve using exactly the same amide I bands. For 25–60 °C results were comparable between the two techniques only showing fluctuations within 2%. For higher temperatures differences increased up to 8.92% for α-helix at 80 °C. The trend to lower helical values by the calibration curve at 70 and 80 °C can be explained by two points. Firstly, the spectral region for α-helix is only 10 wavenumbers wide (from 1660 to 1650 cm^{-1}) leaving a narrow window for quantification. As denaturation of HSA is linked with an overall broadening of the amide I band (Fig. 5) spectral features related to α-helix could shift beyond the limits set by calibration thereby avoiding detection. Second, the shape of the amide I band changes significantly during heating matching no standard defined during calibration. This could increase the error during quantification.

For further validation, data obtained by FTIR spectroscopy was additionally compared to CD measurements (see Table 3). As evaluation by CD did not discern between intramolecular and intermolecular conformations, only one value is available for β-sheet. Freshly dissolved HSA showed a secondary structure of mainly α-helix and only very little β-sheet. In analogy to FTIR, only minor changes occurred until 60 °C resulting in a small decline in α-helix by 5.0% and slight increase in β-sheet content by 1.7%. Increasing temperatures to 80 °C inflicted strong protein damage. Helical content was reduced by 25.8% and β-sheets increased by 18.9%. NRMSD of all evaluations stayed below 0.1 indicating good fitting except for HSA at 80 °C (0.146) which was still considered acceptable. In comparison to FTIR, results from CD gave lower absolute values for helical (up to 10.52% at 25 °C) and β-sheet content (up to 18.64% at 80 °C). This is not surprising as deviancies between both techniques have been previously reported in literature [9]. In our case, the most plausible explanation is that different references were used for calibration of the datasets. The SP175 dataset was created from X-ray structure using the DSSP algorithm [16] while our FTIR dataset was created by peak fitting of infrared spectra and validated with X-ray data created by Levitt and Greer [26]. However, both methods show that about 72% of the loss in α-helix can be associated with an increase in β-sheets during denaturation. This finding supports our approach to quantify protein damage by measuring the transition from native structures to intermolecular (and partially intramolecular) β-sheet.

HSA is known to form soluble aggregates during storage at elevated temperatures. Solubility decreases with increasing size finally leading to protein precipitation [31]. The formation of soluble aggregates during heating was detected by SEC–HPLC and compared to the increase in intermolecular β-sheet detected by FTIR spectroscopy. Until 60 °C only one monomer peak was detectable in the chromatogram after 10 min. Increasing the temperature to 70 °C decreased the height of the monomer peak and gave rise to a new signal after 8 min assigned to aggregated protein molecules. At 80 °C aggregates increased even more at the expense of monomers (Fig. 6). Turbidity analysis for detection of soluble and insoluble aggregates supported the results determined by HPLC. For 25, 40 and 60 °C transmission lay between 0.730 (rel. sdv. = 0.21%) and 0.721 (rel. sdv. = 0.08%). At 70 and 80 °C transmission showed a strong decrease due to protein aggregation down to 0.659 (rel. sdv. = 0.26%) and 0.646 (rel. sdv. = 0.39%), respectively.

Glucagon is a 29-amino acid polypeptide that is known to fibrillate in solution at high concentrations and acidic pH [32]. Depending on protein concentration and polarity of the solvent freshly dissolved glucagon can form a predominately helical or random coil structure that transforms completely into β-sheets during fibrillation [33,34]. For fibrillation analysis glucagon solutions (7 mg/ml) were kept at room temperature and stirred constantly. Samples were taken at 1 h intervals and subsequently analyzed after centrifugation. Immediately after dissolving in 1% AcOH, the FTIR spectrum of glucagon showed a maximum at 1656 cm^{-1} typical for high α-helix content. Within 4 h the spectrum underwent substantial changes giving rise to two new peaks at 1630 and 1614 cm^{-1} assigned to intra- and inter-molecular β-sheets (Fig. 7). Quantification by peak fitting showed that native α-helix dropped quickly from 64.82% to 0% after only 4 h. In the same time intramolecular β-sheet increased from 14.36% to 34.99% and intermolecular β-sheet 0.42% to 16.92%. The remaining 27.69% of the former α-helix were split up between an increase of unordered conformations and β-turns. Quantification by iPLS showed even higher α–β transition. α-Helix content dropped from 60.57% to 0% within 4 h while intramolecular and intermolecular β-sheets rose from 15.88% to 40.53% and 0.33% to 21.96%, respectively. As with HSA, 76% of the loss in helical structures can be attributed to the formation of β-sheets. These results are in good agreement with the CD results from Moran [33] claiming that glucagon at high concentrations assumes a predominantly α-helical structure (48%) that devolves into β-sheet (54%) during aging. As no sufficient signal-to-noise ratio could be obtained from our CD setup, the FTIR results could only be compared to literature.

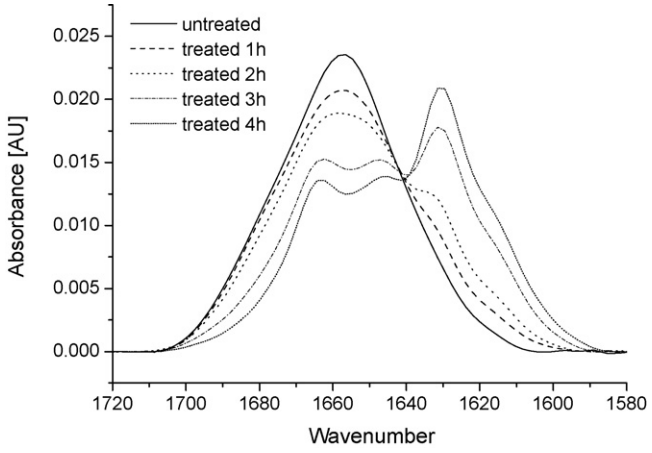

Fig. 7. FTIR spectra of glucagon. During fibrillation the maximum for α-helix (1656 cm^{-1}) completely vanishes. Two new peaks are formed at 1630 cm^{-1} and 1614 cm^{-1} typical for intra- and inter-molecular β-sheet.

4. Conclusion

Evaluation of protein secondary structure by iPLS database analysis leads to objective and reproducible results. Quantification is performed quickly as the only post-processing steps consist of baseline correction and area normalization. Statistical evaluation of the calibration curve indicated close correlation between calculated values and values determined by peak fitting. Validity of the calibration curve is supported by X-ray analysis which gave similar results. Results by CD showed some differences compared to FTIR. However, differences between both techniques have been reported earlier in literature and can be explained by different data sets during calibration. Overall tendencies still concur and thereby support our results. SEC–HPLC and turbidity analysis showed that the formation of HSA aggregates is linked to the formation of inter- and intramolecular β-sheet. This means that shifts from native structures to β-sheet should be usable for quantifying damage to protein secondary structure and possible aggregate formation. By extending the calibration curve to more proteins structural prediction could be improved even more.

References

[1] W. Wang, Instability, stabilization, and formulation of liquid protein pharmaceuticals, Int. J. Pharm. 185 (1999) 129–188.
[2] W. Wang, Lyophilization and development of solid protein pharmaceuticals, Int. J. Pharm. 203 (2000) 1–60.
[3] S. Yoshioka, Y. Aso, Application of accelerated testing to shelf-life prediction of commercial protein preparations, J. Pharm. Sci. 83 (1994) 454–456.
[4] H.R. Costantino, K.G. Carrasquillo, Effect of excipients on the stability and structure of lyophilized recombinant human growth hormone, J. Pharm. Sci. 87 (1998) 1412–1420.
[5] J.F. Carpenter, S.J. Prestrelski, Application of infrared spectroscopy to development of stable lyophilized protein formulations, Eur. J. Pharm. Biopharm. 45 (1998) 231–238.
[6] B.S. Wang, S. Tchessalov, Impact of sucrose level on storage stability of proteins in freeze-dried solids: I. Correlation of protein-sugar interaction with native structure preservation, J. Pharm. Sci. (2008), published online.
[7] J.L. Arrondo, A. Muga, Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy, Prog. Biophys. Mol. Biol. 59 (1993) 23–56.
[8] A. Barth, C. Zscherp, What vibrations tell us about proteins, Q. Rev. Biophys. 35 (2002) 369–430.
[9] D.M. Byler, H. Susi, Examination of the secondary structure of proteins by deconvolved FTIR spectra, Biopolymers 25 (1986) 469–487.
[10] A. Dong, S.J. Prestrelski, Infrared spectroscopic studies of lyophilization- and temperature-induced protein aggregation, J. Pharm. Sci. 84 (1995) 415–424.
[11] M. Jackson, H.H. Mantsch, The use and misuse of FTIR spectroscopy in the determination of protein structure, Crit. Rev. Biochem. Mol. Biol. 30 (1995) 95–120.
[12] B.S. Kendrick, A. Dong, Quantitation of the area of overlap between second-derivative amide I infrared spectra to determine the structural similarity of a protein in different states, J. Pharm. Sci. 85 (1996) 155–158.
[13] K. Rahmelow, W. Hubner, Secondary structure determination of proteins in aqueous solution by infrared spectroscopy: a comparison of multivariate data analysis methods, Anal. Biochem. 241 (1996) 5–13.
[14] L. Whitmore, Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, Biopolymers 89 (2008) 392–400.
[15] N. Sreerema, Estimation of the number of helical and strand segments in proteins using CD spectroscopy, Protein Sci. 8 (1999) 370–380.
[16] J.G. Lees, A reference database for circular dichroism spectroscopy covering fold and secondary structure space, Bioinformatics 22 (2006) 1955–1962.
[17] K.K. Chittur, FTIR/ATR for protein adsorption to biomaterial surfaces, Biomaterials 19 (1998) 357–369.
[18] D.C. Lee, P.I. Haris, Determination of protein secondary structure using factor analysis of infrared spectra, Biochemistry 29 (1990) 9185–9193.
[19] F. Dousseau, M. Pezolet, Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods, Biochemistry 29 (1990) 8771–8779.
[20] N.N. Kalnin, I.A. Baikalov, Quantitative IR spectrophotometry of peptide compounds in water (H_2O) solutions. III. Estimation of the protein secondary structure, Biopolymers 30 (1990) 1273–1280.
[21] R.W. Sarver Jr., W.C. Krueger, Protein secondary structure from Fourier transform infrared spectroscopy: a data base analysis, Anal. Biochem. 194 (1991) 89–100.
[22] L.A. Forato, R. Bernardes-Filho, Protein structure in KBr pellets by infrared spectroscopy, Anal. Biochem. 259 (1998) 136–141.
[23] Y. Wang, R.I. Boysen, Determination of the secondary structure of proteins in different environments by FTIR-ATR spectroscopy and PLS regression, Biopolymers 89 (2008) 895–905.
[24] A. Lorber, A theoretical foundation for the PLS algorithm, J. Chemometr. 1 (1987) 19–31.
[25] M.C. Manning, Underlying assumptions in the estimation of secondary structure content in proteins by circular dichroism spectroscopy—a critical review, J. Pharm. Biomed. Anal. 7 (1989) 1103–1119.
[26] M. Levitt, J. Greer, Automatic identification of secondary structure in globular proteins, J. Mol. Biol. 114 (1977) 181–239.
[27] S. Navea, R. Tauler, Application of the local regression method interval partial least-squares to the elucidation of protein secondary structure, Anal. Biochem. 336 (2005) 231–242.
[28] G. Anderle, R. Mendelsohn, Thermal denaturation of globular proteins. Fourier transform-infrared studies of the amide III spectral region, Biophys. J. 52 (1987) 69–74.
[29] S. Wi, P. Pancoska, Predictions of protein secondary structures using factor analysis on Fourier transform infrared spectra: effect of Fourier self-deconvolution of the amide I and amide II bands, Biospectroscopy 4 (1998) 93–106.
[30] G. Barone, DSC studies on the denaturation and aggregation of serum albumins, Thermochim. Acta 199 (1992) 197–205.
[31] W. Wang, Protein aggregation and its inhibition in biopharmaceutics, Int. J. Pharm. 289 (2005) 1–30.
[32] S. Onoue, Mishandling of the therapeutic peptide glucagon generates cytotoxic amyloidogenic fibrils, Pharm. Res. 21 (2004) 1274–1283.
[33] E.C. Moran, Conformational transitions of glucagon in solution: the alpha to beta transition, Biochem. Biophys. Res. Commun. 77 (1977) 1300–1306.
[34] S. Onoue, Structural transition of glucagon in the concentrated solution observed by electrophoretic and spectroscopic techniques, J. Chromatogr. A 1109 (2006) 167–173.