Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 301 (2023) 122919

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

FTIR-based prediction of collagen content in hydrolyzed protein samples

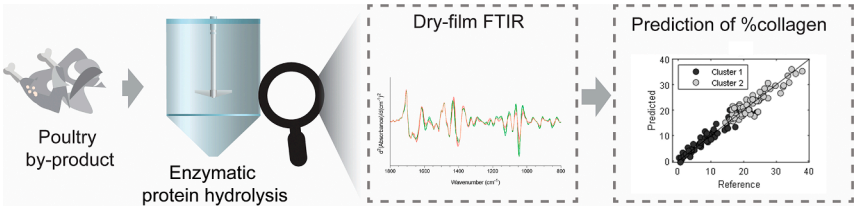
Kenneth Aase Kristoffersen [a,b], Ingrid Måge [a], Sileshi Gizachew Wubshet [a], Ulrike Böcker [a], Katinka Riiser Dankel [a], Andreas Lislelid [a,c], Mats Aksnes Rønningen [a,c], Nils Kristian Afseth [a,*]

[a] Nofima AS – Norwegian Institute of Food, Fisheries and Aquaculture Research, PB 210, NO-1431 Ås, Norway
[b] Faculty of Chemistry, Biotechnology, and Food Science, NMBU – Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Ås, Norway
[c] Department of Mechanical, Electronics and Chemical Engineering, Faculty of Technology, Art and Design, OsloMet – Oslo Metropolitan University, P.O. Box 4, St. Olavs plass, NO-0130 Oslo, Norway

HIGHLIGHTS

- FTIR is a potential tool for quantifying collagen in hydrolyzed proteins.
- Results were validated both with laboratory and industrial test sets.
- With broad variation in collagen, the FTIR response is non-linear.
- The non-linear behavior was improved using hierarchical cluster-based regression.
- This is one of few examples where FTIR is used to quantify protein composition.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Fourier-transform infrared spectroscopy
Collagen content
Hydrolyzed proteins
Hierarchical cluster-based regression
Protein composition

ABSTRACT

Fourier transform infrared spectroscopy (FTIR) is a powerful analytical tool that has been used for protein and peptide characterization for decades. In the present study, the objective was to investigate if FTIR can be used to predict collagen content in hydrolyzed protein samples. All samples were obtained from enzymatic protein hydrolysis (EPH) of poultry by-products providing a span in collagen content from 0.3% to 37.9% (dry weight), and the FTIR analysis was performed using dry film FTIR. Since nonlinear effects were revealed by calibration using standard partial least squares (PLS) regression, Hierarchical Cluster-based PLS (HC-PLS) calibration models were constructed. The HC-PLS model provided a low prediction error when validated using an independent test set (RMSE = 3.3% collagen), while validation using real industrial samples also showed satisfying results (RMSE = 3.2%). The results corresponded well with previously published FTIR-based studies of collagen, and characteristic spectral features for collagen were well identified in the regression models. Covariance between collagen content and other EPH related processing parameters could also be ruled out in the regression models. To the authors' knowledge, this is the first time that collagen content has been systematically studied in solutions of hydrolysed proteins using FTIR. This is also one of few examples where FTIR is successfully used to quantify protein composition. The dry-film FTIR approach presented in the study is expected to be an important tool in the growing industrial segment that is based on sustainable utilization of collagen-rich biomass.

* Corresponding author.
E-mail address: nils.kristian.afseth@nofima.no (N.K. Afseth).

https://doi.org/10.1016/j.saa.2023.122919
Received 6 December 2022; Received in revised form 4 May 2023; Accepted 25 May 2023
Available online 30 May 2023

1. Introduction

Collagen is a class of structural proteins with a characteristic amino acid composition and unique physicochemical properties. For that reason, extracted collagen has been widely utilized in e.g., food, pharmaceuticals, medical products, and cosmetics [1–7]. The abundance of underutilized collagen-rich biomass and the potential for high value applications have made collagen particularly interesting as a degradable biopolymer [8,9]. Attractive applications, ranging from medical use to the development of new degradable plastic-like materials, have increased the demand for collagen with specific qualities [3,10]. To meet these growing and specific demands, development of novel technologies for the extraction and use of collagen from several sources is necessary. Underutilized by-products from processing of meat and fish are in this context regarded as collagen-rich biomass with high potential [11].

In traditional production of collagen-rich products, including partially degraded collagen (i.e., gelatin), the collagen-rich parts of the animals are separated and processed. Collagen extraction methods typically involve solubilization of the collagen in the extraction process, using pH shift, heat energy, hydrolysis under alkaline or acidic conditions, or proteases [3,6]. For smaller animals like poultry and fish, separation of collagen-rich parts is difficult and labor intensive. This can partly explain why collagen products, such as gelatin, mostly originate from bovine and porcine sources. The demand for gelatin is expected to increase in the years to come. As a result, there is a need for technology development for processing underutilized collagen sources from for example poultry meat production [12]. A feasible approach to overcome challenges related to separation of the different components in collagen-rich poultry materials is to use specific proteases and processing parameters to hydrolyze and solubilize the muscle proteins and collagens at different stages in an EPH process. This can be used as a strategy to separate proteins and peptides from muscle and collagen into different fractions [13].

Production of collagen peptides for high-end applications require effective process monitoring and quality characterization technologies. Fourier transform infrared spectroscopy (FTIR) is a powerful analytical tool that has been used for protein and peptide characterization for decades. FTIR can be used for this purpose since the spectra contain detailed structural information on proteins and peptides through nine distinctive infrared absorption bands (i.e., the amide bands) [14,15]. The sensitivity of these vibrational modes also provides a range of possibilities to study parameters related to protein secondary structures, including hydration, solvent effects, pH, degree of hydrolysis (DH%) and peptide size [16–21]. Several studies have also demonstrated that FTIR can be used to monitor proteolytic reactions [22–26], with applications ranging from milk protein hydrolysis to the hydrolysis of poultry-based substrates [27–30]. From these studies it is shown that the amide absorptions (i.e., amide I at ~1650 cm^{-1}), the NH_3^+ deformation (1516 cm^{-1}), and the COO^- stretching (1400 cm^{-1}) are important for prediction of mass-average molar mass and DH% of hydrolyzed proteins [27,28,30]. FTIR has also been used to reveal protein composition in food matrices like milk. However, since the content of protein components is not necessarily directly linked to the structural information of the different amide vibrations, reported predictive performances are usually not high [31]. Eskildsen et al. also showed that for FTIR analysis of milk, protein composition may often be closely related to the protein content [32].

Collagen contains significant amounts of the amino acids hydroxyproline (Hyp) and proline (Pro), which is a unique characteristic for this class of structural proteins, and Hyp is found almost exclusively in collagen. Collagens have therefore been shown to have characteristic features in several absorption bands in FTIR spectra. These include the amide I at ~1655 cm^{-1}, amide II at ~1560 cm^{-1}, and a set of three weaker bands that represent amide III vibrational modes centered at ~1245 cm^{-1} [33]. In denatured and partly hydrolyzed collagen, amide I has been shown to shift to ~1645 cm^{-1} and ~1630 cm^{-1}, which therefore has been linked to Pro and Hyp vibrations [34,35]. When incorporated in a polypeptide chain, Pro and Hyp are tertiary amides. Since the side groups then restrict rotation around the axis of these peptide bonds, the vibrational modes are influenced, resulting in characteristic patterns in the amide bands of the FTIR spectra. In this way, there are characteristic FTIR features that can be used to distinguish collagen containing muscle tissues from tissues with less collagen. Cheheltani et al. utilized these characteristic FTIR features when studying the potential of using FTIR for prediction of collagen and elastin in an *in vitro* model of extracellular matrix degradation in aorta. In this study, both a fiber optic probe directly on aorta samples, as well as microscopic imaging of corresponding aorta sections, were used, and moderate to good predictive performance were obtained for both components [36].

Knowing that FTIR spectra contain quantitative information on collagen in biological tissues, it is of high interest to investigate if FTIR can serve as a generic tool for quantification of collagen during collagen solubilization and degradation. Thus, in the present study, the objective was to investigate if dry film FTIR can be used to predict collagen content in solutions of hydrolyzed proteins. All samples were obtained from EPH of poultry by-products providing a span in collagen content from 0.3% to 37.9% (dry weight). Regression models were constructed using EPH samples and tested on similar but independent lab samples as well as samples from an industrial process. To the best of our knowledge, this is the first time FTIR spectra have been used to predict collagen content in solutions of hydrolyzed protein samples.

2. Materials and methods

2.1. Materials

Poultry raw materials, i.e., turkey tendons (TT) and turkey carcasses (TC), were provided by Nortura (Hærland, Norway). The tendons were manually separated from turkey side-stream materials. Both TT and TC were ground using a Seydelmann SE130 grinder (Stuttgart, Germany), vacuum packed in 350 g packages and stored at −20 °C until use. Proximate analysis and amino acid composition of the raw materials are presented in Supporting Information (SI) Table S-1. The methods used are described in a previously published study [13]. The stem Bromelain powder 1200 GDU was provided by PT. Bromelain Enzyme (Jakarta, Indonesia) and Endocut-02 by Tailorzyme AsP (Herlev, Denmark). Chemicals without further specified origin were all purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2. Samples for calibration and test sets

The calibration sample set was selected from a previously published study by our research group comprising 180 hydrolysate samples [13]. These samples were produced using two poultry raw materials (i.e., TT and TC), two proteases (i.e., stem Bromelain and Endocut-02), and two processing strategies (i.e., thermal inactivation with or without sediment). To ensure an even calibration range, the calibration set in the current study was restricted to samples with collagen content <40%. Samples with high uncertainty in collagen reference measurements (i.e. SD > 2%) were also excluded. Thus, from the original collection of 180 independent hydrolysate samples, 119 samples were used as calibration set in the current study. Proximate analysis and amino acid composition of TT and TC are presented in SI (Table S-1).

An independent test set of 30 samples was produced by sampling upscaled reactions employing similar raw materials, similar proteases and performing sampling at similar timepoints as for the calibration set. For practical reasons, the production of the independent test set samples was performed approximately 1 year after production of the samples for the calibration set. The EPH reactions were performed by mixing 300 g raw material with 600 mL water. The mixture was heated to 50 °C over

40 min before 1% (w/w) protease to raw material was added. The reaction was performed in a Reactor-Ready jacketed reaction vessel (Radleys, Saffron Walden, Essex, United Kingdom) as described by Wubshet et al. [28]. The reaction mixture was sampled and thermally inactivated with or without sediment as described by Kristoffersen et al. [13]. A second test set, subsequently denoted "Industrial test set" was also included. This test set consisted of samples taken after thermal inactivation from an industrial EPH process of poultry raw materials at Bioco (Hærland, Norway) and included 22 samples. For this industrial process, the processing settings and the enzyme used was different than what was used when producing hydrolysates in the laboratory. An overview of number of samples and variation ranges of the data sets used for calibration and testing is provided in Table 1.

2.3. FTIR spectroscopy

Lyophilized samples were dissolved in Milli-Q water (25 mg/mL) and stored overnight at room temperature to rehydrate. The rehydrated samples were stirred for 30 min at 50 °C and filtrated using Millex-HV PVDF 0.45 μm 33 mm filters (MilliporeSigma, Burlington, MA, USA). Aliquots of 7.5 μL were deposited on 96-well IR-transparent Si-plates and dried at room temperature for at least 60 min to form dry films. From each hydrolysate sample, five aliquots were deposited to allow for replicate measurements. FTIR measurements of the dry films were performed using a High Throughput Screening eXTension (HTS-XT) unit coupled to a Tensor 27 spectrometer (Bruker, Billerica, MA, USA), as described by Böcker et al. [27].

Raw spectra of low quality were discarded based on overall absorbance intensity, signal-to-noise ratio, and water vapor content, as described by Måge et al. [37]. For each raw spectrum, a second derivative was calculated using the Savitzky-Golay algorithm with a second order polynomial and a window size of 13 points, followed by standard normal variate (SNV) normalization. Then, spectra from replicate dry-films were averaged to provide one single spectrum per sample, and the region between 1800 cm^{-1} to 800 cm^{-1} was used for further analysis.

2.4. Hydroxyproline analysis

The amino acid Hyp was used to quantify collagen indirectly, and collagen content was calculated based on the assumption that collagen contains 13.5% Hyp per weight [38,39]. The Hyp content of the calibration samples were measured as described by Kristoffersen et al. [13]. The Hyp content in the independent test set and the industrial test set were determined using the same procedure and hydroxyproline assay kit (MAK008-1KT) from Sigma-Aldrich (St. Louis, Mo, USA). All Hyp analysis was performed in triplicates.

2.5. Multivariate calibration

Multivariate calibration was used to establish a model between the FTIR spectra and the collagen content of the samples. Partial least squares (PLS) regression is a well-established method for multivariate calibration that has been used in the domain of spectroscopy and chemometrics for decades [40]. The method extracts a set of latent variables from the spectra, with the criterion that the latent variables have maximum covariance with the response.

The results indicated that a regular PLS regression was non-optimal due to groupings among the samples (see section 3.2). An approach to overcome this challenge is to use Hierarchical Cluster-based PLS (HC-PLS) regression [41]. The method consists of the following steps: 1) Build a global PLS model, 2) Perform a cluster analysis on global score vectors, and 3) Build separate PLS models for each cluster. Using this approach, new samples are predicted by first projecting them onto the global PLS model in step 1 (obtaining the score values), then classifying them to one of the clusters defined in step 2, and finally predicting the response value based on the cluster specific PLS model in step 3. There are several options for the HC-PLS algorithm, see Tøndel et al. for details [41]. In this work the clustering was performed on a reduced rank (3 components) global model, using Fuzzy C-Means clustering with Euclidian distance and fuzzifier = 2.

In PLS calibration, the only parameter to optimize is the number of latent variables. In HC-PLS on the other hand, the number of clusters need to be set first, and then the number of latent variables in each cluster must be optimized independently. In this study, two clusters were selected based on visual inspection of the PLS scores and the fact that the number of samples is limited. Cross-validation with ten random segments was used to select the number of PLS components in all models. After parameter optimization, the predictive performances of the models were assessed by an independent test set. Plots of predicted versus reference values as well as the root mean squared error (RMSE), the Bias, and the bias-corrected Standard Error of Prediction (SEP_b) were reported. The RMSE is an estimate of the expected difference between predictions and reference values for new samples. The RMSE can be split into contributions from bias (mean offset) and bias-corrected prediction error (SEP_b,), giving additional insight into sources of prediction error. The relation between them is $RMSE^2 = Bias^2 + SEP_b^2$. All the multivariate data analysis were performed in MATLAB Release 2021a (The MathWorks, Inc., Natick, MA, USA).

3. Results and discussion

3.1. Collagen content and FTIR profiling

All FTIR analysis was performed using the dry film FTIR approach. This sampling technique was used since the water bands otherwise would be expected to overlap important information found in the amide I region [30]. The samples included in this study had a collagen content ranging from 0.3% to 37.9%. An overview of samples and collagen variation ranges of the data sets used for calibration and testing is provided in Table 1.

Fig. 1A shows the amide I region of second derivative FTIR spectra of three selected EPH samples containing 1.4%, 26.9% and 55.3% collagen, respectively. The second derivative spectral region from 1750 cm^{-1} to 1550 cm^{-1} of an EPH time series produced from turkey tendons is displayed in Fig. 1B [13]. The collagen content of the samples in Fig. 1B ranged from 23.9% to 37.9%.

A gradual shift towards a sharper ~1645 cm^{-1} peak of the amide I band can be observed in the FTIR spectra of the time series. By comparing the second derivate spectra in Fig. 1A to the time series in Fig. 1B it is evident that the observed shift may be related to an increase in collagen content. An increase in the amount of collagen fragments and peptides in a sample will result in an increase in the amount of the amino acids Hyp and Pro which are found in high concentrations in collagen. Polypeptides with a high Hyp and Pro content similar to what is found in collagen have been thoroughly studied using FTIR. Lazarev et al. observed three distinct amide I component peaks (i.e., ~1665 cm^{-1}, ~1645 cm^{-1}, and ~1630 cm^{-1}) related to collagen-like polypeptides [42]. Another report using FTIR to study rat skin collagen and calf skin

Table 1
Overview of collagen content (%) in all data sets used for calibration and testing.

Data set		Number of samples	mean	std	min	max
Calibration set	All	119	16.0	8.9	0.3	37.9
	Cluster 1	55	8.5	5.5	0.3	23.9
	Cluster 2	64	22.4	5.7	14.4	37.9
Independent test set	All	30	14.6	8.4	2.6	35.8
	Cluster 1	22	10.8	5.3	2.6	21.9
	Cluster 2	8	25.3	5.9	15.8	35.8
Industrial test set	All / Cluster 2	22	26.2	3.4	22.0	33.7

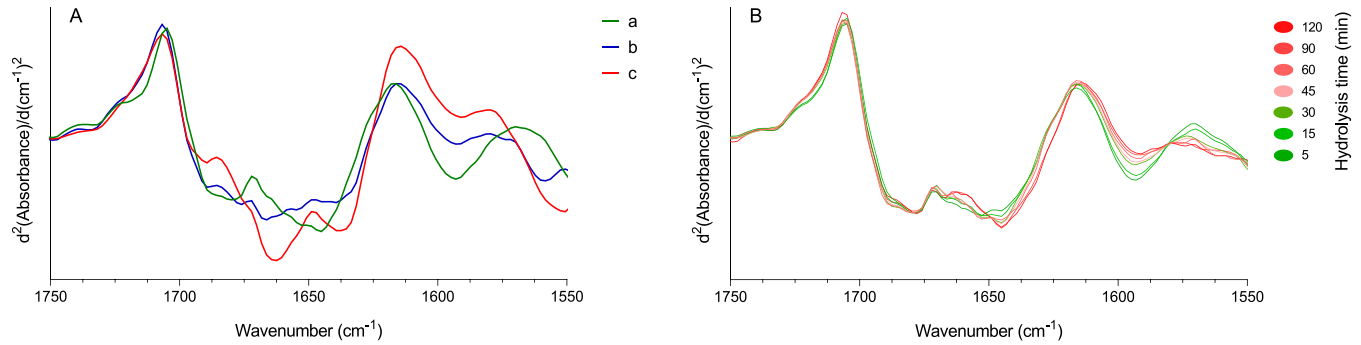


Fig. 1. A) The second derivate FTIR spectra from 1750 cm^{-1} to 1550 cm^{-1} of a) EPH sample (1.4% collagen), b) EPH sample (26.9% collagen) and c) EPH sample (55.3% collagen). The latter sample was added for illustration purposes and is not part of the calibration set. B) The second derivate FTIR spectra from 1750 cm^{-1} to 1550 cm^{-1} of an EPH time series produced using turkey tendons.

gelatin identified three component absorption peaks within the amide I band at 1660 cm^{-1}, 1643 cm^{-1}, and 1633 cm^{-1}. These peaks were common to the spectra irrespective of the degree of triple helical content of the samples, and a shift towards the 1633 cm^{-1} were observed with increasing denaturation levels [34]. The amide I band of samples containing protein fragments and peptides with less Hyp and Pro will not have the same characteristics. Changes in the amide I band related to collagen content are thus expected, and similar features can be observed in both Fig. 1A and 1B.

3.2. Multivariate calibration

The spectral region between 1800 cm^{-1} to 800 cm^{-1} of the second derivative FTIR spectra was used to establish the relationship between FTIR spectra and the corresponding collagen content. A calibration set consisting of 119 samples with a collagen content from 0.3% to 37.9%, was used for model building and parameter optimization. An independent test set, consisting of 30 samples with a collagen content from 2.6% to 35.8%, was used to validate how well the model predicts new samples.

The global PLSR model, with five latent variables, had a cross-validated prediction error ($RMSE_{CV}$) of 3.2% collagen and a test set prediction error ($RMSE_{test}$) of 3.9% collagen. Predicted versus reference values from cross-validation and test set are plotted in Fig. 2A and 2B, respectively. Even if $RMSE_{test}$ is just slightly higher than $RMSE_{CV}$, Fig. 2B shows that samples containing less collagen are systematically over-estimated while collagen-rich samples are slightly under-estimated. This is an indication of non-linear relationships between the spectra and the response.

Closer inspection of the scores and loadings from the global PLS model (See SI Fig. S-1) revealed that the first latent variable spanned a variation mainly present in samples with collagen above approximately 15%, related to several peaks in the area 1700–1550 cm^{-1} and around 1400 cm^{-1}. This is an indication that the relationship between FTIR and

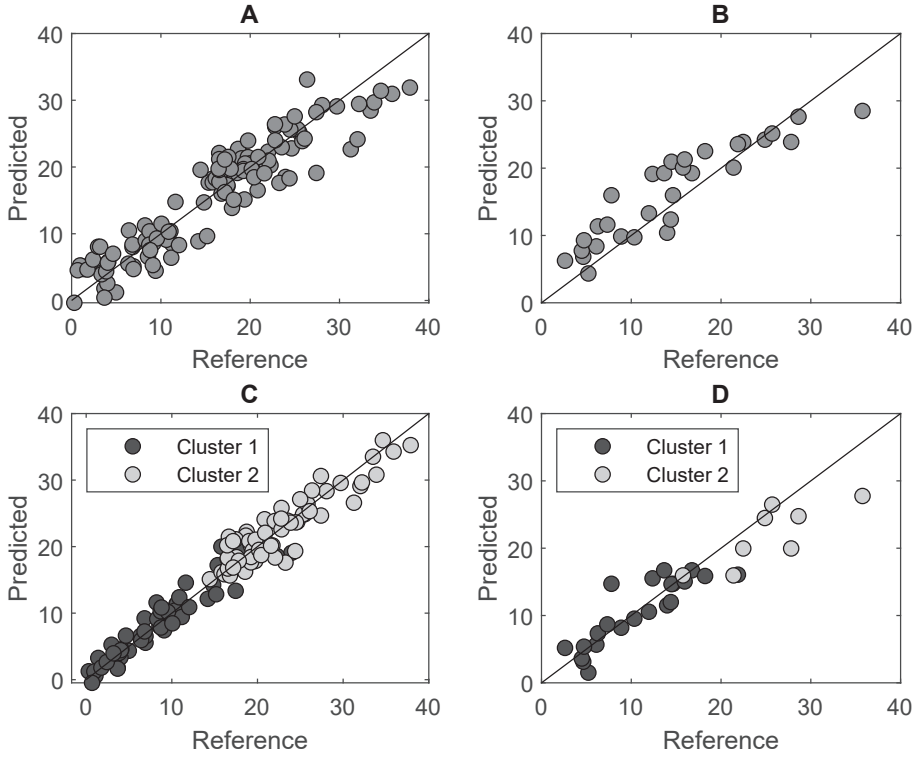


Fig. 2. Predicted versus measured collagen: A) cross-validation, global PLS model, B) test set, global PLS model, C) cross-validation, HC-PLC model, and D) test set, HC-PLS model.

collagen is different for samples containing low and high amounts of collagen, which may explain the challenges in predictive performance illustrated in Fig. 2B. It is also important to note that during EPH of collagen-rich raw materials, there are several processes going on simultaneously: 1) A gradual solubilization of protein components, and 2) A gradual breakdown of the solubilized proteins and peptides. The calibration sample set used in the present study has different mass-average molar mass and DH% values, ranging from low to relatively high, see Kristoffersen et al. [13]. Samples with similar collagen content may therefore exhibit different spectral fingerprints due to differences in degradation degree (i.e., DH%). When the mass-average molar mass value of a sample is low (at high DH%) most of the ordered secondary structure of the protein fragments and peptides are broken down. This shows that different processes will affect the spectral information reflected in amide bands of the samples.

To overcome these challenges, an HC-PLS model was built, clustering the samples into two groups based on the first three latent variables from the global PLS model. The two clusters consisted of 55 and 64 samples with a collagen content ranging from 0.3% to 23.9% (Cluster 1) and 14.4% to 37.9% (Cluster 2). Comparing the mass-average molar mass and DH% data of the two clusters, cluster 1 had on average lower mass-average molar mass and higher DH% values relative to cluster 2 [13]. The optimal PLS calibration model for cluster 1 had four latent variables, while the model for cluster 2 had six latent variables. The RMSE, SEP_{-b} and Bias for the cross-validation and the independent test set is given in Table 2 and predicted versus measured plots are shown in Fig. 2C and 2D. It is clear from Fig. 2 and Table 2 that predictions of the samples in cluster 1, i.e., low collagen, low mass-average molar mass and high DH %, is improved by the cluster-based model, mainly due to reduced bias but also slightly lower SEP_{-b}. Samples in cluster 2, on the other hand, have a larger bias in the cluster-based model while the SEP_{-b} is unaltered. Note that there are only eight test set samples for cluster 2, so these numbers are highly uncertain. This shows that in future development of an FTIR calibration, emphasis have to be put on making the cluster 2 range more robust by e.g. extending the sample size in this region.

The regression coefficients for the three models are plotted in Fig. 3. The global model and the cluster 2 model generally follow the same pattern, but the cluster 2 coefficients have slightly higher intensities in the amide band regions of the FTIR spectra related to collagen. This might explain why the cluster 2 model performs poorer on the test set as large intensities in the regression coefficients make the model more sensitive and thereby less robust to small changes in the spectra related to secondary structures. Also, since the global and cluster 2 models are very similar, it is natural that the global model is more universal since it is based on a larger calibration set.

The regression coefficients for the cluster 1 model deviate from the cluster 2 model in several spectral regions, and differences in the regions associated with both amide bands and the NH_3^+ and the COO^- bands can be observed. The characteristic absorption bands for collagen seems to be important for all the three models. However, the differences seen between the regression coefficients of clusters 1 and 2 can indeed be related to differences in collagen content and DH% of the samples. The samples in cluster 1 contain less than 23.9% collagen and their DH% values are in most cases relatively high compared to cluster 2. As previously discussed, due to the composition (i.e., muscle proteins and peptides relative to collagen proteins and peptides) and differences in protein degradation, the amide I absorption band will be different when comparing FTIR spectra from the two clusters. Some of these differences are shown in Fig. 1A where sample a and b have a stronger absorption bans between 1700 and 1670 cm^{-1} and lower between 1670 and 1630 cm^{-1} compared to sample c (which contains the most collagen). Similar differences are also seen in the regression coefficients of Fig. 3.

3.3. Cage of covariance

The amount of collagen in a hydrolysate is inevitably connected to reaction time, and it is therefore expected to covary with other time-dependent parameters such as DH% and mass-average molar mass [13]. Previous work has shown that several "reaction time" dependent parameters can be identified in FTIR spectra [29,30,43]. Within a time series collagen content usually increases with increase in DH% and decrease in mass-average molar mass caused by protein digestion. As a result the FTIR spectra of a time series will contain changes associated with secondary structure, increase in N-terminals (NH_3^+, ~1516 cm^{-1}) and C-terminals (COO^-, ~1404 cm^{-1}) and distinct peaks associated with increase in tertiary amides (~1630 cm^{-1}) [27,28,33,36]. In such cases, there is a risk that a calibration model performs well due to indirect correlations instead of chemical signals caused by the desired response itself. This phenomenon is referred to as the cage of covariance [44,45].

The studied calibration set was deliberately designed to create variation in collagen independent of DH% and mass-average molar mass by using raw materials with very different collagen content, different proteases, and thermal inactivation with or without sediment [13]. The correlation between collagen content (%) and DH% within the same raw material and with separation of the water phase before inactivation, was positive (0.6 and 0.8 for tendons and carcass respectively). Correlation among samples that were inactivated with sediment was negative (-0.6). The overall correlation in the full calibration set was −0.5 (see SI Fig. S-2).

To investigate this further, a new model was fitted using a subsample of our calibration set with an almost constant DH% but with varying collagen content. The subset contained 42 samples that spanned the full range of collagen but only a narrow range of DH% and mass-average molar mass (see SI Fig. S-2). The regression coefficients were very similar to those from the global model, and the RMSE for the test set was 3.4% collagen (see SI Fig. S-3). Based on these observations it was concluded that the calibration model for collagen is not caused by indirect correlations with DH% or changes in mass-average molar mass.

3.4. Prediction of collagen content in industrial test set samples

The calibration model developed was tested on 22 samples collected from an EPH factory that processes poultry by-products. The collagen content in these samples varied from 22.0% to 33.7%. They were all classified to cluster 2 in the HC-PLS model. The RMSE were 4.5% and

Table 2
Model performances from cross-validation (CV), test set validation, and on the industrial test set. The statistics are the Root Mean Squared Error (RMSE), the bias-corrected Standard Error of Prediction (SEP_{-b}) and the bias. The CV and test set statistics correspond to Fig. 2, while industrial test set predictions are plotted in Fig. 4.

		All samples			Cluster 1			Cluster 2		
		RMSE	SEP_{-b}	bias	RMSE	SEP_{-b}	bias	RMSE	SEP_{-b}	bias
Global PLS	CV	3.2	3.2	0	2.9	2.9	0	3.5	3.5	0
	Test set	3.9	3.9	−1.8	4.1	3.0	−2.8	3.4	3.4	1.1
	Industry set	4.5	3.0	3.5				4.5	3.0	3.5
HC-PLS	CV	2.8	2.8	0	2.2	2.2	0	3.2	3.2	0
	Test set	3.3	3.2	1.1	2.6	2.7	0.2	4.7	3.4	3.4
	Industry set	3.2	2.7	1.9				3.2	2.7	1.9

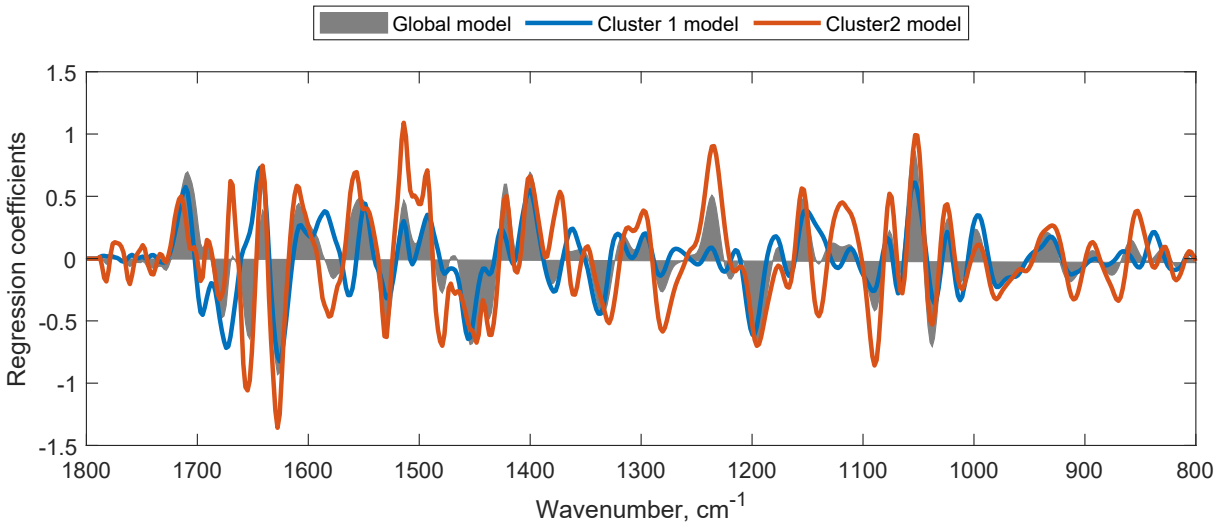


Fig. 3. Regression coefficients from global PLS (grey) and HC-PLS (red and blue).

3.2% in the global and cluster specific model, respectively (See Table 2 and Fig. 4). The improvement in prediction error by the HC-PLS model is mainly caused by a lower bias (−3.5 versus −1.9). It is not surprising to observe a large Bias in this case, since the majority of the raw material in the industrial EPH process originated from chicken while the calibration set is based on turkey only. The factory also uses a different protease, which is known to affect the protein degradation pattern observed in the FTIR spectra [29]. Including relevant industrial samples in the calibration is therefore expected to improve the prediction performance, but the current results still indicate the industrial potential of using FTIR to predict collagen content in hydrolyzed protein samples.

3.5. General discussion

Extracted collagen can be used for many applications dependent on the extract's quality. Collagen source and processing are important factors influencing the physicochemical properties of extracts [3,6]. Variation in the composition of the collagen-rich by-products that are used as a collagen source will also influence the quality of the extracts making it challenging to keep a stable product quality for specific applications. An example is shown in a previous publication demonstrating that collagen from poultry by-products is solubilized to some extent during an EPH process and that the thermal inactivation step is a critical step in solubilization and extraction of the collagen [13]. It was concluded that separation of the water and lipid phase prior to thermal inactivation could be a strategy to separate collagens from muscle proteins in industrial EPH processes. This approach will make it possible to tailor the composition of the hydrolysates with regard to collagen content of the hydrolysates. The challenge, however, is that variation in the composition of the collagen-rich material can affect collagen solubilization during the EPH process. Therefore, development of tools to monitor the amount of collagen dissolved in the water phase during an EPH process will be critical for implementing this strategy in industry.

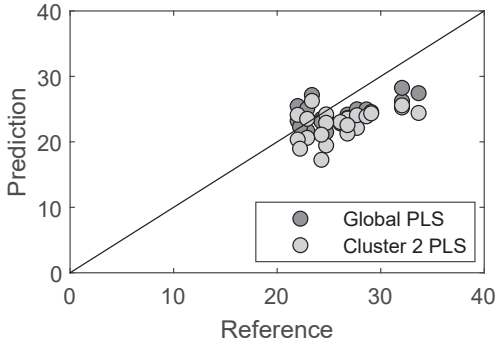


Fig. 4. The predicted collagen content of industry samples using a global and cluster specific model.

The presented study demonstrates the possibility of using FTIR-based prediction of collagen content in EPH samples from poultry, which is a collagen source with high potential for several applications, including gelatin production [12,46]. But it is expected that an FTIR approach also will find use in applications where other collagen-rich raw materials are processed, like e.g. marine by-products. It is also important to note that the reference methodology used in the current study, i.e., Hyp analysis, is associated with high estimation errors [13], highlighting the promising FTIR results of the study. Thus, together with previous studies on collagen quantification using FTIR, the methodology developed here is expected to be applicable for EPH-based processing of collagen-rich materials in general. Since the dry film FTIR technique used also can generate FTIR spectra with only a few minutes delay when applying quick drying methods described by e.g. Ayvaz and Temizkan [47], dry film FTIR is clearly a promising tool for monitoring and controlling collagen content in protein hydrolysates from EPH processes.

4. Conclusion

The present study shows the potential of FTIR for prediction of collagen content in solutions of hydrolyzed poultry proteins. With the dry-film FTIR approach, the characteristic spectral features of collagen and collagen peptides is revealed in the spectra, enabling sound quantitative regression models for collagen content measurements. Since the calibration set contained samples with a relatively large collagen content range, nonlinear effects were revealed. By employing an HC-PLS model the prediction performance of samples with lower collagen content improved significantly. However, for samples with higher collagen content the prediction performance remained similar to the global PLS approach. To the authors knowledge, the current study is one of few examples where FTIR is successfully used to quantify protein composition. Based on previously published studies and the presented work, the dry film FTIR approach is expected to be an important tool in the growing industrial segment that is based on sustainable utilization of collagen-rich biomass.

CRediT authorship contribution statement

Kenneth Aase Kristoffersen: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Ingrid Måge:** Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Sileshi Gizachew Wubshet:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. **Ulrike Böcker:** Data curation, Methodology, Writing – review & editing. **Katinka Riiser Dankel:** Data curation, Methodology, Writing – review & editing. **Andreas Lislelid:** Data curation, Writing – review & editing. **Mats Aksnes Rønningen:** Data curation, Writing – review & editing. **Nils Kristian Afseth:** Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Financial support from the Norwegian Fund for Research Fees on Agricultural Products through the project “Precision Food Production” (no. 314111) and from the Norwegian Research Council through the projects “Notably” (no. 280709) and “SFI Digital Food Quality” (no. 309259) is greatly acknowledged. Internal financing from Nofima through the project “Peptek” is also greatly acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.saa.2023.122919.

References

[1] T.H. Silva, et al., Marine Origin Collagens and Its Potential Applications, Mar. Drugs 12 (12) (2014) 5881–5901.
[2] M.C. Gomez-Guillen, et al., Functional and bioactive properties of collagen and gelatin from alternative sources: A review, Food Hydrocoll. 25 (8) (2011) 1813–1827.
[3] D. Liu, et al., Collagen and gelatin, Annu. Rev. Food Sci. Technol. 6 (2015) 527–557.
[4] S. Chattopadhyay, R.T. Raines, Collagen-based biomaterials for wound healing, Biopolymers 101 (8) (2014) 821–833.
[5] H. Hong, et al., Preparation of low-molecular-weight, collagen hydrolysates (peptides): Current progress, challenges, and future perspectives, Food Chem. 301 (2019), 125222.
[6] P. Hashim, et al., Collagen in food and beverage industries, Int. Food Res. J. 22 (1) (2015) 1–8.
[7] M.I.A. Rodriguez, L.G.R. Barroso, M.L. Sanchez, Collagen: A review on its sources and potential cosmetic applications, J. Cosmet. Dermatol. 17 (1) (2018) 20–26.
[8] M.D. Shoulders, R.T. Raines, Collagen Structure and Stability, Annu. Rev. Biochem 78 (1) (2009) 929–958.
[9] W. Yang, M.A. Meyers, R.O. Ritchie, Structural architectures with toughening mechanisms in Nature: A review of the materials science of Type-I collagenous materials, Prog. Mater Sci. 103 (2019) 425–483.
[10] L. Cen, et al., Collagen tissue engineering: Development of novel biomaterials and applications, Pediatr. Res. 63 (5) (2008) 492–496.
[11] T. Aspevik, et al., Valorization of Proteins from Co- and By-Products from the Fish and Meat Industry, Top. Curr. Chem. 375 (3) (2017) 53.
[12] A. Abedinia, et al., Poultry gelatin: Characteristics, developments, challenges, and future outlooks as a sustainable alternative for mammalian gelatin, Trends Food Sci. Technol. 104 (2020) 14–26.
[13] K.A. Kristoffersen, et al., Post-enzymatic hydrolysis heat treatment as an essential unit operation for collagen solubilization from poultry by-products, Food Chem. (2022), 132201.
[14] U. Böcker, et al., Salt-induced changes in pork myofibrillar tissue investigated by FT-IR microspectroscopy and light microscopy, J. Agric. Food Chem. 54 (18) (2006) 6733–6740.
[15] P. Gelfand, et al., Characterization of Protein Structural Changes in Living Cells Using Time-Lapsed FTIR Imaging, Anal. Chem. 87 (12) (2015) 6025–6031.
[16] A. Barth, The infrared absorption of amino acid side chains, Prog. Biophys. Mol. Biol. 74 (3) (2000) 141–173.
[17] N. Perisic, et al., Monitoring Protein Structural Changes and Hydration in Bovine Meat Tissue Due to Salt Substitutes by Fourier Transform Infrared (FTIR) Microspectroscopy, J. Agric. Food Chem. 59 (18) (2011) 10052–10061.
[18] P.V. Andersen, E. Veiseth-Kent, J.P. Wold, Analyzing pH-induced changes in a myofibril model system with vibrational and fluorescence spectroscopy, Meat Sci. 125 (2017) 1–9.
[19] S.H. Arabi, et al., Serum albumin hydrogels in broad pH and temperature ranges: characterization of their self-assembled structures and nanoscopic and macroscopic properties, Biomater. Sci. 6 (3) (2018) 478–492.
[20] G. Martra, et al., The Formation and Self-Assembly of Long Prebiotic Oligomers Produced by the Condensation of Unactivated Amino Acids on Oxide Surfaces, Angew. Chem. Int. Ed. 53 (18) (2014) 4671–4674.
[21] H.-F. Okabayashi, H.-H. Kanbe, C.J. O'Connor, The role of an L-leucine residue on the conformations of glycyl-L-leucine oligomers and its N- or C-terminal dependence: infrared absorption and Raman scattering studies, Eur. Biophys. J. 45 (1) (2016) 23–34.
[22] C. Ruckebusch, et al., Hydrolysis of hemoglobin surveyed by infrared spectroscopy II. Progress predicted by chemometrics, Anal. Chim. Acta 396 (2–3) (1999) 241–251.
[23] C. Ruckebusch, L. Duponchel, J.P. Huvenne, Degree of hydrolysis from mid-infrared spectra, Anal. Chim. Acta 446 (1–2) (2001) 257–268.
[24] C. Ruckebusch, et al., On-line mid-infrared spectroscopic data and chemometrics for the monitoring of an enzymatic hydrolysis, Appl. Spectrosc. 55 (12) (2001) 1610–1617.
[25] G. Guler, et al., Real time observation of proteolysis with Fourier transform infrared (FT-IR) and UV-circular dichroism spectroscopy: Watching a protease eat a protein, Spectrochim. Acta A Mol. Biomol. Spectrosc. 79 (1) (2011) 104–111.
[26] G. Guler, et al., Proteolytically-induced changes of secondary structural protein conformation of bovine serum albumin monitored by Fourier transform infrared (FT-IR) and UV-circular dichroism spectroscopy, Spectrochim. Acta Part a-Mol. Biomol. Spectrosc. 161 (2016) 8–18.
[27] U. Böcker, et al., Fourier-transform infrared spectroscopy for characterization of protein chain reductions in enzymatic reactions, Analyst 142 (15) (2017) 2812–2818.
[28] S.G. Wubshet, et al., FTIR as a rapid tool for monitoring molecular weight distribution during enzymatic protein hydrolysis of food processing by-products, Anal. Methods 9 (29) (2017) 4247–4254.
[29] K.A. Kristoffersen, et al., FTIR-based hierarchical modeling for prediction of average molecular weights of protein hydrolysates, Talanta 205 (2019), 120084.
[30] N.A. Poulsen, et al., Predicting hydrolysis of whey protein by mid-infrared spectroscopy, Int. Dairy J. 61 (2016) 44–50.
[31] T. Bresolin, J.R.R. Dórea, Infrared Spectrometry as a High-Throughput Phenotyping Technology to Predict Complex Traits in Livestock Systems, Front. Genet. 11 (2020).
[32] C.E. Eskildsen, et al., Quantification of bovine milk protein composition and coagulation properties using infrared spectroscopy and chemometrics: A result of collinearity among reference variables, J. Dairy Sci. 99 (10) (2016) 8178–8186.
[33] B. de Campos Vidal, M.L.S. Mello, Collagen type I amide I band infrared spectroscopy, Micron 42 (3) (2011) 283–289.
[34] K.J. Payne, A. Veis, Fourier transform ir spectroscopy of collagen and gelatin solutions: Deconvolution of the amide I band for conformational studies, Biopolymers 27 (11) (1988) 1749–1760.
[35] M.A. Bryan, et al., FTIR studies of collagen model peptides: complementary experimental and simulation approaches to conformation and unfolding, J. Am. Chem. Soc. 129 (25) (2007) 7877–7884.
[36] R. Cheheltani, et al., Fourier transform infrared spectroscopy to quantify collagen and elastin in an in vitro model of extracellular matrix degradation in aorta, Analyst 139 (12) (2014) 3039–3047.
[37] I. Måge, et al., Fourier-transform infrared (FTIR) fingerprinting for quality assessment of protein hydrolysates, LWT 152 (2021), 112339.
[38] I. Stoilov, et al., Chapter 7 - Measurement of elastin, collagen, and total protein levels in tissues, in: R.P. Mecham (Ed.), Methods in Cell Biology, Academic Press, 2018, pp. 133–146.
[39] D.D. Cissell, et al., A Modified Hydroxyproline Assay Based on Hydrochloric Acid in Ehrlich's Solution Accurately Measures Tissue Collagen Content, Tissue Eng. Part C Methods 23 (4) (2017) 243–250.
[40] H. Martens, T. Naes, Multivariate calibration, John Wiley & Sons, 1989.
[41] K. Tøndel, et al., Hierarchical Cluster-based Partial Least Squares Regression (HC-PLSR) is an efficient tool for metamodelling of nonlinear dynamic models, BMC Syst. Biol. 5 (1) (2011) 90.
[42] Y.A. Lazarev, B.A. Grishkovsky, T.B. Khromova, Amide I band of IR spectrum and structure of collagen and related polypeptides, Biopolymers 24 (8) (1985) 1449–1478.
[43] K.A. Kristoffersen, et al., Average molecular weight, degree of hydrolysis and dry-film FTIR fingerprint of milk protein hydrolysates: Intercorrelation and application in process monitoring, Food Chem. 310 (2020), 125800.
[44] D.T. Berhe, et al., Prediction of total fatty acid parameters and individual fatty acids in pork backfat using Raman spectroscopy and chemometrics: Understanding

the cage of covariance between highly correlated fat parameters, Meat Sci. 111 (2016) 18–26.
[45] C.E. Eskildsen, et al., Cage of covariance in calibration modeling: Regressing multiple and strongly correlated response variables onto a low rank subspace of explanatory variables, Chemom. Intel. Lab. Syst. 213 (2021), 104311.
[46] K.G. Grønlien, et al., Collagen from Turkey (Meleagris gallopavo) tendon: A promising sustainable biomaterial for pharmaceutical use, Sustain. Chem. Pharm. 13 (2019), 100166.
[47] H. Ayvaz, R. Temizkan, Quick vacuum drying of liquid samples prior to ATR-FTIR spectral collection improves the quantitative prediction: a case study of milk adulteration, Int. J. Food Sci. Technol. 53 (11) (2018) 2482–2489.