

**This is an electronic reprint of the original article.  
This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Rieppo, Lassi; Kokkonen, Harri T.; Kulmala, Katariina A. M.; Kovanen, Vuokko; Lammi, Mikko J.; Töyräs, Juha; Saarakkala, Simo

**Title:** Infrared microspectroscopic determination of collagen cross-links in articular cartilage

**Year:** 2017

**Version:**

**Please cite the original version:**

Rieppo, L., Kokkonen, H. T., Kulmala, K. A. M., Kovanen, V., Lammi, M. J., Töyräs, J., & Saarakkala, S. (2017). Infrared microspectroscopic determination of collagen cross-links in articular cartilage. *Journal of Biomedical Optics*, 22(3), Article 035007.  
<https://doi.org/10.1117/1.JBO.22.3.035007>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

# Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

## Infrared microspectroscopic determination of collagen cross-links in articular cartilage

Lassi Rieppo  
Harri T. Kokkonen  
Katariina A. M. Kulmala  
Vuokko Kovanen  
Mikko J. Lammi  
Juha Töyräs  
Simo Saarakkala

**SPIE.**

Lassi Rieppo, Harri T. Kokkonen, Katariina A. M. Kulmala, Vuokko Kovanen, Mikko J. Lammi, Juha Töyräs, Simo Saarakkala, "Infrared microspectroscopic determination of collagen cross-links in articular cartilage," *J. Biomed. Opt.* **22**(3), 035007 (2017), doi: 10.1117/1.JBO.22.3.035007.

# Infrared microspectroscopic determination of collagen cross-links in articular cartilage

Lassi Rieppo,<sup>a,b,\*</sup> Harri T. Kokkonen,<sup>c</sup> Katariina A. M. Kulmala,<sup>b</sup> Vuokko Kovanen,<sup>d</sup> Mikko J. Lammi,<sup>e,f</sup> Juha Töyräs,<sup>b,g</sup> and Simo Saarakkala<sup>a,h,i</sup>

<sup>a</sup>University of Oulu, Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, Oulu, Finland

<sup>b</sup>University of Eastern Finland, Department of Applied Physics, Kuopio, Finland

<sup>c</sup>South Karelia Central Hospital, Department of Radiology, Lappeenranta, Finland

<sup>d</sup>University of Jyväskylä, Department of Health Sciences, Jyväskylä, Finland

<sup>e</sup>Umeå University, Department of Integrative Medical Biology, Umeå, Sweden

<sup>f</sup>Health Science Center of Xi'an Jiaotong University, School of Public Health, Key Laboratory of Trace Elements and Endemic Diseases, National Health and Family Planning Commission, Xi'an, China

<sup>g</sup>Kuopio University Hospital, Diagnostic Imaging Center, Kuopio, Finland

<sup>h</sup>Oulu University Hospital, Department of Diagnostic Radiology, Oulu, Finland

<sup>i</sup>University of Oulu and Oulu University Hospital, Medical Research Center Oulu, Oulu, Finland

**Abstract.** Collagen forms an organized network in articular cartilage to give tensile stiffness to the tissue. Due to its long half-life, collagen is susceptible to cross-links caused by advanced glycation end-products. The current standard method for determination of cross-link concentrations in tissues is the destructive high-performance liquid chromatography (HPLC). The aim of this study was to analyze the cross-link concentrations nondestructively from standard unstained histological articular cartilage sections by using Fourier transform infrared (FTIR) microspectroscopy. Half of the bovine articular cartilage samples ( $n = 27$ ) were treated with threose to increase the collagen cross-linking while the other half ( $n = 27$ ) served as a control group. Partial least squares (PLS) regression with variable selection algorithms was used to predict the cross-link concentrations from the measured average FTIR spectra of the samples, and HPLC was used as the reference method for cross-link concentrations. The correlation coefficients between the PLS regression models and the biochemical reference values were  $r = 0.84$  ( $p < 0.001$ ),  $r = 0.87$  ( $p < 0.001$ ) and  $r = 0.92$  ( $p < 0.001$ ) for hydroxylysyl pyridinoline (HP), lysyl pyridinoline (LP), and pentosidine (Pent) cross-links, respectively. The study demonstrated that FTIR microspectroscopy is a feasible method for investigating cross-link concentrations in articular cartilage.

© The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.22.3.035007](https://doi.org/10.1117/1.JBO.22.3.035007)]

Keywords: articular cartilage; collagen; cross-links; infrared spectroscopy; multivariate analysis.

Paper 160692R received Oct. 6, 2016; accepted for publication Feb. 22, 2017; published online Mar. 14, 2017.

## 1 Introduction

Articular cartilage is a specialized avascular and aneural connective tissue that covers the ends of long bones. Articular cartilage has an important role in reducing the contact stresses directed to bone ends, and, together with synovial fluid, it provides a nearly frictionless contact between the articulating bones. Articular cartilage is mainly composed of water, type II collagen, proteoglycans, and chondrocytes.<sup>1</sup> Collagen forms a highly organized fibrous network in articular cartilage to provide the tissue tensile stiffness. Osteoarthritis, the most common joint disease, is characterized by degradation of cartilage tissue. Age is one of the most significant risk factors of osteoarthritis. Collagen has an exceptionally long half-life<sup>2</sup> and, therefore, it is susceptible to cross-links caused by advanced glycation end-products (AGEs).<sup>3</sup> This type of cross-linking of collagen increases the stiffness of articular cartilage making it more brittle and susceptible to injuries. Therefore, accumulation of AGEs may be one of the reasons why age increases the risk of osteoarthritis.<sup>3</sup>

Fourier transform infrared (FTIR) spectroscopy is a technique that provides information on the chemical composition of the investigated sample. Infrared light is absorbed by a

molecule at characteristic frequencies, i.e., at frequencies that correspond to one of the vibrational modes of the molecule. Therefore, an infrared absorption spectrum is often called a chemical fingerprint of the sample. FTIR microspectroscopy, a combination of spectroscopy and microscopy, is an intriguing method, as it enables determination of collagen and proteoglycan distributions from unstained histological articular cartilage sections. This is conducted either by calculating integrated areas of absorption peaks<sup>4</sup> or by calibrating multivariate models, e.g., principal component<sup>5</sup> or partial least squares (PLS)<sup>6,7</sup> regression models, against reference information about the investigated compound. The capability of FTIR microspectroscopy to determine collagen and proteoglycan contents has been demonstrated and validated against multiple reference methods.<sup>5-9</sup> Currently, the standard method to determine the cross-link concentrations in tissues is high-performance liquid chromatography (HPLC). As it is a destructive method, dedicated sample blocks are needed for the analysis. In principle, as chemical information exhibits in FTIR spectra, FTIR microspectroscopy could be used to determine the cross-link concentrations from standard histological sections.

Glycation of type I collagen has been investigated by FTIR<sup>10</sup> and Raman spectroscopy.<sup>11</sup> The studies concluded that the carbohydrate region of the spectrum can be used as a marker

\*Address all correspondence to: Lassi Rieppo, E-mail: [lassi.rieppo@oulu.fi](mailto:lassi.rieppo@oulu.fi)

of the glycation level. In tissue level studies, the ratio of the carbohydrate region to the amide I and amide II regions was utilized for studying the glycation in cardiac tissue.<sup>12</sup> Recently, the same approach was used to study AGE components in articular cartilage.<sup>13</sup> However, the same parameter has also been used to assess the ratio of proteoglycan content to collagen content in articular cartilage.<sup>14</sup> As proteoglycans are a major component in articular cartilage, it is likely that influence of AGEs to the carbohydrate region is mostly masked by contribution of proteoglycans.

L-threose, a 4-carbon sugar, is a degradation product of ascorbic acid.<sup>15</sup> Threose modifies lysine residues in proteins and forms a number of characteristic AGEs.<sup>3</sup> Therefore, threose can be used to mimic age-related increase in cross-link concentration in articular cartilage. The first aim of this study was to separate intact and threose-treated bovine articular cartilage samples from each other using FTIR microspectroscopy and PLS regression-based discriminant analysis. The second aim was to determine the concentrations of specific enzymatic and nonenzymatic cross-links in articular cartilage using FTIR microspectroscopy combined with PLS regression. In particular, we studied the use of different variable selection algorithms in combination with PLS regression to find the most suitable method for quantification of cross-links in articular cartilage.

## 2 Methods

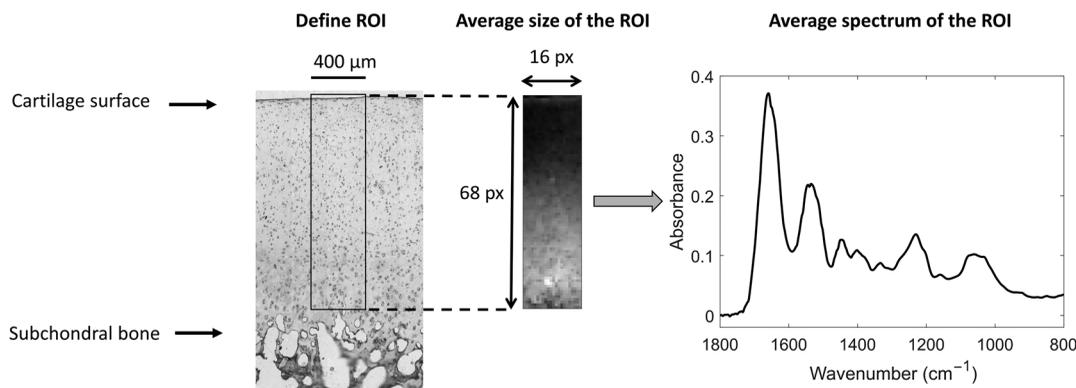
### 2.1 Sample Preparation

The samples of this study were originally collected and prepared for two other studies.<sup>16,17</sup> Intact bovine patellae ( $n = 14$ ) were obtained from a local abattoir (Atria Oyj, Kuopio, Finland). Ethical approval was not required, as the animals were not slaughtered for the purposes of this study. The bovine knees were stored at room temperature after slaughtering. The knee joints were delivered within a few hours of slaughter and were prepared immediately after delivery. Four osteochondral samples (diameter = 6 mm) were prepared from lateral upper quadrants of each patella except for one, from which only two samples were prepared. Taken together, a total of 54 samples were used in this study. Half of the samples of the patella

(altogether  $n = 27$ ) were treated with threose to increase the collagen cross-linking while the other half (altogether  $n = 27$ ) were used as control samples. Control samples were incubated for 7 days at 37°C in cell culture medium (DMEM low glucose 1 g/l, Lonza Cologne AG, Belgium) with 100 unit/ml penicillin, 100 µg/ml streptomycin, 10 mM HEPES buffer solution (HEPES, EuroClone S.p.A., Italy), 1 mM L-glutamine and 10 mM vitamin C. Cross-link formation was induced by adding 100 mM of threose (Sigma Aldrich Co., St. Louis, Missouri) into the incubation media of the threose group. Finally, the samples were split in half, and one half was fixed in 10% formalin for 48 h, decalcified with ethylene diamine tetraacetic acid, dehydrated and immersed in liquid paraffin. Subsequently, the paraffin was hardened by cooling. The other half was used for biochemical reference analyses.

### 2.2 Infrared Microspectroscopic Measurements

For FTIR microspectroscopy, 5-µm-thick sections were cut with a microtome. For dewaxing, the sections were immersed in series of xylene baths (3 × 5 min) and subsequently washed with ethanol. The sections were placed onto 2-mm-thick ZnSe windows and were let dry in room air. The sections were measured using the Perkin Elmer Spotlight 300 FTIR imaging system (Perkin Elmer, Shelton, Colorado) with spectral resolution of 4 cm<sup>-1</sup>, pixel size of 25 µm, and using four scans per pixel. A rectangular region of interest (ROI) extending from cartilage surface to cartilage–bone junction was imaged from each section (Fig. 1). The ROI width was set to 400 µm, whereas the ROI height, which varied depending on the cartilage thickness, was on average ( $\pm$ SD) 1693  $\pm$  484 µm. On average, 1088 spectra were collected per sample. A dry air purge (Parker Balston, Haverhill, Massachusetts) was used to minimize the variation in the measurement conditions. The spectra of each section were first averaged to obtain one mean spectrum. Thereafter, resonant Mie scattering correction (RMieSC) algorithm<sup>18,19</sup> was used to remove scattering-related effects from the spectra. After RMieSC, the spectra were vector normalized. Second derivative spectra were calculated using Savitzky–Golay algorithm with 11 smoothing points. In this paper, we use the following definitions for spectral regions: the amide I (1590 to 1720 cm<sup>-1</sup>), the amide II (1500 to 1590 cm<sup>-1</sup>), the mixed region (1200 to 1500 cm<sup>-1</sup>),



**Fig. 1** A rectangular ROI extending from cartilage surface to cartilage–bone junction was imaged from each section. The ROI width was set to 400 µm, whereas the ROI height, which varied according to the cartilage thickness, was on average ( $\pm$ SD) 1693  $\pm$  484 µm. Therefore, the average size of the ROI was 16 × 68 pixels, which results in ~1088 spectra/section. These spectra were averaged to obtain an average spectrum for the section.

and the carbohydrate region (800 to 1200  $\text{cm}^{-1}$ ).<sup>20</sup> All data analysis was done in MATLAB (R2015a, MathWorks, Inc., Natick, Massachusetts).

### 2.3 Univariate Analysis

To evaluate the overall changes in the spectra due to the increase in cross-linking, the difference spectrum (mean spectrum of threose-treated samples minus mean spectrum of control samples) was calculated. In addition, Pearson's correlation coefficients between each variable (i.e., absorbance at each wavenumber) in the FTIR spectra of the samples and the concentrations of HP, LP, and Pent were calculated to find the relationship between the variables of the FTIR spectrum of AC and the cross-link concentrations.

### 2.4 Partial Least Squares Regression

PLS regression, a method based on latent variables, was used to predict cross-link concentrations from FTIR spectra. Spectral region of 800 to 1800  $\text{cm}^{-1}$  was used for PLS regression models. Root-mean-square error of cross validation (RMSECV) was used as the term to be minimized in PLS regression. Leave-one-out cross-validation was used when deciding the number of PLS components for the full spectrum PLS regression model. All models were validated using leave-one-out cross-validation.

### 2.5 Backward Iterative Partial Least Squares Regression

Backward iterative partial least squares (biPLS) is a variable selection algorithm that evaluates the importance of different spectral regions in PLS regression problems.<sup>21</sup> The spectra were divided into 25 equal-sized spectral windows (20  $\text{cm}^{-1}$  window size). The effect of each spectral window to the model was tested by building the model without the window. The window whose removal improved the model the most was removed in each cycle. The biPLS procedure was continued until there was only one window remaining. Finally, the combination of spectral windows that produced the lowest RMSECV was selected as the best biPLS model. The whole procedure was conducted for PLS components from 1 to 10. The optimal number of components for biPLS model was chosen based on the minimum of RMSECV.

### 2.6 Backward Iterative Partial Least Squares Discriminant Analysis

PLS is most often used for regression analysis. However, PLS can also be used for classification tasks. For this purpose, a dummy variable describing the categories is used as the response variable. biPLS-DA algorithm was applied to second derivative spectra to classify the samples into control and threose-treated groups. The parameters of biPLS-DA were set to be the same as for biPLS. A binary dummy variable with a value of zero and one for control and threose-treated groups, respectively, was used as the response variable. The samples with predicted class values of under 0.5 were assigned to group 0 (control group) and with values of over 0.5 to group 1 (threose-treated group). The optimal number of PLS components was found experimentally. The model was validated using leave-one-out cross-validation.

### 2.7 Genetic Algorithm for Wavenumber Selection

Genetic algorithms (GAs) are optimization methods that mimic the process of natural selection. GA was used in two ways: directly to FTIR spectra and in combination with biPLS (biPLS-GA). In biPLS-GA, biPLS was first used to remove 18 out of 25 spectral windows to reduce the number of variables before running GA. GA was then applied to the variables that were left after biPLS.

In GA, a chromosome is a binary vector with a size that corresponds to the number of variables in the spectrum. Variables of the chromosome, genes, indicate whether (1) or not (0) the corresponding variable in the spectrum is selected to the model. The GA is run for a predefined number of generations. In the beginning of GA, a population of randomly generated chromosomes is created. A PLS regression model is built using each of the chromosomes of the population, and RMSECV is used to evaluate the performance of the variables selected by each chromosome. The best chromosome is included in the next population, whereas the other chromosomes of the population are obtained by recombining the chromosomes of the earlier population by using cross-over of chromosomes and mutation of genes. Final generation contains the best chromosome of a single GA run. GA is run multiple times, and the importance of individual variables is evaluated based on their selection frequency in the best chromosome. In this study, GA was run for 50 times. Finally, the variables are added to the final model according to their selection frequency. The variable combination that results in the smallest RMSECV is chosen as the final model. The maximum number of PLS components was limited to 10. Other parameters of GA are summarized in Table 1. A more complete description of the algorithms can be found in earlier studies.<sup>22,23</sup>

### 2.8 Competitive Adaptive Reweighted Sampling Partial Least Squares Regression

Competitive adaptive reweighted sampling (CARS)<sup>24</sup> is a variable selection that uses the absolute values of regression coefficients as estimates for the importance of each variable. An exponentially decreasing function is used to define how many variables are retained in each of the sampling runs. The parameters of the exponentially decreasing function are selected so that all variables are retained in the first sampling run, whereas only two variables are left in the final sampling run. In every iteration, the least important variables are removed (variables with the smallest absolute regression coefficient values). A new calibration model is then built using the remaining

**Table 1** Parameters used in the GA for selecting spectral variables.

Population size	50
Gene initialization probability	5%
Mutation method	One-point cross-over
Cross-over probability	80%
Mutation probability	1%
Number of generations	50
Response to be minimized	RMSECV of the prediction

variables, and the regression coefficients of the new model are used as new indicators of their importance. Removal of variables is continued for a predefined number of sampling runs. In this study, the number of sampling runs was set to 100. Finally, the best variable combination is searched and used for the final model, which is validated by leave-one-out cross-validation. A more detailed description of the algorithm can be found in an earlier study.<sup>24</sup> The RMSECV was first monitored as a function of the number of PLS components in CARS-PLS. However, a clear minimum was not found, as the RMSECV steadily decreased as the number of PLS components was increased. Therefore, the maximum number of PLS components was limited to 10. CARS-PLS analyses were conducted in MATLAB using libPLS toolbox.<sup>25</sup>

## 2.9 Biochemical Analysis

A more detailed description of the biochemical analysis is found in previous studies.<sup>16,17,26</sup> Briefly, the amount of the collagen specific amino acid, hydroxyproline, was measured spectrophotometrically to estimate the total amount of collagen.<sup>27</sup> HPLC was used to separate hydroxylysyl pyridinoline (HP), lysyl pyridinoline (LP), and pentosidine (Pent).<sup>28</sup> Pure compounds of HP, LP, and Pent with known concentrations were used as standards and for converting the measurement results into concentration values. The results are expressed as mol of cross-link per mol of collagen. The HPLC system used consisted of a Quaternary Gradient Pump unit, PU-2089 Plus, Intelligent AutosamplerAS-2057 Plus, and Intelligent Fluorescence Detector, FP-2020 by Jasco (Jasco Scandinavia AB, Mölndal, Sweden). Data were analyzed using Jasco Chrompass software (Jasco, Sweden). The LiChroCART<sup>®</sup> 125-4 column was from Merck Hitachi (Merck KGaA, Darmstadt, Germany)

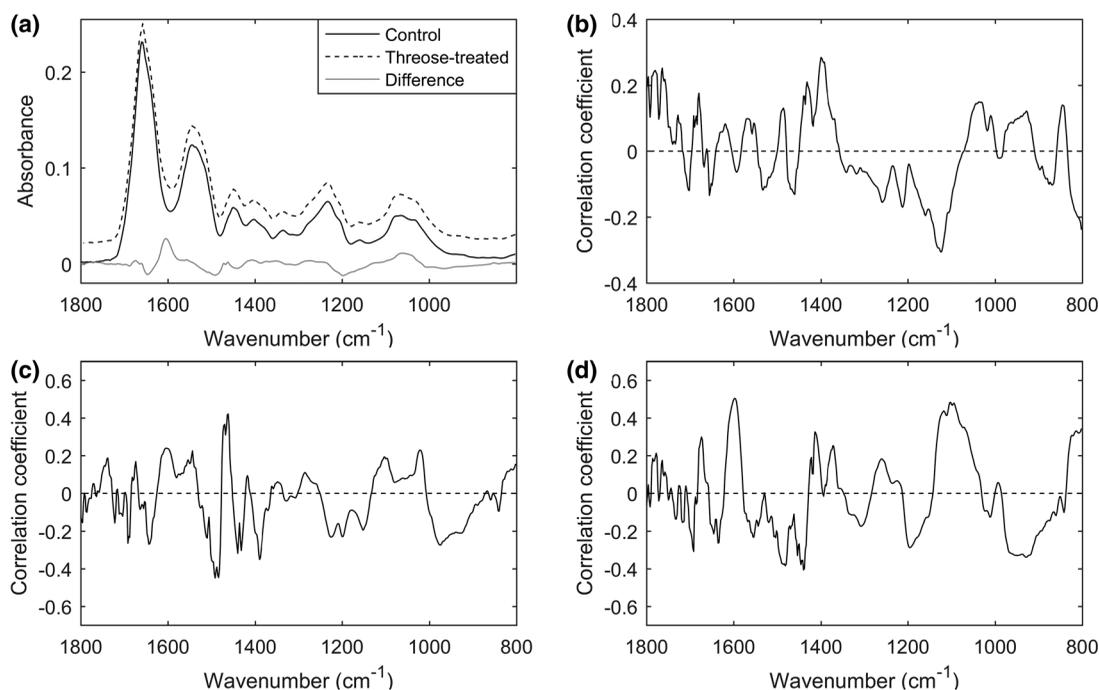
## 2.10 Statistical Analysis

The values predicted by the PLS models were compared to the reference information by using the Pearson's correlation analysis. Furthermore, the Spearman's correlation analysis was also conducted as the Anderson-Darling test revealed that LP and Pent values did not follow the normal distribution. The limit for statistically significant correlation was set to  $p < 0.05$ . All statistical analyses were conducted using MATLAB (R2015a, MathWorks, Inc.).

## 3 Results

### 3.1 Difference Spectrum and Univariate Analysis

Mean absorbance spectra of control and threose-treated samples are shown in Fig. 2(a). The difference spectrum, i.e., mean spectrum of threose-treated samples ( $n = 27$ ) minus mean spectrum of control samples ( $n = 27$ ) shows a negative peak in the amide I (1700 to 1590  $\text{cm}^{-1}$ ) and a positive peak in the carbohydrate region (1000 to 1100  $\text{cm}^{-1}$ ). Pearson's correlation coefficients between the variables in the FTIR spectra of the samples and the concentrations of HP, LP, and Pent were calculated to find the best variables for univariate analysis. For HP, the highest absolute correlation coefficient is found in the carbohydrate region ( $\sim 1120 \text{ cm}^{-1}$ ) ( $r = -0.31$ ,  $p < 0.05$ ) [Fig. 2(b)]. Some samples were excluded from the LP ( $n = 14$ ) and Pent ( $n = 16$ ) correlation coefficient analysis as the biochemical analysis did not detect any LP or Pent cross-links in some of the control samples due to the limitations of the HPLC method. The correlation coefficients of LP [Fig. 2(c)] and Pent [Fig. 2(d)] follow the shape of the difference spectrum. The highest absolute correlation coefficient values are  $r = -0.45$  ( $p < 0.001$ ) in the amide II region



**Fig. 2** (a) Mean absorbance spectra of control samples (solid black line) and threose-treated samples (dashed black line). There is an added offset between the mean spectra for better clarity. The difference spectrum (multiplied by a factor of five for better visualization) is shown as a solid gray line. Pearson's correlation coefficients between the variables in the FTIR spectra of the samples and the cross-link concentrations are shown for (b) HP, (c) LP, and (d) Pent.

and  $r = 0.50$  ( $p < 0.001$ ) in the amide I and the carbohydrate regions for LP and Pent, respectively.

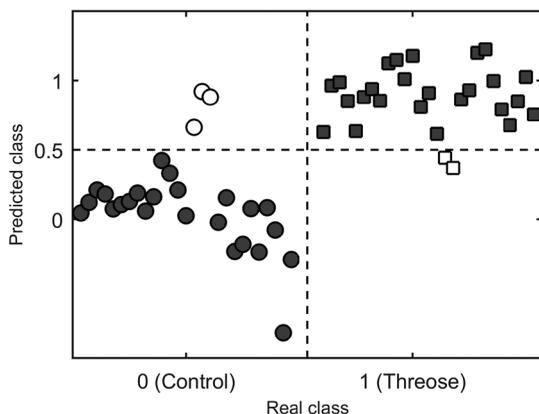
### 3.2 Backward Iterative Partial Least Squares Discriminant Analysis

BiPLS-DA was applied to second derivative spectra for separating the threose-treated samples from intact ones. Two PLS components were used for classification. The sensitivity (true positive rate) and specificity (true negative rate) for threose-treated samples were 92.5% and 88.9%, respectively. The accuracy (percentage of correctly classified samples) was 90.7%. The class predictions are shown in Fig. 3.

### 3.3 Partial Least Squares Regression

The biochemical analysis did not detect any LP or Pent cross-links in some of the control samples. These samples were excluded from the PLS regression models when predicting LP or Pent concentrations. The results of raw spectrum [Fig. 4(a)] models are summarized in Table 2 and Fig. 4. In general, equal or better results were obtained when using second derivative spectra [Fig. 5(a)] instead of raw spectra. Therefore, we focus on the results obtained using second derivative spectra.

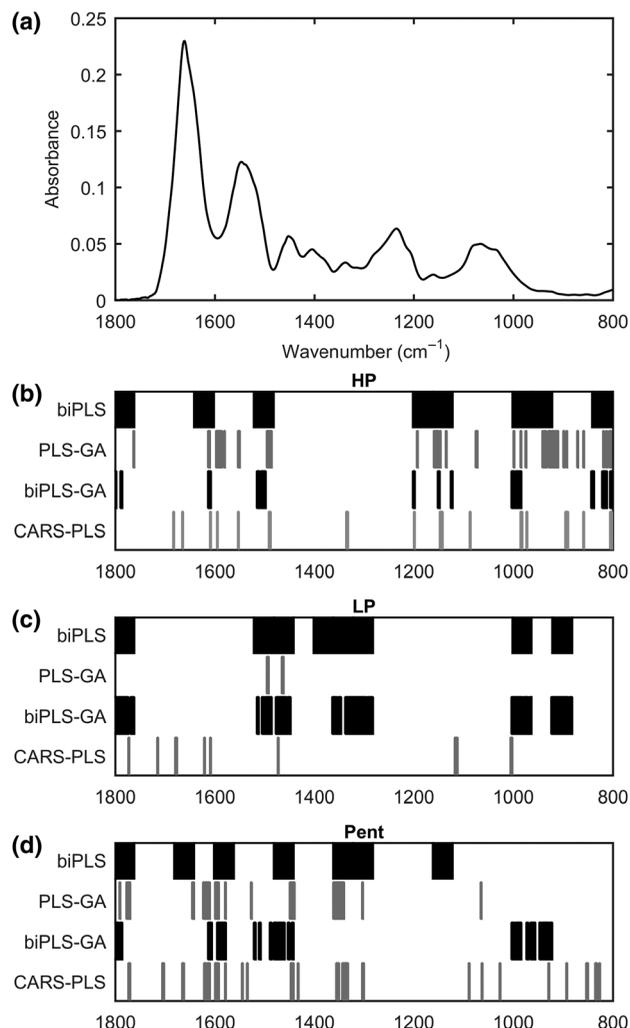
Four PLS components produced the best model ( $r = 0.41$ ,  $p < 0.01$ ) when the full spectrum (800 to 1800  $\text{cm}^{-1}$ , 500 variables) was used for prediction of the HP concentration, whereas the other models used 9 or 10 PLS components. Variable selection using biPLS improved the correlation coefficient to  $r = 0.70$  ( $p < 0.001$ ). The model utilized portions of the amide II region, the mixed region, and the carbohydrate region [Fig. 5(b)]. The correlation coefficient between the HP concentration and the concentration predicted by the PLS-GA regression model was  $r = 0.62$  ( $p < 0.001$ ). PLS-GA selected variables quite evenly throughout the spectral region of 800 to 1700  $\text{cm}^{-1}$  [Fig. 5(b)]. BiPLS-GA utilized mainly the carbohydrate region and, to a smaller extent, the amide II and the mixed regions [Fig. 5(b)]. The correlation coefficient between the HP concentration and the concentration predicted by the biPLS-GA regression model was  $r = 0.83$  ( $p < 0.001$ ). CARS-PLS



**Fig. 3** The predicted class memberships obtained from biPLS-DA. The control and the threose-treated samples are marked with circles and squares, respectively. Correctly and incorrectly classified samples are displayed with solid and open markers, respectively. Two PLS components were used in the final biPLS-DA model to separate the control and threose-treated groups from each other. The samples were classified with an accuracy of 90.7%.

selected variables throughout the spectral region of 800 to 1600  $\text{cm}^{-1}$  [Fig. 5(b)] and resulted in a correlation coefficient of  $r = 0.83$  ( $p < 0.001$ ) [Figs. 6(a) and 6(d)].

When predicting the LP concentration, two PLS components produced the best model ( $r = 0.20$ ,  $p > 0.05$ ) in the full spectrum (800 to 1800  $\text{cm}^{-1}$ , 500 variables) model. Also BiPLS, PLS-GA and biPLS-GA used only two PLS components. BiPLS improved the correlation coefficient to  $r = 0.47$  ( $p < 0.01$ ). The model relied on the carbohydrate and mixed regions [Fig. 5(c)]. The correlation coefficient between the LP concentration and the concentration predicted by the PLS-GA regression model was  $r = 0.52$  ( $p < 0.001$ ). PLS-GA selected variables from all regions [Fig. 5(c)]. In addition to the regions used in biPLS, BiPLS-GA used the additional spectral window of 800 to 840  $\text{cm}^{-1}$  [Fig. 5(c)]. The correlation coefficient between the LP concentration and the concentration predicted by the biPLS-GA regression model was  $r = 0.51$  ( $p < 0.001$ ). CARS-PLS used nine PLS components for prediction of the LP concentration. CARS-PLS used variables mainly from the amide I, the amide II, and the carbohydrate regions [Fig. 5(c)]. The correlation coefficient between the LP concentration and



**Fig. 4** (a) The mean raw spectrum of the dataset, and the variables selected from the spectra by the studied variable selection algorithms for the prediction of (b) HP, (c) LP, and (d) Pent concentrations, respectively.

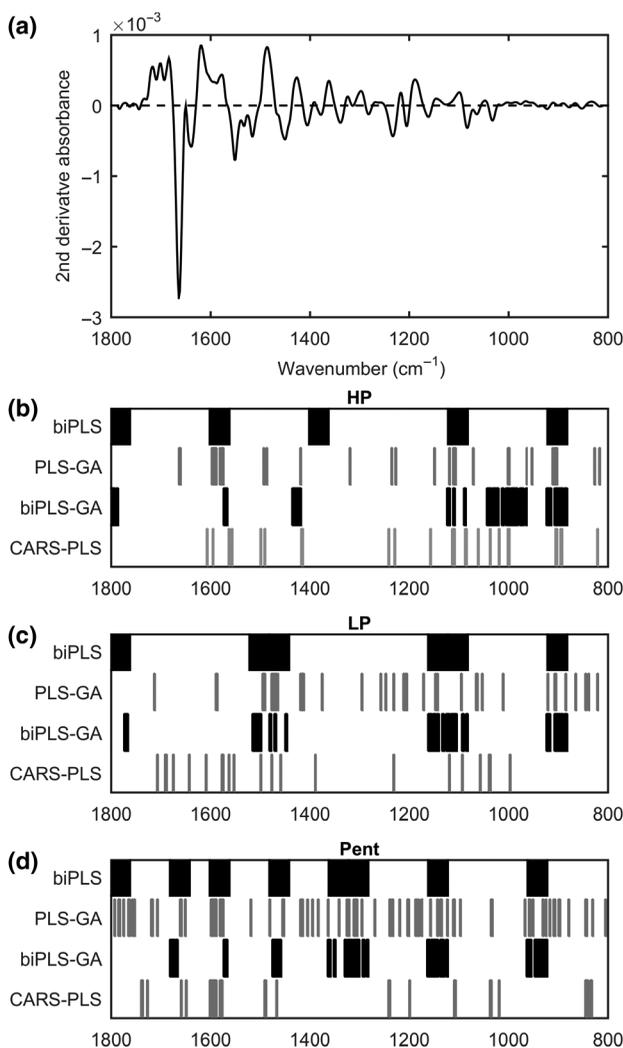
**Table 2** The results in predicting cross-link (HP, LP, and Pent) concentrations using PLS regression with different variable selection techniques from the raw spectrum. The number of PLS components (Comp), the number of used variables (Var), the Pearson's correlation coefficient ( $r$ ), and the Spearman's correlation coefficient ( $r_s$ ) between the predicted values and the reference values, and RMSECV are shown for each PLS method.

	HP					LP					Pent				
	Comp	Var	$R$	$r_s$	RMSECV	Comp	Var	$r$	$r_s$	RMSECV	Comp	Var	$r$	$r_s$	RMSECV
PLS	5	500	0.30 <sup>†</sup>	0.36 <sup>**</sup>	0.3672	1	500	0.02	0.02	0.0410	6	500	0.51 <sup>***</sup>	0.45 <sup>**</sup>	0.0072
biPLS	8	160	0.57 <sup>***</sup>	0.58 <sup>***</sup>	0.3114	2	160	0.40 <sup>**</sup>	0.32 <sup>†</sup>	0.0359	8	220	0.75 <sup>***</sup>	0.71 <sup>***</sup>	0.0053
GA	10	48	0.63 <sup>***</sup>	0.64 <sup>***</sup>	0.2905	2	2	0.46 <sup>**</sup>	0.42 <sup>**</sup>	0.0344	7	42	0.82 <sup>***</sup>	0.79 <sup>***</sup>	0.0045
biPLS-GA	8	26	0.68 <sup>***</sup>	0.69 <sup>***</sup>	0.2695	2	88	0.50 <sup>***</sup>	0.44 <sup>**</sup>	0.0335	7	43	0.84 <sup>***</sup>	0.81 <sup>***</sup>	0.0042
CARS-PLS	9	23	0.72 <sup>***</sup>	0.73 <sup>***</sup>	0.2563	7	18	0.72 <sup>***</sup>	0.73 <sup>***</sup>	0.0273	7	19	0.87 <sup>***</sup>	0.83 <sup>***</sup>	0.0038

<sup>†</sup> $p < 0.05$  (Pearson's or Spearman's correlation analysis)

<sup>\*\*</sup> $p < 0.01$  (Pearson's or Spearman's correlation analysis)

<sup>\*\*\*</sup> $p < 0.001$  (Pearson's or Spearman's correlation analysis)



**Fig. 5** (a) The mean second derivative spectrum of the dataset, and the variables selected from the second derivative spectra by the studied variable selection algorithms for the prediction of (b) HP, (c) LP, and (d) Pent concentrations, respectively.

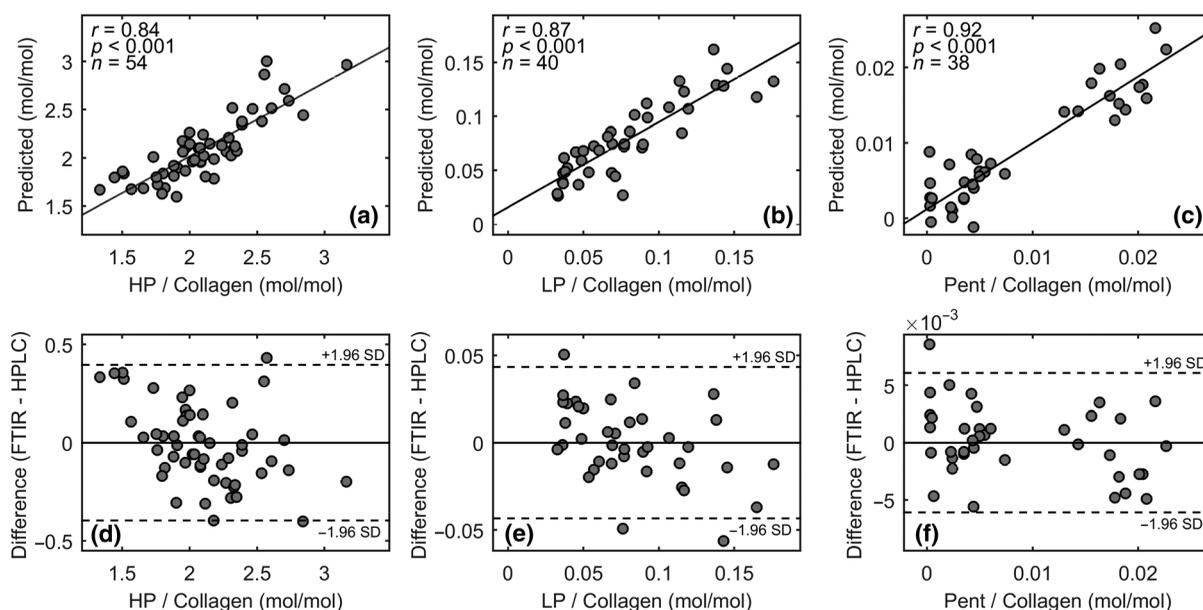
the CARS-PLS model was  $r = 0.87$  ( $p < 0.001$ ) [Figs. 6(b) and 6(e)].

Nine PLS components produced the best model ( $r = 0.66$ ,  $p < 0.001$ ) in the full spectrum ( $800$  to  $1800$   $\text{cm}^{-1}$ , 500 variables) model when predicting the Pent concentration. In cases of biPLS and biPLS-GA, the selected spectral regions were very similar [Fig. 5(d)]. PLS-GA selected a relatively high number of variables throughout the whole spectrum [Fig. 5(d)]. The correlation coefficients between the Pent concentration and biPLS, PLS-GA and biPLS-GA were  $r = 0.84$  ( $p < 0.001$ ),  $r = 0.90$  ( $p < 0.001$ ) and  $r = 0.93$  ( $p < 0.001$ ), respectively. CARS-PLS selected variables from all regions of the spectrum [Fig. 5(d)] and resulted in a correlation coefficient of  $r = 0.92$  ( $p < 0.001$ ) [Figs. 6(c) and 6(f)]. The results of all second derivative spectrum-based PLS regression models are summarized in Table 3.

## 4 Discussion

In the present study, the capabilities of FTIR microspectroscopy to determine cross-links in articular cartilage were investigated. The first aim of this study was to separate intact and threose-treated samples from each other based on their FTIR spectra. Mean spectra of control and threose groups displayed only minor differences [Fig. 2(a)]. However, the difference spectrum revealed negative peaks in the amide I ( $1700$  to  $1590$   $\text{cm}^{-1}$ ) and the amide II ( $1590$  to  $1450$   $\text{cm}^{-1}$ ) regions and a positive peak in the carbohydrate region ( $1000$  to  $1100$   $\text{cm}^{-1}$ ). Even though the absolute differences between the groups were small, biPLS-DA separated the threose-treated and the control samples from each other with a reasonable accuracy, as the sensitivity and specificity were 92.6% and 88.9%, respectively, i.e., five samples were incorrectly classified. These few misjudged samples further indicate that the spectroscopic features caused by cross-linking are minute, which makes the FTIR spectroscopic cross-link analysis challenging.

The second aim was to quantify the cross-link concentrations in articular cartilage using FTIR microspectroscopy. For this purpose, PLS regression combined with different variable selection methods (biPLS, PLS-GA, biPLS-GA, and CARS-PLS) were used. The best results were obtained with CARS-PLS, which gave accurate results for all studied cross-links. Furthermore, as expected, the use of the second derivative spectra



**Fig. 6** Scatter plots between the biochemically determined concentrations and the concentrations predicted by CARS-PLS for (a) LP, (b) HP, and (c) Pent cross-links. (d-f) Corresponding Bland-Altman plots are shown below each correlation plot.

**Table 3** The results in predicting cross-link (HP, LP, and Pent) concentrations using PLS regression with different variable selection techniques from the second derivative spectrum. The number of PLS components (Comp), the number of used variables (Var), the Pearson's correlation coefficient ( $r$ ), and the Spearman's correlation coefficient ( $r_s$ ) between the predicted values and the reference values, and RMSECV are shown for each PLS method.

	HP					LP					Pent				
	Comp	Var	$r$	$r_s$	RMSECV	Comp	Var	$r$	$r_s$	RMSECV	Comp	Var	$r$	$r_s$	RMSECV
PLS	4	500	0.41**	0.44***	0.3474	2	500	0.20	0.17	0.0399	9	500	0.66***	0.59***	0.0062
biPLS	10	100	0.70***	0.68***	0.2654	2	120	0.47**	0.27	0.0343	7	160	0.84***	0.79***	0.0042
GA	9	32	0.62***	0.62***	0.2923	2	41	0.52***	0.40*	0.0332	5	81	0.90***	0.79***	0.0033
biPLS-GA	10	50	0.83***	0.79***	0.2050	2	44	0.51***	0.42**	0.0333	9	50	0.93***	0.83***	0.0028
CARS-PLS	10	26	0.84***	0.81***	0.2002	9	21	0.87***	0.87***	0.0191	7	31	0.92**	0.81***	0.0031

\* $p < 0.05$  (Pearson's or Spearman's correlation analysis)

\*\* $p < 0.01$  (Pearson's or Spearman's correlation analysis)

\*\*\* $p < 0.001$  (Pearson's or Spearman's correlation analysis)

improved the results. The scatter plots (Fig. 6) showed a significant amount of deviation around the regression line. This is somewhat expected, since these cross-links form only a small fraction of the mass of articular cartilage dry matrix. As the FTIR spectrum of articular cartilage is a sum spectrum of its all constituents, it is dominated by the collagen and the proteoglycan spectra. Nevertheless, the Bland-Altman plots did not display any obvious bias in the prediction of cross-link concentrations, as the absolute error remained similar throughout the whole range of reference values in case of all cross-links.

All variable selection methods improved the predictions compared to the full spectrum models. This result indicates that it is essential to use variable selection algorithms with PLS regression models to obtain the best results. CARS-PLS and biPLS-GA performed equally in prediction of HP and Pent

concentrations. Surprisingly, CARS-PLS clearly outperformed the other studied methods in prediction of LP concentration. In general, the variables selected by PLS-GA and CARS-PLS were more deviated throughout the whole spectrum than the variables selected by biPLS and biPLS-GA. The obvious reason for this is that instead of individual variables, the biPLS estimates the effect of spectral windows for the prediction. PLS-GA and CARS-PLS, which do not use the windowed approach, selected variables mostly from the same regions for the prediction of HP concentration, but bigger differences are seen in prediction of LP and Pent concentrations. The windowed approach of biPLS may miss some important variables from narrow regions. On the other hand, CARS-PLS, which produced the best results in this study, is able to select important individual variables without being limited to certain spectral windows. The probability of missing important variables could be reduced by studying the

results of multiple biPLS algorithm runs with variable sizes of windows.<sup>21</sup> This could improve the performance of biPLS and biPLS-GA compared to the fixed spectral window width approach used in this study.

Glycation of type I collagen has been shown to result in a strong increase in the carbohydrate region absorbance.<sup>10</sup> In a recent study, the ratio of the integrated absorbance of the carbohydrate region to the amide I absorbance was used to study the differences in AGEs between control and ribose-treated articular cartilage samples.<sup>13</sup> In the present study, the maximum correlation coefficient between the Pent concentration and the variables in the carbohydrate region was  $r = 0.50$ . The samples in both groups of this study were similar in terms of the collagen and the proteoglycan contents, as the increased cross-linking was achieved by a threose treatment. Since the glycosaminoglycans of proteoglycans display strong absorption in the carbohydrate region, it is likely that changes in the proteoglycan content of articular cartilage introduce larger spectroscopic changes than the differences in AGE components. Therefore, it is not encouraged to use the absorbance of the carbohydrate region as a marker of AGE components if significant differences exist in the proteoglycan content between the studied samples.

Ideally, the training dataset for multivariate models should contain a relatively even distribution of reference values over the investigated range. However, this was not fulfilled in the cases of LP and Pent cross-links, as there were very little LP or Pent cross-links in the control samples. In reality, it is likely that these samples also contain LP and Pent cross-links, but the HPLC method was not sensitive enough to detect the small amounts. For this reason, these particular samples were excluded from LP and Pent analyses. Nevertheless, two distinct groups in Pent scatter plot are still seen after exclusion of these samples. The results in the prediction of Pent cross-links would likely be improved if a sample set with a more even range of cross-link concentrations was used.

The samples of this study were formalin-fixed, which preserves the samples by creating cross-links in the tissue. Formalin fixation of tissues alters the amide I band position compared to fresh tissue.<sup>29</sup> Another consequence of the sample processing is that lipids are removed from the samples.<sup>29</sup> Therefore, the built models may not work well with fresh tissue that still contains lipids. However, as standard sample processing protocols were used for all samples of the study, the effect of sample processing to spectra can be regarded as equivalent for all samples. Earlier studies have also shown that formalin fixation does not change the HP, LP, or Pent concentrations in articular cartilage.<sup>30,31</sup> Thus, it is feasible to study the cross-link concentrations of formalin-fixed tissues.

In conclusion, cross-linked articular cartilage samples were successfully separated from the control samples using FTIR microspectroscopy. Furthermore, the concentrations of enzymatic and nonenzymatic cross-links were successfully predicted from articular cartilage FTIR spectra. Unlike traditional HPLC analysis, FTIR microspectroscopy can be conducted nondestructively on standard histological tissue sections. In principle, FTIR microspectroscopy also enables determination of distribution of cross-links within histological sections. However, validation of cross-link distributions would require conducting the reference analyses separately for different layers of cartilage. Furthermore, the methodology should be validated using preferably human articular samples from patients of different ages to

guarantee that changes in all compositional parameters are within normal physiological variation.

### Disclosure

The authors declare no conflicts of interest.

### Acknowledgments

The technical assistance by Kaisa-Leena Tulla (University of Jyväskylä) in HPLC analyses is acknowledged. The financial support from the Academy of Finland (Grant Nos. 268378 and 273571); Sigrid Juselius Foundation; European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 336267; the strategic funding of the University of Oulu; and the strategic funding of the University of Eastern Finland is acknowledged. The funding bodies had no role in the study design, implementation, or writing of the manuscript.

### References

1. J. A. Buckwalter and H. J. Mankin, "Articular cartilage: tissue design and chondrocyte-matrix interactions," *Instrum. Course Lect.* **47**, 477–486 (1998).
2. N. Verzijl et al., "Effect of collagen turnover on the accumulation of advanced glycation end products," *J. Biol. Chem.* **275**(50), 39027–39031 (2000).
3. N. Verzijl et al., "Crosslinking by advanced glycation end products increases the stiffness of the collagen network in human articular cartilage: a possible mechanism through which age is a risk factor for osteoarthritis," *Arthritis Rheum.* **46**(1), 114–123 (2002).
4. N. P. Camacho et al., "FTIR microscopic imaging of collagen and proteoglycan in bovine cartilage," *Biopolymers* **62**(1), 1–8 (2001).
5. J. Yin, Y. Xia, and M. Lu, "Concentration profiles of collagen and proteoglycan in articular cartilage by Fourier transform infrared imaging and principal component regression," *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **88**, 90–96 (2012).
6. A. Hanifi et al., "Fourier transform infrared imaging and infrared fiber optic probe spectroscopy identify collagen type in connective tissues," *PLoS One* **8**(5), e64822 (2013).
7. L. Rieppo et al., "Optimal variable selection for Fourier transform infrared spectroscopic analysis of articular cartilage composition," *J. Biomed. Opt.* **19**(2), 027003 (2014).
8. L. Rieppo et al., "Infrared spectroscopic analysis of human and bovine articular cartilage proteoglycans using carbohydrate peak or its second derivative," *J. Biomed. Opt.* **18**(9), 097006 (2013).
9. L. Rieppo et al., "Fourier transform infrared spectroscopic imaging and multivariate regression for prediction of proteoglycan content of articular cartilage," *PLoS One* **7**(2), e32344 (2012).
10. R. Roy, A. Boskey, and L. J. Bonassar, "Processing of type I collagen gels using nonenzymatic glycation," *J. Biomed. Mater. Res. Part A* **93A**(3), 843–851 (2010).
11. M. Guilbert et al., "Probing non-enzymatic glycation of type I collagen: a novel approach using Raman and infrared biophotonic methods," *Biochimica et Biophysica Acta* **1830**(6), 3525–3531 (2013).
12. G. Birarda et al., "Synchrotron infrared imaging of advanced glycation endproducts (AGEs) in cardiac tissue from mice fed high glycemic diets," *Biomed. Spectrosc. Imaging* **2**(4), 301–315 (2013).
13. J. M. Fick et al., "In vitro glycation of articular cartilage alters the biomechanical response of chondrocytes in a depth-dependent manner," *Osteoarthritis Cartilage* **22**(10), 1410–1418 (2014).
14. M. Kim et al., "Fourier transform infrared imaging spectroscopic analysis of tissue engineered cartilage: histologic and biochemical correlations," *J. Biomed. Opt.* **10**(3), 031105 (2005).
15. R. H. Nagaraj and V. M. Monnier, "Protein modification by the degradation products of ascorbate: formation of a novel pyrrole from the Maillard reaction of L-threose with proteins," *Biochim. Biophys. Acta* **1253**(1), 75–84 (1995).
16. H. T. Kokkonen et al., "Computed tomography detects changes in contrast agent diffusion after collagen cross-linking typical to natural aging

- of articular cartilage," *Osteoarthritis Cartilage* **19**(10), 1190–1198 (2011).
17. K. A. Kulmala et al., "Diffusion of ionic and non-ionic contrast agents in articular cartilage with increased cross-linking—contribution of steric and electrostatic effects," *Med. Eng. Phys.* **35**(10), 1415–1420 (2013).
  18. P. Bassan et al., "Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples," *Analyst* **135**(2), 268–277 (2010).
  19. P. Bassan et al., "Resonant Mie scattering in infrared spectroscopy of biological materials—understanding the 'dispersion artefact'," *Analyst* **134**(8), 1586–1593 (2009).
  20. H. Najbjerg et al., "Monitoring cellular responses upon fatty acid exposure by Fourier transform infrared spectroscopy and Raman spectroscopy," *Analyst* **136**(8), 1649–1658 (2011).
  21. R. Leardi and L. Nørgaard, "Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions," *J. Chemometrics* **18**(11), 486–497 (2004).
  22. R. Leardi, "Application of genetic algorithm PLS for feature selection in spectral data sets," *J. Chemometrics* **14**(5–6), 643–655 (2000).
  23. L. Rieppo et al., "Prediction of compressive stiffness of articular cartilage using Fourier transform infrared spectroscopy," *J. Biomech.* **46**(7), 1269–1275 (2013).
  24. H. Li et al., "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration," *Anal. Chim. Acta* **648**(1), 77–84 (2009).
  25. H. Li, Q. Xu, and Y. Liang, "libPLS: an integrated library for partial least squares regression and discriminant analysis," *PeerJ PrePrints* **2**, e190v1 (2014).
  26. M. Kongsgaard et al., "Corticosteroid injections, eccentric decline squat training and heavy slow resistance training in patellar tendinopathy," *Scand. J. Med. Sci. Sports* **19**(6), 790–802 (2009).
  27. L. B. Creemers et al., "Microassay for the assessment of low levels of hydroxyproline," *Biotechniques* **22**(4), 656–658 (1997).
  28. R. A. Bank et al., "Sensitive fluorimetric quantitation of pyridinium and pentosidine crosslinks in biological samples in a single high-performance liquid chromatographic run," *J. Chromatogr. B Biomed. Sci. Appl.* **703**(1–2), 37–44 (1997).
  29. V. Zohdi et al., "Importance of tissue preparation methods in FTIR micro-spectroscopical analysis of biological tissues: 'traps for new users'," *PLoS One* **10**(2), e0116491 (2015).
  30. M. Abe et al., "The changes in crosslink contents in tissues after formalin fixation," *Anal. Biochem.* **318**(1), 118–123 (2003).
  31. K. W. Fishbein et al., "Effects of formalin fixation and collagen cross-linking on T2 and magnetization transfer in bovine nasal cartilage," *Magn. Reson. Med.* **57**(6), 1000–1011 (2007).

Biographies for the authors are not available.