DISSERTATIONS IN
FORESTRY AND
NATURAL SCIENCES

YEVGENIYA KOBRINA

Infrared Microspectroscopic Cluster Analysis of Bone and Cartilage

PUBLICATIONS OF THE UNIVERSITY OF EASTERN FINLAND
Dissertations in Forestry and Natural Sciences

UNIVERSITY OF
EASTERN FINLAND

YEVGENIYA KOBRINA

Infrared Microspectroscopic Cluster Analysis of Bone and Cartilage

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences
No 137

Academic Dissertation
To be presented by permission of the Faculty of Science and Forestry for public examination in the Auditorium L3 in Canthia Building at the University of Eastern Finland, Kuopio, on June, 17, 2014, at 12 o'clock noon.

Department of Applied Physics

Grano Oy
Kuopio, 2014
Editors: Prof. Pertti Pasanen, Prof. Pekka Kilpeläinen,
Prof. Kai Peiponen, Prof. Matti Vornanen
Distribution:
Eastern Finland University Library / Sales of publications
P.O.Box 107, FI-80101 Joensuu, Finland
tel. +358-50-3058396
http://www.uef.fi/kirjasto

ISBN: 978-952-61-1457-6 (nid.)
ISBN: 978-952-61-1458-3 (PDF)
ISSNL: 1798-5668
ISSN: 1798-5668
ISSN: 1798-5676 (PDF)

Author's address: University of Eastern Finland
Department of Applied Physics
P.O.Box 1627
70211 KUOPIO
FINLAND
email: ye.kobrina@gmail.com

Supervisors: Professor Jukka Jurvelin, Ph.D.
University of Eastern Finland
Department of Applied Physics
P.O.Box 1627
70211 KUOPIO
FINLAND
email: jukka.jurvelin@uef.fi

Associate Professor Hanna Isaksson, Ph.D.
Lund University
Division of Solid Mechanics
P.O.Box 118
221 00 LUND
SWEDEN
email: hanna.isaksson@solid.lth.se

Professor Markku Hauta-Kasari, Ph.D.
University of Eastern Finland
School of Computing
P.O.Box 111
80101 JOENSUU
FINLAND
email: markku.hauta-kasari@uef.fi

Reviewers: Adjunct Associate Lisa Miller, Ph.D.
Stony Brook University
Department of Chemistry and Biomedical Engineering
P.O.Box 5000
NY 11973, Upton
USA
email: lmiller@bnl.gov

Professor Yang Xia, Ph.D.
Oakland University
Department of Physics
P.O.Box 4487
MI 48309, Rochester
USA
email: xia@oakland.edu

Opponent: Professor Ela Claridge
School of Computer Science
The University of Birmingham
Edgbaston Birmingham, B15 2TT
United Kingdom
email: E.Claridge@cs.bham.ac.uk

Abstract

The composition and structure of bone and articular cartilage (AC) change during development, and in response to physical exercise or pathology. Fourier-transform infrared microspectroscopy (FTIR-MSP) has a great potential for determining the molecular composition, concentration and spatial distribution of biochemical compounds in bone and AC. Usually, univariate parameters are used to analyze infrared spectra. However, their use is restricted to some tissue components due to the complex overlapping infrared spectra.

Unsupervised multivariate cluster analysis plays an important role in many areas of science and has been used successfully for classification and discrimination of various biological tissues, fluids and cells based on spectral differences. This technique can be used to group samples objectively based on their spectral features, and it does not rely on knowledge or expertise of the operator. Earlier, such multivariate methods as principal component regression and partial least squares were successfully applied to FTIR-MSP of AC to estimate a true quantitative concentration and spatial variations of its main components using pure chemical spectral libraries. However, no previous studies had investigated compositional changes in bone and AC with FTIR-MSP and unsupervised cluster analysis.

This thesis work demonstrates the feasibility of clustering to capture subtle differences in infrared spectra of bone and AC. In particular, it could reveal the variations in different age groups of maturing bone, in normal and repaired AC with good accuracy. Further, it could be used to identify histological zones in intact AC. The results also demonstrated the improvements in performance of fuzzy clustering in comparison to "hard" and hierarchical clustering methods. This method calculates the

strength of connection between each spectrum and obtained clusters. Thus, provided that the boundaries between tissue regions and disease stages are not sharp, fuzzy clustering allows meticulous investigation of continuously changing features. This study also introduces an algorithm for processing FTIR-MSP using cluster analysis.

Taken together, this thesis provides a basis for future studies towards employing multivariate cluster analysis and FTIR-MSP in medical diagnostics of musculoskeletal diseases, such as osteoporosis and osteoarthritis.

National Library of Medicine Classification: QY 90, WE 200, WE 300

Medical Subject Headings: Bone and Bones; Cartilage, Articular; Aging; Spectroscopy, Fourier Transform Infrared; Spectrum Analysis; Cluster Analysis; Multivariate Analysis; Classification; osteoporosis/diagnosis; osteoarthritis/diagnosis

Yleinen suomalainen asiasanasto: luu; nivelrusto; ikääntyminen; spektroskopia; spektrianalyysi; klusterianalyysi; diagnostiikka

To Andrey and Veronika

Acknowledgments

This study was carried out in the Biophysics of Bone and Cartilage (BBC) research group in the University of Eastern Finland during the years 2009–2013.

I would like to thank my principal supervisor Professor Jukka Jurvelin for the opportunity to work in his world-famous research group and for his understanding, support and professional guidance throughout all my work. I want to offer my sincere thanks to my second supervisor, Associate Professor Hanna Isaksson, Ph.D., for her invaluable encouragement and endless support. I appreciate the help and contribution given by my third supervisor Professor Markku Hauta-Kasari. I am grateful to Associate Professor Simo Saarakkala, who expressed an interest in the discrimination methods I used in my Master's work and noticed a good scientific researcher in me.

I would like to express my deepest gratitude to all of my co-authors for their contributions. I would also like to acknowledge the staff of the Institute of Biomedicine, Anatomy and BioMater Centre for their efforts on the sample preparation, measurements and all related issues. I would like to thank Professor Ilkka Kiviranta for providing me a workspace and support during the time I worked in Helsinki.

The study was funded by the University of Eastern Finland, National Doctoral Program of Musculoskeletal Disorders and Biomaterials and Emil Aalto foundation, whose support is acknowledged.

I am grateful to the official reviewers of this thesis, Adjunct Associate Lisa Miller and Professor Yang Xia, Ph.D., for their constructive criticism and suggestions to improve the thesis. I would also like to thank Ewen MacDonald for linguistic review.

I spent most of my time working outside the university and, unfortunately, I spent very little time with my colleagues at the BBC group. I thank them for their friendliness, support and understanding during the times we met. Especially, I am grateful to my spectroscopy brothers – Drs Lassi Rieppo and Mikael Turunen, who taught me basics of FTIR microspectroscopy and guided me throughout my laboratory experiments. Special thanks also to Dmitry Semenov for his support and for the long effective discussions on spectroscopy and clustering.

Finally, I wish to thank my mother Lyudmila and grandmother Margarita for their support and love they have given to me during my whole life. My deepest gratitude goes to my loving husband Andrey, who never stopped encouraging me and would not let me lose my faith. And our lovely daughter Veronika, the sun of our life, thank you for the happiness you bring to our lives.

Espoo, May 2014,

Yevgeniya Kobrina

ABBREVIATIONS

ABC	absorption spectra
AC	articular cartilage
ACI	autologous chondrocytes implantation
AI	amide I
AII	amide II
BMD	bone mineral density
CHO	carbohydrate
CG	collagen gel repair
CO2	carbon dioxide
DA	discriminant analysis
DZ	deep zone
DF	discriminant function
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
FC	fibrocartilage
FCM	Fuzzy c-means clustering
FTIR	Fourier transform infrared
FTIR-MSP	Fourier transform infrared microspectroscopy
GAGs	glycosaminoglycans
HCA	hierarchical clustering analysis
IFOP	infrared fiber optic probe
LDA	linear discriminant analysis
MRI	magnetic resonance imaging
MSE	mean squared error
MZ	middle zone
OA	osteoarthtosis
OCT	optical coherence tomography
OP	osteoporosis
PC	principal component
PCA	principal component analysis
PCR	principal component regression
Ph	phosphate
PLM	polarized light microscopy
PLS	partial least squares
PMMA	polymethyl methacrylate

PG	proteoglycan
ROI	region of interest
SNR	signal-to-noise ratio
SSE	sum of squared errors
SZ	superficial zone
SB	subchondral bone
TR	transmittance spectra

SYMBOLS AND NOTATIONS

k	number of clusters
l	number of iterations
L	partition
n	number of objects
p	dimensionality of the vector, data
R	Rand index

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on data presented in the following articles, referred to by the Roman numerals I-IV.

I Kobrina Y, Isaksson H, Sinisaari M, Rieppo L, Brama PA, van Weeren R, Helminen HJ, Jurvelin JS, Saarakkala S. Infrared spectroscopy reveals both qualitative and quantitative differences in equine subchondral bone during maturation, *Journal of Biomedical Optics 15(6): 067003, 2010.*

II Kobrina Y, Turunen M, Saarakkala S, Hauta-Kasari M, Jurvelin JS, Isaksson H. Cluster analysis of infrared spectra of rabbit cortical bone samples during maturation and growth, *Analyst 135(12): 3147-55, 2010.*

III Kobrina Y, Rieppo L, Jurvelin JS, Saarakkala S, Isaksson H. Clustering of infrared spectra reveals histological zones in intact articular cartilage, *Osteoarthritis and Cartilage 20(5):460-8, 2012.*

IV Kobrina Y, Rieppo L, Saarakkala S, Pulkkinen HJ, Tiitu V, Valonen P, Kiviranta I, Jurvelin JS, Isaksson H, Pulkkinen. Cluster analysis of infrared spectra can differentiate intact and repaired articular cartilage, *Osteoarthritis and Cartilage 21(3): 462-9, 2013.*

AUTHOR'S CONTRIBUTION

The author of this thesis is the principal author in the studies I-IV. The author has contributed to the development of spectral analysis techniques and has carried out spectroscopic measurements for studies III and IV, and analyses. The multivariate methods used in the studies and their applications were proposed by the author.

Contents

1 Introduction

The composition and structure of biological tissues change with growth, maturation, disease, trauma and physical activity [1]. Bone and articular cartilage (AC) are connective tissues, which cooperate together to provide mechanical strength and support for body motion. Each tissue has a unique structure and composition. Characterization and prediction of tissue behavior under specific circumstances are important in the diagnostics and monitoring of the tissue health.

AC cushions the bones in joints, allowing the joints to move smoothly without pain. In general, bone and AC consist of cells and extracellular matrix (ECM), which is secreted by the cells. ECM comprises protein fibers (collagen) and ground substance, containing proteoglycans (PGs) and hyaluronic acid [2, 3].

Bone matrix is mineralized (60% of dry weight) by calcium phosphate that makes bone a hard tissue, providing it with rigidity and compressive strength [4]. The organic matrix is dominated by type I collagen fibers (40% of dry weight) and this confers resilience to bone. There are four different types of cells in bone that synergistically support tissue development, growth and remodeling [4]. The blood supply regulates the bone life cycle by supplying nutrients, and it also transports waste products away from the bone. Disruption of the blood vessels network in injury can lead to impaired bone healing [4].

1, in contrast, is an avascular tissue with a distinct layered organization [2]. The composition and structure of AC constituents change with depth and determine the unique functional properties of AC, such as its compressive strength and tensile resistance. Only 5-10% of dry weight is occupied by

AC cells, the chondrocytes, which produce collagen and PGs. The size, morphology and arrangement of the cells vary accordingly to their depth in AC [2]. Fine collagen fibers, predominantly of type II, occupy 50-75% of AC dry weight and provide AC with flexibility and high tensile strength. PGs (15-30% of dry weight) play an important role in compressive strength of AC and contribute to its low permeability [5].

The lack of a blood supply means that AC has a low ability to heal itself after injury and progressive degeneration over time often results in the development of osteoarthrosis (OA) [2, 6, 7]. The metabolism of bone depends on the activity of cells and is regulated by several hormones. Metabolic activity changes throughout life, and formation and growth are the most active during childhood. Loss of bone mass, measured clinically as the change in bone mineral density (BMD), is progressive with age. It is considered as a serious risk factor for bone fragility [8]. Disorders of bone metabolism lead to bone diseases, like osteoporosis (OP) or osteomalacia [1, 4, 9]. Even very subtle compositional deviations may be the evidence of early symptoms of tissue disease and, therefore, should ideally be identified as early as possible to allow successful treatment. Later, the follow-up is essential for identifying proper healing and, if relevant, making a decision about supplemental treatments.

Sensitive techniques are required to rapidly determine alterations in the composition and the structure of bone and AC. Most of the recent imaging modalities for monitoring tissue health, like magnetic resonance imaging (MRI) and ultrasound [10-13], may be qualitative in nature, with too low resolution for detailed investigation of the tissue composition. Quantitative biochemical analysis permits a precise measurement of composition of the tissue, but is not able to provide spatial distribution of the tissue components.

Fourier transform infrared (FTIR) imaging and microspectroscopy (FTRI-MSP) have been applied successfully in bone and AC research both for assessment of tissue composition and spatial organization [14-18]. By using univariate or multivariate analyzing techniques, FTIR-MSP can reveal minor variations in tissue composition [19-21]. Total and relative collagen, PG, mineral contents, mineral-to-matrix ratio and mineral crystallinity can be evaluated by quantitative assessment[1] of areas under the particular regions of infrared spectra [15, 17, 22]. However, utilization of several variables in the analysis permits the complex, unsupervised assessment of the subtle changes occurring in the FTIR spectra. These changes correspond to the alterations of the tissue's absolute or relative composition. Multivariate analysis techniques have been used successfully to investigate quantitative, as well as qualitative, changes in the infrared spectra and to discriminate biological tissues, fluids and cells based on their spectral differences [21, 23].

The main aim of this thesis is to examine the potential of using FTIR-MSP to investigate the advantages of cluster analysis for detecting quantitative and qualitative changes in composition of bone and AC. The first two studies examined the potential of cluster analysis to differentiate bone of different ages. Furthermore, the most accurate clustering method was determined. The following two studies focused on the changes

[1] Terms "quantitative" and "qualitative" are used in this thesis to sub-divide two types of analysis. Cluster analysis represents "qualitative" analysis, since it takes in use only qualitative features of infrared spectra, e.g, its shape. "Quantitative" analysis follows clustering and implies an estimation of areas under the infrared peaks and their ratios and comparison of these values between different clusters. Although an integrated absorbance indicates no concentration of tissue components in real units, it correlates with true concentration according to Lambert-Beer law. No true quantitative assessment of component's concentration in bone and AC, like calibration with pure chemicals, was conducted.

in the composition of normal and repaired AC. Cluster analysis was used to reveal the histological structure of AC in two species. Finally, healthy and repaired AC were differentiated based on subtle changes in FTIR spectra. Thus, this thesis represents a foundation for future studies towards employing cluster analysis and FTIR-MSP in medical diagnostics of cartilage and bone diseases.

2 Bone and articular cartilage

This chapter describes structure and composition of bone and cartilage tissues. Moreover, alterations in bone during its development, as well as evaluation of a repair of AC are reviewed.

2.1 BONE

2.1.1 Composition and functions of bone

Bone is a hard tissue; nevertheless, it is metabolically very active and dynamic, constantly adapting its shape and structure to the mechanical forces applied on the tissue. The main functions of bone are to provide mechanical support, to protect organs and bone marrow from damage, to transform muscle contractions into motions, to act as a mineral reservoir and to produce most of the blood components, e.g. red blood cells [24].

At the micromolecular level, bone tissue consists of an organic (20% of wet weight) and an inorganic (65%) matrix that can amount to 90% of the tissue volume, water (10%), and cells [4]. The mechanical properties of bone are closely associated with the structure, volume fraction of the bone and its ECM. Collagen accounts for 90% of the organic matrix and provides bone with its tensile strength and the ECM for the deposition of mineral [25, 26]. There is predominantly collagen type I in bone, but a small amount of types III and V are also present [4, 26].

Most of the minerals in the body are located in the inorganic matrix (mainly as hydroxyapatite crystals) of the bone, which

provide bone with resistance to compression, stiffness and strength [4]. Mineralization of bone occurs in the organic matrix as a transformation from soluble to solid phases of crystals [4].

There are four types of cells present in bone: osteoblasts, osteocytes, osteoclasts, and undifferentiated mesenchymal stem cells [4, 24]. Osteoblasts are densely packed rounded cells lying on the surface of bones. They synthesize the bone organic matrix, whereas other cells, osteoclasts, are responsible for the bone resorption. They are developed from the osteoclast precursor cells when stimulated by specific hormones and growth factors [4]. The most abundant and long-living cells in bone are osteocytes (90% of the total number of cells). They are surrounded by the bone matrix [4]. Altogether these cells form a complex network and are responsible for the sensitive mechanism of bone remodeling and coordination of bone life cycle.

2.1.2 Structure of bone

At the macromolecular level, the central fatty bone marrow is surrounded by two primary forms of bone tissue: first, trabecular (or cancellous) and then cortical (or compact) bone [4] (Figure 2.1). The bone marrow produces blood cells and comprises a net of blood vessels. The integrity of these vessels is crucial for bone health. Both types of bone have similar compositions and material properties, but the cortical bone has a higher density and lower porosity [24]. There are more cells per volume unit in the cortical bone and they are closely surrounded by the matrix. Cells in the cancellous bone are located on the surface of the trabeculae, which forms a porous net.

Cortical bone surrounds the bone marrow and cancellous bone. It provides support for the thin layer of the subchondral bone

(SB), which underlies AC in joints. SB can be subdivided into the SB plate and trabecular bone.

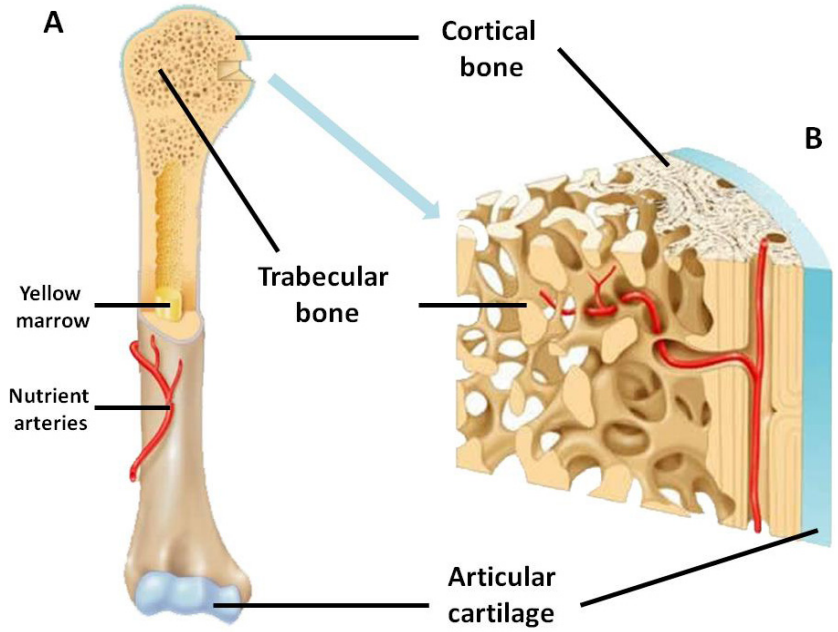

***Figure 2.1**: Schematic representation of bone structure (modified from [27]), showing A) the location and B) closer view of cortical and trabecular bone. Articular cartilage covers the ends of a long bone.*

2.1.3 Developing bone

Bone is a metabolically very active tissue, especially at young ages. In addition to modeling and remodeling during growth and maturation, physical activity, hormonal factors, bone diseases and artificial implants can influence the bone metabolism [1, 4]. In estimation, 10–15% of the bone in the whole body is replaced with new bone every year [24].

When the skeleton is newly formed, it consists of woven (or primary) bone, which is later almost entirely replaced by the lamellar (or secondary) bone [4]. Woven bone has an irregular structure of collagen fibrils and a very high rate of metabolic activity. Mineralization is a relatively fast process once it begins; and most of the mineral forms within hours. It results in the formation of strong and rigid lamellar bone with highly organized collagen structure and high BMD. Defects in the bone mineralization process can lead to osteomalacia, or a low rate mineralization. And under these conditions bone will weaken and be easily deformed.

In general, the structure, composition and mechanical properties of bone change with age [26]. Aging affects different types of bone differently [28]. Cancellous bone has a higher rate of metabolic activity and remodeling than cortical bone, and, thus, responds more quickly to mechanical loads [4]. A decrease in density of the cancellous bone can be detected earlier than an increase in porosity of the cortical bone [4]. Age-related fractures occur more often in the cancellous bone sites. A reduction of the mechanical strength of bone correlates with the decrease in collagen content [26].

The metabolism of bone collagen is the most active in SB. This is indicated by the gradual arrangement of the collagen network and remodeling of SB during maturation [8, 29-32]. There are studies describing biochemical changes in the levels of mineral, collagen, and collagen cross-links during growth and maturation of equine SB [30, 32]. According to these observations, major and rapid changes in equine SB occur during the first months of life after which further adaptation becomes slower, and skeletal maturation in horses is reached around the age of four years.

BMD, collagen content, amount of collagen cross links, mineral content and mechanical strength have been shown to increase in cortical bone during early growth in rabbits [33, 34]. Moreover,

maturation of the collagen network was followed by the mineralization process, which continued after the collagen network had become totally mature. The bone growth rate differs among locations in the body and depends on the gender and physical activity of a subject [28, 35]. Moreover, an age-related loss of bone mass and reduction in bone strength has been revealed in the elderly [4, 26]. This process was accompanied by thinning of the trabeculae in cancellous bone and increasing porosity in cortical bone [4].

2.2 ARTICULAR CARTILAGE

2.2.1 Composition and histological structure

AC is a thin layer (usually less than 3 mm)[36] of soft tissue that plays an important role in reducing friction and distributing loads across the joint surface. It contains no nerves or blood vessels [2]. The chondrocytes, AC cells that are responsible for the maintenance and repair of cartilage, are surrounded by an ECM consisting primarily of water, type II collagen, PGs, and glycoproteins [2, 7]. Water is a major component in cartilage, comprising 60-80% of its total weight [7]. The collagen in AC, which is mostly of type II, is represent as a meshwork of thin oriented fibers (15-22% of the AC wet weight) [7]. They provide tensile and shear resistance for the AC. PG (4-7% of the AC wet weight) is composed of a protein core and highly negatively charged glycosaminoglycans (GAGs). PGs contribute to the compressive stiffness of the tissue, mostly because of their charge and ability to attract water.

The thickness of AC varies between anatomical locations and species. It can be sub-divided into four main histological zones based on the orientation of the collagen fibrils, distinctive shapes of the cells and the biochemical composition of the ECM, i.e., superficial zone (SZ), middle zone (MZ), deep zone (DZ), and calcified zone [2, 7](Figure 2.2).

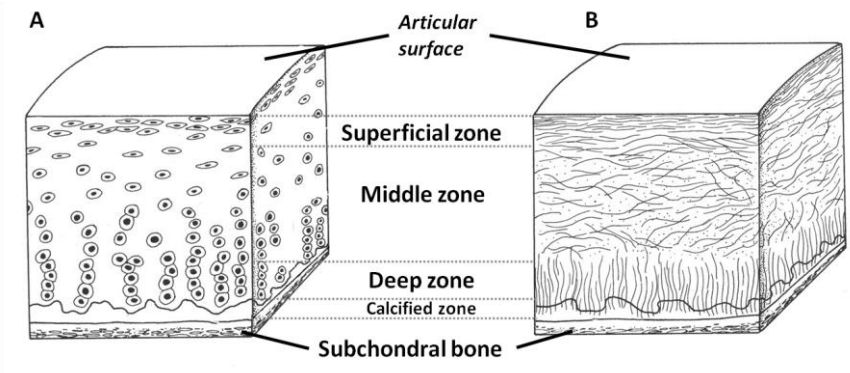

Figure* 2.2: Histological structure of AC (adapted from [37]). A) Schematic image demonstrating chondrocytes organization; B) Cross-sectional illustration of collagen fiber architecture.*

Thickness of the zones varies between species and joints [2]. The collagen fibers in the thin SZ (10-20% of the total AC thickness) are oriented in parallel to the AC surface. This arrangement helps cartilage to distribute the forces during mechanical loading. The SZ has the lowest PG content; PG content increases with depth in the cartilage and has reaches its zenith in the DZ. In the MZ (approximately 60% of total thickness) collagen fibrils are mostly randomly organized while in the DZ they are oriented perpendicular to the AC surface. The size and activity of chondrocytes also vary with depth from the small size and relatively inactive cells in the SZ to clusters of more active cells in the DZ. The tidemark, the line separating the calcified cartilage zone, is characterized by the absence of PGs, and by having rounded chondrocytes and perpendicular collagen fibers in the calcified matrix [2]. This unique spatial distribution defines the main functional properties of AC. When a force is applied to the joint, AC deforms, which causes flow of the tissue fluid and results in a swelling pressure [7]. Network of collagen fibers balances the swelling pressure of the water-PG gel, creating a composite with unique biomechanical properties.

2.2.2 Cartilage repair

Although AC has a highly organized layered structure and can resist high compressive stresses, it can be damaged either mechanically or chemically [7]. Injury or diseases lead to deterioration of AC and the formation of focal lesions in the tissue. Without treatment, small lesions increase in size with time and may result in full thickness lesions reaching the SB plate [38]. The avascular nature of AC and the immobility of chondrocytes result in a tissue with very limited capacity to heal spontaneously [7, 36, 38]. When the defects penetrate into the bone, a blood clot is formed, initiating inflammation and more extensive reparative processes [7, 36]. Small (<3 mm in diameter) osteochondral defects can heal partially, remaining stable or developing distinctive degradation patterns over time [7, 38]. However, larger defects (>6mm) or small partial-thickness defects lack the ability to completely heal [6, 7]. Therefore, much effort has been exerted into finding ways to repair AC defects. This leaded to the introduction of several surgical techniques focused primarily on transplantation of new viable cells capable of chondrogenesis and/or on improving access to a vascular supply [7]. Many methods have been examined in animal and clinical studies with various degrees of success. Drilling, shaving of AC, implantation of autologous chondrocytes (ACI), mesenchymal stem cells embedded in various gels, implants and growth stimulating factors have all been described in the literature [7, 36, 38, 39]. Some of these techniques were claimed to produce good quality cartilage and have entered into clinical practice (like ACI [40, 41]). However, long-term follow-up of the treatment revealed no complete filling of defects. Some reports have described the continuous replacement of fibrous tissue with fibrocartilaginous tissue (FC) showing high collagen type I to type II ratio. Later, a partial replacement with hyaline-like cartilage has been reported, which in most cases was followed by the onset of degenerative changes occurring as early as 10-12 weeks after implantation [7, 40, 42]. The degradation of repaired AC was attributed to cell death, poor integration of repaired

tissue with surrounding normal tissue and filling of the superficial layer of AC with FC, which structure and morphology are rather dissimilar to AC [38, 39].

Collagen content, integrity and orientation of collagen fibers, as well as PG content are crucial determinants of the AC integrity [2]. It is necessary to monitor the repaired tissue to understand the mechanisms of the healing process and to evaluate repair quality. Special guidelines for repair studies have been developed [43] with the aim being to standardize the experimental setup and assessment. The structure, composition, integrity and organization of the repaired tissue have been evaluated using histological staining and scoring, as well as polarized light microscopy (PLM) [18, 43, 44]. Several other imaging techniques proved their utility in the non-invasive evaluation of AC, e.g. high resolution MRI [10, 45], optical coherence tomography (OCT) [46], ultrasound imaging [11, 13] and infrared fiber optic probe (IFOP) [47, 48]. Experimental human and animal studies employed imaging techniques and revealed an increase in collagen integrity during the stage of early repair [49, 50]. Nonetheless, the structure of the collagen network and distribution of PG and collagen across the AC differed from that found in intact tissue [50, 51].

3 Fourier Transform Infrared Microspectroscopy

This chapter describes the basic principles and advantages of the spectroscopic imaging technique used to collect data from bone and AC. Further, an overview of the applications of FTIR-MSP in bone and cartilage research will be presented, followed by a review of the multivariate data analysis methods used to analyze FTIR-MSP data.

3.1 BASIC PRINCIPLES OF FTIR-MSP

Fourier transform infrared microspectroscopy (FTIR-MSP) is a vibrational spectroscopic technique that is capable of producing biochemical microscopic images of tissue sections. In FTIR-MSP, the fraction of infrared light intensity (amount of energy) transmitted through the sample is measured point-by-point from the microscopic section at each frequency in the mid-infrared region of electromagnetic spectra (500-4000 cm^{-1}) (Figure 3.1), producing an interferogram. Additionally, a background spectrum is measured at a sample-free area, which is then subtracted from the original spectra in order to remove any instrument and mounting substrate characteristics from the spectral information of the true sample [52]. Fourier transformation is applied to the corrected interferogram at each pixel to obtain the desired infrared spectra [52]. Thus, a three-dimensional data matrix with two spatial and one spectral dimension is produced as an output.

The absorption spectra (ABS) is calculated later from the transmittance spectra (TR) according to eq. 3.1 [52]:

$$ABS = \log_{10}\left(\frac{I_0}{I}\right) = \log_{10}\left(\frac{100}{TR}\right) \quad (3.1)$$

where I and I_0 are the intensity of radiation transmitted from or incident to the sample, respectively.

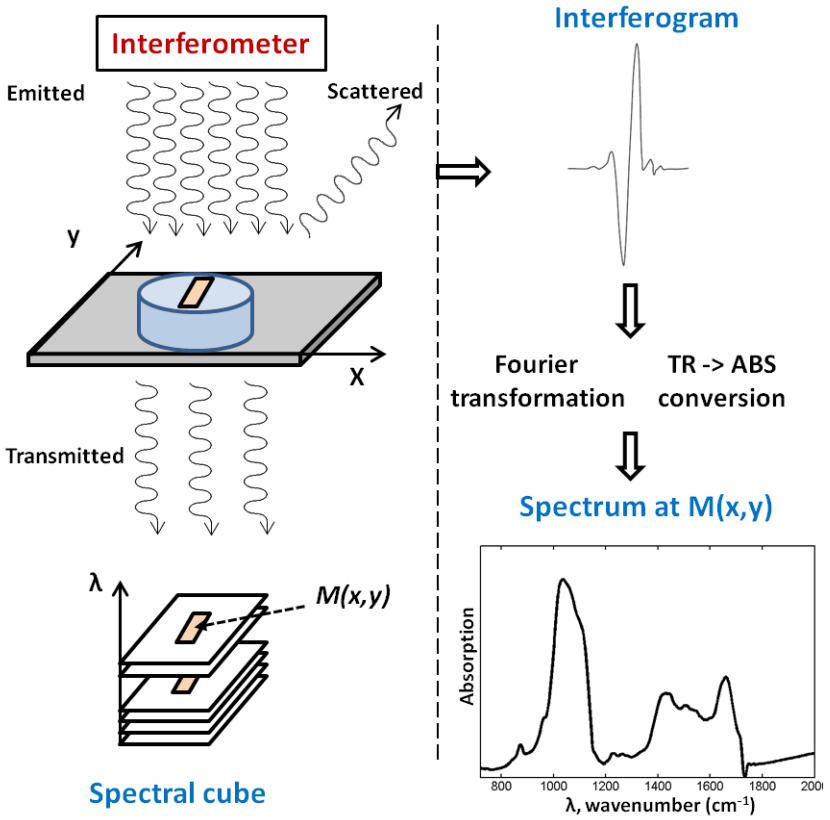

Figure 3.1: *Process of acquiring infrared spectra. Infrared energy is emitted by the source and is transmitted (TR) through or reflected off the sample surface. The beam is then passed to the detector, which measures an interferogram signal. Fourier transform is then applied, and the signal is converted into absorption (ABS). A spectral matrix with two spatial (x,y) and one spectral (λ) dimensions is constructed, consisting of a spectrum at each coordinate (x,y).*

The infrared ABS shows the absorption bands, which originate from the interaction between specimen molecules and the

energy of infrared radiation. Different motions of the molecules, such as rotation and vibration, can be discriminated in the spectra [1, 52] (Figure 3.2).

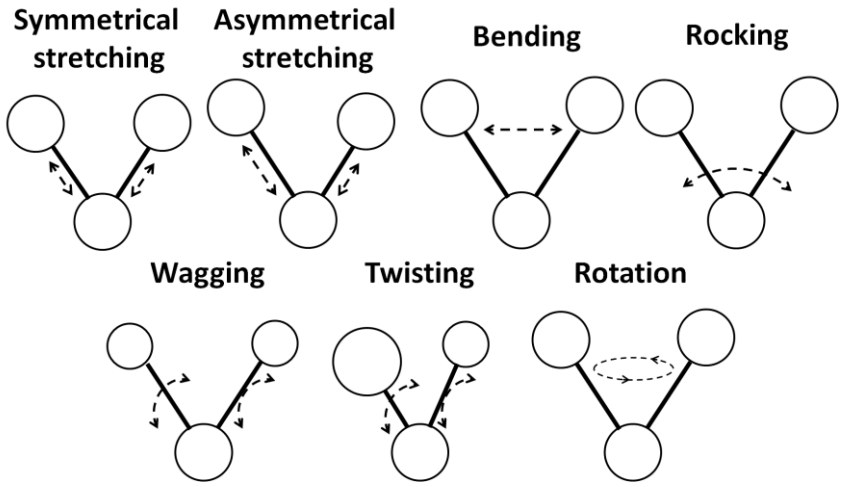

***Figure 3.2**: Schematic representation of different types of molecular vibrations*

Different atoms and molecules absorb infrared energy and undergo particular motions at a specific wavelength [1, 52-54]. Thus, the groups and structures of the molecules can be identified from the infrared spectrum by accessing the position of a particular infrared band (Table 3.1) [54, 55]. Hence, the infrared spectrum can be thought of as a fingerprint of the underlying molecular structure.

***Table 3.1**: Characteristic assignment of infrared bands*

Wavenumber (cm^{-1})	**Band assignment**	**Description**	
1202	**amide III (AIII)**	CH_2 wagging vibration from the glycine backbone and proline sidechain [56]	
1228-1230		SO_3- asymmetric stretching vibration of sulfated GAGs [57] / CN stretching/ NH bending [58]	
1340		CH_2 side chain vibrations of collagen [15, 56]	
1514	**amide II (AII)**	C-N stretching/N-H bending/ C-C stretching of collagens [15, 56, 58]	
1548-1550			
1638-1644	**amide I (AI)**	C=O stretching [15, 56]	β-sheet [56, 58], amide I from proteoglycans [15] O–H bending of water [56]
1659-1660			possibly non-reducible collagen cross-links [15, 59]
1660-1668			3_{10} Helix [56] / β-Turn [58]
1692			β-sheet [56, 58]
1677, 1695			possibly reducible collagen cross-links [15, 59]
1030	**carbohydrate (CHO)/ phosphate**	C-O stretching vibrations of the carbohydrate residues in collagen and PGs [56]	
1062			SO_3^- symmetric stretching vibration of sulfated GAGs [57]
1080			
850-890	**carbonate**	out-of-plane bending modes of CO_3^{2-} [22]	

When a complex biological specimen is measured, the infrared spectra represent the superposition of vibrational modes of the molecules [52, 55]. Therefore, biochemical composition, molecular structures and concentrations, conformations, and molecular interactions can be analyzed from the infrared absorption spectra at each pixel. The protein and mineral constituents of the biological tissues, such as bone and AC, produce intense, structure sensitive infrared modes.

Other advantages of the FTIR-MSP are that it is non-destructive and requires no staining to reveal the spatial organization and concentration of the tissue constituents [60]. Instead, the map of a particular component is constructed by cutting the spectral cube in the spectral dimension at a specific wavelength, or integrating over the spectral region [52](Figure 3.3).

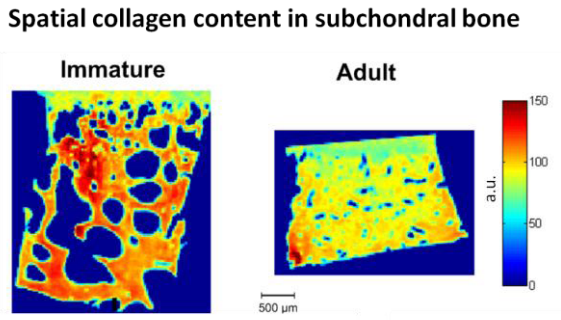

***Figure 3.3**: Absorption maps, constructed by integrating the area under the amide I absorption peak of the infrared spectra. The images show the spatial distribution and concentration of collagen, and different structure of samples in subchondral bone obtained from horses. Articular cartilage tissue can be seen at the top as a light-blue area, containing less collagen.*

Moreover, the technique practically has no restrictions related to the origin of the sample or its state and the specimen requires no or very little preparation before measurements [52, 60].

Therefore, FTIR-MSP has been successfully used in various fields of science due to its outstanding features, like speed, sensitivity, simplicity for identification, quality control, quantitative analysis and differentiation [19, 21, 53-55, 61].

3.2 PRACTICAL CONSIDERATIONS FOR FTIR-MSP MEASUREMENTS

Despite the obvious advantages of FTIR-MSP, there are also several important points one must consider before conducting the measurements [60]. The biological samples must be thoroughly prepared. First, the tissue section must be thin enough for transmission measurements. Usually, 2-5 μm and 5-10 μm thick sections are used when bone or AC is measured, respectively [22, 61, 62]. The uniformity of sample thickness is a critical issue when conducting quantitative analysis [60]. However, this aspect is not so critical when qualitative analysis is conducted after vector normalization of spectra. Second, soft biological samples must be either cryosectioned or embedded into resin, usually paraffin, after formalin fixation and decalcification [60, 62]. Hard plastic, such as polymethyl methacrylate (PMMA), can be used without decalcification of hard tissues [62]. Third, only dehydrated materials can be successfully measured with FTIR-MSP due to the high overlap of water infrared vibrations with the most information-rich regions in the infrared spectra [52]. Last, special infrared transparent windows must be used to avoid any extra contribution of the infrared absorption of the mounting material to the sample spectra.

The quality of the spectral data is an important consideration for conducting a multivariate analysis; thus, the signal-to-noise ratio (SNR) of the spectra must be as high as possible [55]. To achieve this goal, high spatial resolution, spectral resolution and adequate averaging of several scans are required, and this unfortunately can considerably increase the measurement times [60].

The next essential complication one must overcome during the measurement is to remove any contributions of water vapor and carbon dioxide (CO_2) from the sample surrounding air. Therefore, sample chamber need to be purged with N_2-gas or dried air and the concentration of water vapor is continuously monitored. Moreover, the temperature in the laboratory should be kept constant to standardize the measurement conditions over time [60].

3.3 FTIR-MSP OF BONE AND CARTILAGE

The specific ability of FTIR-MSP to reveal composition and structural organization of the complex tissues makes it capable of providing information on chemical alterations in tissue composition, e.g. that resulting from natural processes, like aging or degradation [53, 63]. When the tissue is affected by a disease, its structure and composition may change, and it evokes changes in the infrared spectra. Sensitive analysis techniques can reveal those subtle specific "fingerprint" changes [64]. Furthermore, the sensitivity and specificity of FTIR-MSP for identifying the nature of the specimen permit the discrimination of different types of sample materials. These advantages make FTIR-MSP a useful tool in biomedical research [19, 53].

The assessment of the tissue quality is a critical issue when investigating bone-related diseases like OP [65]. One aspect of bone quality is its composition. FTIR-MSP has been used to determine the local chemical composition, the relative amounts of bone constituents, their molecular nature, distribution, and orientation [1, 14, 16, 22, 59, 64, 66-68]. The most intensive peaks in the FTIR-MSP spectra of bone are produced by the mineral (carbonate and phosphate) and collagen protein (amide I (AI) and amide II (AII) collagen) constituents (Figure 3.4).

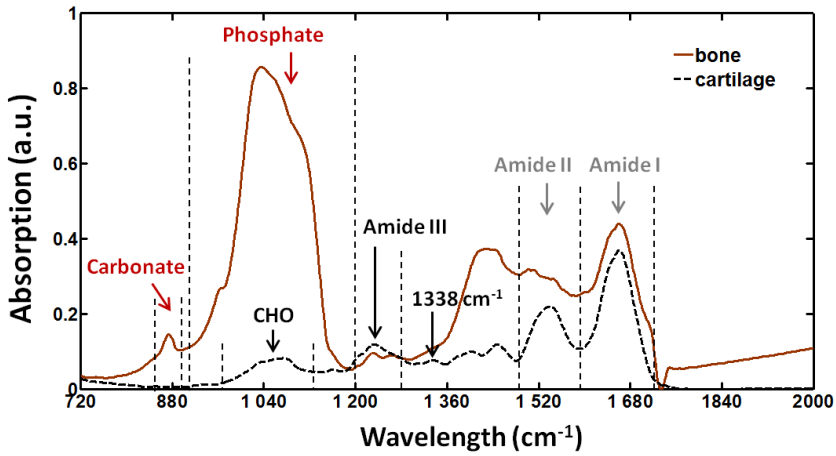

***Figure 3.4**: Examples of typical FTIR spectra of bone and cartilage. The absorption bands of interest are indicated.*

Both the quantity and quality of bone components and their ratios can be assessed and analyzed from FTIR-MSP data [1, 63]. The mineral-to-matrix ratio Ph/AI is calculated as the ratio between the integrated areas under phosphate (900-1200 cm^{-1}) and amide I (1584-1720 cm^{-1}) infrared bands. These areas are directly proportional to the amount of mineral and collagen, respectively. Ph/AI has been used as a measure of BMD; it describes whether the bone tissue is normal, or hyper- or hypomineralized [1]. The mineral maturity (or crystallinity) is accessed as a ratio of the 1030/1020 cm^{-1} sub-bands under the phosphate peaks, which are determined with second derivative analysis. In addition, the maturity of collagen (crosslinks) can be accessed from the ratio of sub-bands intensities at 1660/1690 cm^{-1} [59].

The age-dependent changes in infrared spectral parameters of healthy and diseased human bone have earlier been summarized by Boskey et al. [14, 22]. In addition, significant correlations have been reported between absorption bands and their relative intensity in organic bone matrix with aging [22, 31]. Based on the quantitative analysis, it has been shown that the AI and AII components of proteins undergo frequency and

intensity changes as a result of changes in the protein secondary structure. Both the collagen and mineral content increase rapidly during the development and growth of bone, and remain more stable during adult life. Later, Paschalis et al. [1] demonstrated how FTIR-MSP and FTIR imaging can provide information related to bone quality. In particular, changes of crystallinity and spatial variation of collagen crosslink ratio parameters allowed discrimination between normal and osteoporotic bone.

In cartilage research, FTIR-MSP has been proved to be efficient in imaging the spatial distribution and in estimation of the concentration of the two main solid components of AC, i.e., collagen and PG, in healthy and diseased cartilage [15, 17, 22, 48, 69-72]. Estimated collagen and PG contents are analyzed by calculating the integrated area under the AI peak and ratio of CHO/AI, respectively. Moreover, the collagen integrity, or the absorbance at the 1338 cm^{-1}, has been utilized to estimate the extent of degenerative cartilage [22]. Importantly, decreasing collagen, PG contents and collagen integrity parameter, as well as altered collagen fibril organization, have been shown to correlate with the progression of OA [48-50, 73]. These changes were detected in laboratory studies with FTIR imaging and *in vivo* with IFOP [48, 49] and have been shown to correlate well with the histological assessment.

Moreover, the sensitivity of FTIR-MSP at detecting the orientation of collagen fibers in AC has been demonstrated when polarized light was used [15, 18, 44, 50, 51, 74]. The studies revealed changes in the intensities of the major collagen-specific infrared bands, like AI and AII peaks, when the measurements were affected by the linear polarization or when a sample with anisotropic properties was rotated [15, 18, 44, 51, 74, 75]. This behavior was attributed to the dichroism of the two amide bands molecular vibrations, causing transitional moments, which were almost perpendicular to each other [44, 51]. These properties have been used for the estimation of the

orientation of collagen fibrils and determination of the AC zonal organization by quantifying the ratio of AI/AII integrated absorbencies [15, 18, 51]. In addition, FTIR-MSP has been successfully used in the investigation of protein secondary structure, proteins concentrations, stability and dynamic properties [58].

The power of FTIR-MSP to describe spatial and compositional changes in AC can be utilized in the assessment of the quality of repair AC tissue after surgery or a natural healing process. One study has shown the potential of FTIR imaging to characterize the structure of repair cartilage at the molecular level [76]. Kim et al. [10] conducted a short term follow-up of enzymatic treatment of surgically created AC osteochondral defects with FTIR imaging and correlated FTIR-derived quantitative parameters with T2 mapping parameters of MRI. They could detect changes in the collagen and PG contents with time and treatment using FTIR-MSP [10].

3.4 ANALYSIS METHODS IN FTIR-MSP

The most widely used analytical techniques in FTIR studies are those that access only one variable at a time, or so-called univariate analysis methods [20]. They permit the estimation of the quantity or concentration of tissue constituents by quantifying the areas under specific infrared peaks and their ratios.

Univariate analysis methods are simple, but their sensitivity has been reported to be restricted [71]. Moreover, when complex tissues like bone or AC are analyzed it may be essential to use more complicated multivariate analysis techniques to improve the specificity of parameters. Multivariate analysis methods, in contrast to their univariate counterparts, can process a large region of the spectrum at a time. Univariate methods require one to define a specific part of the spectrum in the calculations.

In contrast, multivariate methods use all available information and reveal regions of spectra with discriminate spectral features. Thus, the use of multivariate methods is essential when one wished to analyze complex overlapping spectra.

A number of data analysis techniques used to obtain similarities or dissimilarities in spectra were employed in FTIR research. The most commonly used multivariate methods in biomedical research are as follows: curve fitting, principal component analysis (PCA), linear discriminant analysis (LDA) [77, 78], artificial neural networks, partial least squares regression (PLS) [20, 21, 64, 73], principal component regression (PCR) [72, 79].

The absorption peaks generally consist of overlapping signals (sub-peaks) from different constituents. Curve-fitting can help to isolate these sub-peaks. With this technique, PG-specific sub-peaks have been identified in the FTIR-MSP spectra of bovine AC. They correlated with results obtained in histology [71]. In bone research, curve-fitting has been used to detect the underlying phosphate bands and to estimate crystallinity/maturity.

All other methods on the list belong to the pattern recognition techniques, and have been employed in diagnostics to detect diseases and their stages [21, 23], as well as in differentiating between types of microorganisms and tissues [21]. PLS and PCR can predict one unknown feature based on data obtained during the learning phase on a test set of samples. With use of these methods, one can estimate relative concentrations of collagen or chondroitin sulfate in native [72, 79] and digested AC [17], as well as their distribution in different cartilage layers [73]. Li et al. detected a correlation of IFOP values with the Collins visual grading scale of cartilage [47].

The multivariate methods can be either unsupervised or supervised [64]. Unsupervised methods, in contrast to the supervised, require no *a priori* knowledge about the structure

and labeling of the data points. This feature is useful when the grouping of samples must be undertaken based solely on the spectral features and not on any knowledge or special expertise of the operator [64]. By studying detailed differences in the overall shape or position of the infrared absorption peaks, it is possible to extract novel information about the molecular structure.

4 Cluster analysis

This chapter provides an insight into the classification of clustering approaches and their algorithms. Moreover, the pros and cons associated with the different methods are assessed, and cluster validity is considered. Finally, the applications of cluster analysis in bone and cartilage research are reviewed.

4.1 PRINCIPLES OF CLUSTER ANALYSIS

Clustering is a technique for statistical data analysis, used in many fields, e.g., image analysis and pattern recognition. Cluster analysis is a generic name for a wide variety of discrimination procedures [80, 81]. In these techniques, "clusters", or classes, are formed as groups from highly similar entities. Discrimination is an important task in many medical applications and serves two goals: to differentiate between diseases, which require different treatments, and to provide a basis for aetiology [82]. When no *a priori* information on class labels is used during clustering, this clustering is called unsupervised.

Discrimination of objects by clustering is done based on differences in their properties, expressed in numbers or textual features. In FTIR-MSP, an infrared spectrum can be thought as a set of features, which describe properties of an object, e.g., a composition of bone or AC. Thus, the input data matrix is a [n,p] multivariate matrix X, which contains columns of variable values for each of objects in a row (eq.4.1)

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \quad (4.1)$$

where $i = \overline{1,n}$ is a number of an object, $j = \overline{1,p}$ is a number of a variable and X_{ij} gives the value of the *j*th variable of *i*th object. In terms of spectra, this matrix represents n spectra of dimension p.

The similarity (distance) between the objects is calculated to evaluate how close or far objects are to each other. Many different distance measures can be used to estimate similarity [80]. The choice is dependent on the type of the variables (categorical or continuous), whether objects were measured once or repeated measurements were performed, and whether the distances were measured between individual objects or between groups of objects [82]. The most commonly used measure for continuous data is the Euclidean distance measure [80, 83] (eq. 4.2).

$$d_{ij} = \left[\sum_{k=1}^{p}(x_{ik} - x_{jk})^2\right]^{1/2} = \left\|x_i - x_j\right\|_2 \quad (4.2)$$

where x_{ik} and x_{jk} are the *k*th variable value of the p-dimensional observations for objects i and j. This measure is interpreted as a physical distance between two points in the Euclidean space.

The calculated distances between n objects are presented in the form of a dissimilarity matrix (eq. 4.3):

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nn} \end{pmatrix} \quad (4.3)$$

The larger the distance d_{ij}, the less similar objects i and j are from each other. This matrix is used by clustering algorithms to construct clusters.

In most clustering applications, data is partitioned in disjoint clusters, where an individual object belongs to a single cluster [82]. However, in some situations, overlapping clusters can

provide a more reliable solution [82]. In general, clustering methods are divided into hierarchical and partitional, and into "hard" and fuzzy approaches [80, 83]. The grouping of clustering methods into ''hard'' or fuzzy is based on the number of clusters to which the object can belong simultaneously [83]. In fact, "hard" clustering methods are a special case of fuzzy methods. In this situation, each object is given a membership degree value of 0 (no membership) or 1 (full membership). Instead, in the fuzzy approach, the object can be assigned to multiple clusters with a membership degree somewhere between 0 and 1.

Hierarchical cluster analysis (HCA) is conducted in two approaches: agglomerative and divisive. In divisive clustering, all objects are first assigned to the same clusters and at each step they split into groups until every object is in its own cluster. In the agglomerative clustering, which is used more frequently, the clusters are generated by a sequence of merging operations. The algorithm starts by initializing each data vector as a separate cluster. Two clusters are merged at each step by applying some rule to compare distances from the matrix D. The process is repeated until the desired number of clusters is obtained. Different rules for characterization of similarity between pairs of objects are used to construct clusters. Among them, Ward's algorithm of minimum-variance [84] is one of the most popular and, in general, highly efficient [85]. Instead of operating on pairs of clusters in a sequence, it first computes distances between all possible pairs of clusters and the overall variance produced by this partition. The combination of clusters with the lowest variance is then chosen [83].

A *dendrogram* is computed in order to visualize the obtained structure of clustering. Dengrogram is a tree with objects located at the ends of branches. The length of the branch corresponds to the distance between the clusters. The desired number of clusters is obtained by "cutting" the tree at a certain level. The optimal cutting level could be determined finding the greatest jump between the stages of clustering construction, namely, the

branches' length. A large jump indicates that lower and upper stage clusters are relatively far apart from each other, and the tree is cut at the level of these long branches [83]. Thus, hierarchical clustering requires no setting for a number of clusters before running the algorithm. This could be beneficial when no information is available beforehand about the structure of the data [83].

In partitional clustering approach, a single partition instead of the clustering structure is constructed [86]. A common feature of the methods in this group is the start of the clustering procedure from an initial solution. This can be a random choice or initial guess defined by the user [83]. It uses an iterative algorithm to update the selected initial cluster centers randomly. Unfortunately, the final solution could be sensitive to the initial setup [83]. This type of clustering requires that the number of clusters has to be defined in advance. Thus, one needs to try to guess the possible structure of the data and give the number of clusters as an input to the clustering algorithm. One can also run the algorithm several times using different number of clusters and select that one, which produces the smallest total variance of the final clustering structure.

Partitional clustering has advantages over the hierarchical approach, when a large set of data is clustered. In this case, the construction of the dendrogram becomes computationally very complex, both in terms of space and time [83]. Moreover, in partitional methods, the objects can change clusters during the computation. In this respect, this type of clustering is more dynamic.

The most frequently used presentation of the clusters is a computation of the clusters' centroids. The means of the clusters can be used to interpret the differences between the obtained data classes. In the case of FTIR-MSP, clustering regroups the spectra with similar spectral characteristics. Hence, ideally, spectra in different classes can demonstrate different spectral (chemical) signatures.

Additionally, if clustering is done point-by-point on an image of size $a*b*p$, where $a*b=n$ is a total number of objects and p is a dimensionality as stated in X (eq. 4.1.), clustering maps could be constructed using the cluster memberships. The false color coding is used to visualize the assignment of the object to a particular cluster. These maps are useful when one is analyzing a structured entity, such as a section of a tissue.

4.2 "HARD" CLUSTERING METHODS

In this thesis, the two commonly used "hard" clustering methods are described, i.e., partitional k-means analysis and agglomerative HCA.

In the computer science and pattern recognition community, the k-means clustering algorithm is well known as the generalized Lloyd algorithm or "hard" c-means algorithm [83].

Algorithm 4.1 Basic K-means algorithm

Select k points as initial centroids
REPEAT
    Compute distances between each point and all centroids
    Form k clusters by assigning each point to its closest centroid
    Re-compute centroids
UNTIL
    Centroids do not change or total error changes less than a predefined minimum

The algorithm starts with a random solution (or predefined by the user initial partition). It classifies the data points into a predefined number of classes by iteratively recomputing means of clusters until the criteria are satisfied. The criterion function, which is most frequently used by the partitional clustering methods, is minimizing a sum of squared errors (SSE) [86]. To

compute the SSE, the error of each object in a partition L is calculated as the Euclidean distance between the object x and its closest centroid c and then the sum of values are found [80] (eq. 4.4):

$$SSE(L) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left\| x_i{}^j - c_j \right\|^2 \quad (4.4)$$

where k is the total number of clusters in the partition L, j is a particular cluster $j = \overline{1,k}$, and i is an object assigned to the cluster j. The sign $\|.\|$ denotes the Euclidean distance measure.

In this case, interclass variance has a maximum value, and the intraclass variance is minimal. The output of the algorithm includes the cluster membership map and the centroids of each cluster. Centroids are calculated as an average value of all objects in the cluster [83]. The popularity of the k-means algorithm can be traced to its easy implementation and a low time complexity. The larger the number of objects being used, the longer the computational procedure will take [81]. This method inherits properties from the group of partitional clustering methods, such as sensitivity to the initial partition [86]. This sometimes leads to convergence to a local instead of global minimum of the criterion function. However, this disadvantage could be overcome by repeating clustering several times with different initial solutions and selecting the best clustering solution with the minimum SSE. In this case, the number of runs needs to be large enough to enable convergence to the global minima, but also adequate in terms of the time spent. Another critical issue is the sensitivity of k-means to outliers. It is recommended that any outliers should be identified and removed before clustering is run [83, 86]. Moreover, k-means and its variations tend to have limits in finding some type of clusters in the data. To deal with this problem, one should either accept that clustering sometimes cannot identify "natural" clusters in the data, or try a larger number of clusters [83].

The second type of "hard" clustering used in this thesis project is the agglomerative HCA. HCA represents a sequence of partitions where each partition is nested into the next partition in a sequence [86]. It groups data patterns recursively into a tree (i.e., dendrogram). The basic HCA is:

Algorithm 4.2 Basic agglomerative HCA

Compute proximity matrix P using matrix D (eq. 4.3.)
REPEAT
    Merge the closest two clusters i and j
    Update P
UNTIL
    Only one cluster remains

The proximity matrix in this notation denotes the dissimilarity measure, which can be defined differently based on the choice of pairs of objects or clusters [80, 83]. In terms of proximity between the objects, single link, complete link or group average approaches are defined. The distance between two closest, two furthest points in clusters or between averages pairwise distances between points from different clusters are considered, respectively. When each cluster is represented by a centroid, instead of the set of objects, the proximity is defined as a distance between the cluster centroids. Ward's method selects the cluster pair to be merged so that the merge increases the SSE value by the least extent.

HCA has no global objective function and in this way it differs from the partitional methods. However, it encounters no problems with local minima or difficulties in selecting the initial solution. The computational complexity of agglomerative methods is their main disadvantage. When the time and space complexities for k-means clustering are $O(nkl)$ and $O(k+n)$, respectively, for agglomerative clustering they become non-linear, or $O(n^2 \log n)$ and $O(n^2)$ [80, 81, 83, 86]. Here, l is the

number of iterations taken by the algorithm to converge. In addition, they are quite sensitive to noise and outliers [86].

4.3 FUZZY C-MEANS

FCM clustering belongs to the group of partition-based clustering methods. It seeks to minimize an objective function by exploiting the fact that each object has some graded membership to each cluster [86]. The clustering criterion allows each object to be assigned to multiple clusters. The algorithm starts from an initial (random or defined by the user) partition by selecting class membership matrix U (eq. 4.5):

$$U = \begin{pmatrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nk} \end{pmatrix} (4.5)$$

where u_{nk} is the grade membership of object x_n in cluster c_k. Typically, $u_{nk} \in [0,1]$. Each object's memberships to different clusters always sum to 1 (eq. 4.6):

$$\sum_k u_{nk} = 1, (4.6)$$

The membership values indicate the strength of association of an object with the cluster. FCM clustering iteratively updates the cluster centroids, estimates and updates the U matrix.

The criterion function in fuzzy clustering L becomes:

$$SSE(L,U) = \sum_{j=1}^{k} \sum_{i=1}^{n} u_{ij} \left\| x_i - c_j \right\|^2 (4.7)$$

where $c_j = \sum_{i=1}^{n} u_{ij} x_i$ is a fuzzy cluster centroid of the *i*th cluster.

FCM is the most popular technique among all fuzzy clustering methods. It performs better than k-means in avoiding local minima, but it can still converge to this value. The critical point in FCM is to select the initial membership function [80]. The

fuzzy clustering results could be converted to "hard" clustering by assigning each object to the cluster with the highest membership value.

The number of clustering maps produced by the algorithm is equal to the number of clusters. Within the map, each point marked with a false color corresponding to the membership degree of belonging to this cluster. In this thesis, "hard" clustering maps were calculated from the combination of fuzzy-maps, by assigning each pixel (object) to the cluster with the maximum membership degree value.

The advantage of fuzzy clustering methods over the "hard" clustering approach lies in their ability to obtain degrees of uncertainty of belonging to each class. This feature of fuzzy clustering makes it more reliable when one is studying biological tissue samples, provided that the boundaries between the features are not sharp and furthermore careful investigation of the continuously changing feature is essential [86].

4.4 VALIDATION OF CLUSTERING

The Rand index R is used as a measure of similarity between the clustering results and the known actual classification [81, 83, 87]. It calculates the relative number of agreements between two partitions over the total number of pairs of elements (eq. 4.8):

$$R = \frac{Number\ of\ agreements\ between\ clusterings}{number\ of\ agreements + number\ of\ disagreements} \quad (4.8)$$

The Rand index values lie between 0 and 1, with 0 indicating that the two data clusterings do not agree at all and 1 meaning that they are identical. Thus, the closer the Rand index is to 1, the better is the clustering result as compared to the actual data classification.

Mean squared error (MSE) indicates how well the clustering model represents the actual data. It decreases monotonically as the number of clusters increases. Smaller MSE value indicates more compact and reliable clustering. When MSE values are compared for different numbers of clusters, a sharp change in the slope of MSE graph can reveal the most adequate number of clusters for an algorithm [83].

4.5 CLUSTER ANALYSIS APPLIED TO FTIR-MSP

The multivariate cluster analysis has been successfully used for differentiation and discrimination purposes in many fields, like biochemistry, biomedicine, nutritional science, imaging, object recognition, etc [64, 80, 82, 83, 88]. It has been applied for the identification and differentiation of biological micro-organisms such as bacteria [55, 89, 90], normal and cancerous tissues [61, 91-93] and other diseased tissues [19, 21, 94].

In bone and cartilage research, cluster analysis was mostly exploited in MRI studies, e.g., to differentiate normal and degraded nasal cartilage [95]. It has rarely been used in the FTIR research. Clustering was applied to the bovine muscle tissue for spatial separation of connective tissue and myofiber spectra [23]. In FTIR-MSP analysis of AC cross sections, clustering on the second derivative infrared spectra was applied [96]. The method was able to reveal major changes in chemical composition, as revealed by a shift in the AI peak to lower wavenumbers. This was consistent with collagen denaturation (antibody-induced damage), which had been conducted prior to the analysis.

In another AC study, a combination of hierarchical and fuzzy clustering techniques was applied on Raman spectra [97]. Together they differentiated cells and ECM, and detected variations in the ECM. Furthermore, fuzzy clustering highlighted the inhomogeneity of the ECM and was able to provide a detailed description of continuous variation in composition of ECM.

5 Aims of the study

The main aim of this thesis was to combine FTIR-MSP with multivariate unsupervised cluster analysis in order to reveal spatial variations and changes in the composition of biological tissues, such as bone, articular cartilage and engineered cartilage.

The specific aims of this thesis were:

- to detect qualitative and quantitative changes in the composition of SB and cortical bone with age;
- to compare the performance of different types of cluster analysis algorithms and identify the most suitable and accurate approach for analyzing bone and cartilage;
- to establish an effective algorithm for processing FTIR-MSP data of bone and cartilage with cluster analysis;
- to test the power of cluster analysis to differentiate between bone of different ages and between normal and repaired AC by utilizing only qualitative differences in the spectra;
- to evaluate the ability of cluster analysis to identify the histological zonal structure within intact AC.

6 Materials and methods

This thesis consists of four independent studies (I-IV). PLM and histological data used in studies II and IV were obtained from earlier studies by Rieppo et al. [98] and Pulkkinen et al. [99], respectively. The rest of the data is original. The rabbit AC in studies III and IV originates from the same set of samples. The materials and methods used in the studies are summarized in Table 6.1. In this thesis, all data and images were analyzed with Matlab software (v. 7.8-7.10, The Mathworks Inc. Natick, MA, USA).

Table 6.1: *Summary of the materials and methods used in studies I-IV.*

Study	Tisue	Species/ location	n	Methods
I	subchondral bone	equine metacarpo-phalangeal joint	*n*=29	quantitative analysis, k-means
II	cortical bone	rabbit humerus	*n*=35	k-means, HCA and FCM; DA; validity analysis
III	articular cartilage	bovine patella rabbit patella	*n*=8 *n*=7	quantitative analysis, FCM, PLM
IV	articular cartilage	rabbit patella	*n*=12	quantitative analysis, FCM, histological analysis

In this chapter, measurement setup along with the protocols used for preparation of samples and FTIR-MSP data for data analysis is explained. Further, the details of comparison of quantitative and qualitative analyses, as well as results with reference data are described.

6.1 SAMPLE PREPARATION

In study I, equine SB samples were taken from two differently loaded sites at the articular surface of the proximal phalanx in the normal left metacarpophalangeal joint (Figure 6.1). The cartilage at *Site 1* was subjected to intermittent high peak loading conditions. In contrast, cartilage at *site 2* was exposed to lower-level but constant joint loading [100].

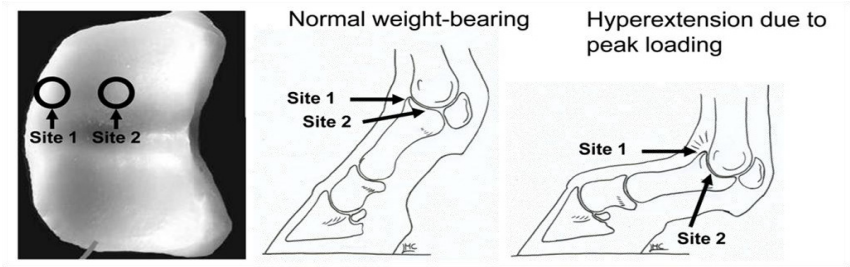

Figure 6.1*: Sampling sites for cylindrical osteochondral plugs from the articular surface of the proximal phalangeal bone of the left metacarpophalangeal joint. Located at the medial dorsal margin of the joint surface, site 1 is not loaded during standing or in a slowly moving animal, but is subjected to high intermittent peak loading during overextension when an animal moves at high speeds and jumps. Site 2, in the mediocentral area, is continually loaded during weight bearing, but experiences lower peak forces than site 1 [100].*

Newborn (n = 6), immature (age range 5 to 11 months, n = 15), and adult horses (age range 6 to 10 years, n=8) were used. Osteochondral plugs (diameter 4 mm) were removed from the two predefined locations using a custom-built hollow drill in a random orientation, but with the long axis perpendicular to the articular surface of the proximal phalangeal bone. The plugs were subsequently sectioned into two equal halves consisting of AC and approximately 10 to 20 mm of SB. One half, used for the analysis in the present study, was decalcified in 10% ethylenediaminetetraacetic acid (EDTA) supplemented with 4% formaldehyde and 0.1M sodium phosphate buffer (pH 7.4) for 12 days at 4°C, and subsequently dehydrated in ethanol and

embedded in paraffin [30]. The study protocol was approved by the ethical committees of Utrecht University, the Netherlands, and Massey University, New Zealand.

In study II, cortical bone samples taken from the mid-diaphysis of the humerus of New Zealand white rabbits (five different age groups: newborn, 11 days, 1 month, 3 months and 6 months, 7 samples per group) were used. The samples were cut (8 mm), dehydrated in ethanol and embedded in polymethylmethacrylate (PMMA). The study protocol was approved by the Animal Care and Use Committee of University of Eastern Finland.

In study III, two separate sets of AC samples were investigated. The first sample set consisted of bovine AC samples (n=8), which were obtained from the local abattoir. The animals were 1-3 years old, and the samples were taken from the lateral upper quadrants of the patellae (13 mm diameter, length 2-3 cm). This sample set originated from an earlier study by Rieppo et al [98]. The second set consisted of AC samples (4 mm diameter, 3 mm depth) from the patellar grooves of 6-months-old New Zealand rabbits. The intact AC (n=7) from the left knee was used in study III and as a control group for the repaired osteochondral lesions (4 mm in diameter, 3 mm in depth, n=6) that were surgically created in the right knee of animals in study IV. In the first repaired group of samples in study IV, the lesions were left empty to heal spontaneously (n=6). In the second group, the lesions were surgically repaired using autologous pre-cultivated chondrocytes in human type II collagen gel (CG)(n=6). The surgical procedures have been described in more detail by Pulkkinen et al [99, 101]. The animals were sacrificed 6 months post-operatively and osteochondral specimens were collected from the repair sites. In the second group, control AC samples were harvested from the intact AC at an adjacent site to the lesion.

All samples in studies III-IV were fixed in 10% formalin, decalcified with EDTA, dehydrated and embedded in paraffin as described earlier. Paraffin was dissolved from all samples with xylene prior to the FTIR-MSP measurements.

6.2 FTIR-MSP MEASUREMENTS

In the FTIR-MSP measurements, the samples were mounted on either ZnSe or BaF_2 windows that were transparent to infrared light. All measurements were performed in the transmittance mode with a FTIR imaging system (Perkin Elmer Spotlight 300, Perkin Elmer, Shelton, Colorado). This system comprises a classical FTIR spectrometer coupled with a light microscope, a computer-controlled sample stage, and a linear array detector. The humidity of the measurement chamber was continuously monitored to be 0% using CO_2-free air to standardize the measurement conditions (Parker Balston, Haverhill, MA, USA). For background correction, one measurement of an empty site with only the window present was used. Several scans were averaged. After the measurements, all spectra were converted to ABS values for further analysis using Cytospec software (Cytospec Inc., Boston, USA). In all studies, spectra of pixels containing no tissue were set to zero and excluded from the analysis.

6.2.1 FTIR-MSP of SB (Study I)

Paraffin was dissolved from 5-μm-thick sections with xylene prior to the measurements. Infrared spectra were acquired with spatial resolution of 25 μm and spectral resolution of 4 cm^{-1}. Two repeated scans were used for data collection with the spectra being collected over the region of 2000 to 700 cm^{-1}.

In the analysis, two rectangular regions of interest (ROIs) were extracted from each sample (Figure 6.2). Region 1 (R1) was

chosen just below the cartilage-bone interface, and region 2 (R2) was chosen in the deeper SB.

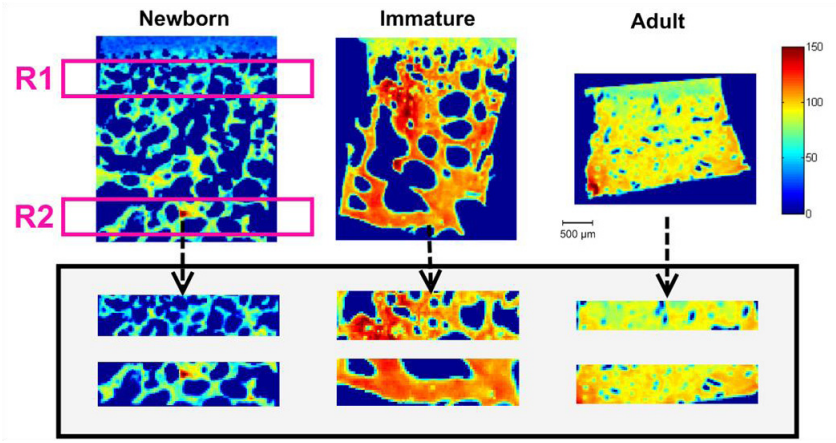

***Figure 6.2**: Amide I absorbance images of newborn (left), immature (middle), and adult (right) equine subchondral bone samples. Region 1 (R1) was chosen just below the articular cartilage-bone interface, and region 2 (R2) was located in the deeper subchondral bone.*

An equal number of spectra from each region (N=300) was used for analysis. When the regions contained more than 300 bone spectra, the pixels were chosen randomly from the data matrix of the region. Spectra from all samples were pooled together into one data matrix. In quantitative analysis, average spectra were calculated for each region.

6.2.2 FTIR-MSP of cortical bone (Study II)

For each 3-µm-thick sample, five rectangular measurement areas (on average 100 x 100 mm) were manually selected based on visual observation of the whole sample section. A pixel size of 6.25 µm and spectral resolution of 4 cm^{-1} were used for FTIR-MSP measurements. Thus, over 1125 spectra covering an area of more than ~44000 mm^2 were collected from five measurements

on each sample. Eight repeated scans were performed, and the spectra were collected over the region of 4000– 720 cm^{-1}.

6.2.3 FTIR-MSP of AC (Study III, IV)

Infrared spectra were acquired from 5-μm-thick sections of rabbit and bovine AC. The pixel size for measurements was set to 25 μm and the spectral resolution to 4 cm^{-1}. The numbers of repeated scans were eight for the bovine samples and four for the rabbit samples, and the measured spectral region was 4000-720 cm^{-1}. An area covering the whole prepared section for each sample was analyzed. After the quality tests and removal of the data containing bone spectra, the spectral map of each rabbit AC sample consisted of 500-5300 spectra. For bovine AC samples, a 400-mm-wide area was measured from the cartilage surface to cartilage-bone junction using the same setup.

6.3 PREPROCESSING PROCEDURES

6.3.1 Quality tests

Quality tests were performed on AC samples in Studies III and IV. The tests consisted of three separate procedures: 1) removal of outliers using 3-sigma rule; 2) confirmation that the SNR of each spectra was high enough; and 3) confirmation that the maximum absorbance in the region 1800-1600 cm^{-1} was higher than 0.1. The last step was conducted to eliminate spectra of low absorbance.

6.3.2 Mathematical removal of bone tissue from samples

Bone spectra were removed from images in studies III and IV by applying PCA. PCA is based on an evaluation of the total variance in the data and projection of initial variables to a new

coordinate system of orthogonal principal components (PCs). It is used for reduction of data dimensionality and classification into a new PC-based coordinate system. PCs are sorted in order of descending percentage of explained variance. The first PC is the most important one and accounts most of the variance in data. PCA was applied on each image using the CHO spectral region of 986-1140 cm^{-1}, which is known to be attributable to the PG vibrations (Figure 6.3).

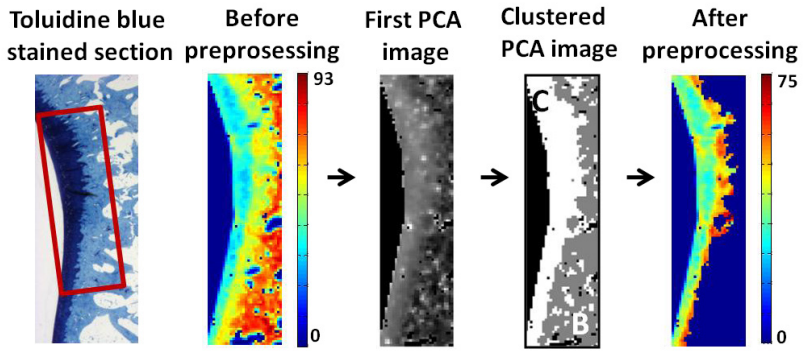

***Figure 6.3**: Removing of bone spectra and outliers using principal component analysis (PCA). The first PCA image was built using the CHO (968-1140 cm^{-1}) region of infrared spectra and clustered using the FCM. Pixels assigned to bone (B) and outliers were removed remaining only cartilage (C). Integrated collagen absorbance images show the sample before and after preprocessing.*

The first PC was most discriminative between cartilage and bone. A gray-scaled projection PC image was created for each sample based on first PC, and was clustered with FCM clustering algorithm using two clusters (represented AC and bone). Pixels, which were assigned to bone tissue, and pixels of the AC cluster located in the bone area were removed.

6.3.3 Removal of PMMA contribution from infrared spectra of cortical bone

In study II, absorbance of the pure PMMA was used to normalize and minimize the effect of section thickness. This was done by scaling the PMMA spectra to the bone spectra using the PMMA peak at 1728 cm^{-1} and subtracting it from the bone spectra [102].

6.3.4 Correction for scattering effects

Cartilage spectra in studies III and IV were corrected for scattering effects using Resonant Mie Scattering algorithm [103]. A reference spectrum for correction was calculated as the global average spectrum of all samples (separately for the rabbit and bovine sample sets).

6.3.4 Correction for water vapor and CO_2

The remaining contribution of water vapor and CO_2 to the AC spectra in studies III and IV were removed with the algorithm described by Bruun et al [104]. This was achieved by measuring a 2550x2875 μm^2 area from an empty window under gradually decreasing water vapor and CO_2 concentrations, resulting in ~11700 infrared absorption spectra in total. The CO_2 gas purging was started after closure of the sample chamber, resulting in a reduction in the water vapor concentration from 22% to 0%. In addition, a slight increase of the temperature inside the chamber (~0.6ºC) was observed during the measurements. The collected spectra were used for estimation of two primary gas spectra, which were subsequently used for correction as explained by Bruun et al [104].

6.3.5 Normalization and derivation of spectra

Only the spectral region that included the main absorption bands of bone or AC at the fingerprint region was chosen for the analysis. Hence, spectra were restricted to 720–2000 cm^{-1} in all studies.

In studies I and II, second derivative spectra were calculated from the raw spectra using the Savitzky-Golay algorithm with nine-smoothing points [105]. Subsequently, derivative spectra were scaled so that the sum of squared deviation over the indicated wavelength equaled unity (vector normalization). The second derivative was used to resolve and locate peaks in the spectra for cluster analysis. In Studies III and IV, the raw spectra were normalized to the vector length before the cluster analysis.

6.4 UNIVARIATE ANALYSIS

In the quantitative FTIR analysis, the representative spectrum was first extracted from the data by averaging the set of spectra and it was then baseline-corrected. Spectra of pixels containing no tissue (background spectra) were set to zero and excluded from the calculations of the average spectrum. In study I, average spectra for each sample were extracted from two rectangular ROIs (Figure 6.2). In studies III and IV, mean spectra were calculated for each cluster, constructed based on the AI spectral region.

Further, a quantitative estimation of basic compositional parameters of bone and AC was done by integrating representative spectra over the particular peak. In the AC samples, the collagen content was quantified by measuring the integrated absorbance of the AI peak (1584–1720 cm^{-1}), the AII peak (1500–1584 cm^{-1}); and the PG content – from CHO peak (968-1140 cm^{-1}). Additionally, averaged CHO/AI and AI/AII peak ratios were calculated in order to estimate relative PG

content and orientation of collagen fibers, respectively. A peak at 1338 cm^{-1}, i.e., calculation of 1338 cm^{-1}/AII ratio, was used to estimate the collagen integrity parameter [15].

The composition of bone samples was analyzed by measuring spatial collagen content using the AI and AII peaks. Additionally, the collagen maturity, or the ratio of immature to mature collagen-cross links, was estimated by using a combination of the second derivative spectra and peak fitting to determine the sub-peaks at 1660 cm^{-1} and 1690 cm^{-1} [59].

6.5 STATISTICAL ANALYSIS

Statistical significance of differences in estimated values between different sample groups was tested using the non-parametric Mann-Whitney U-test in SPSS 11.5-17.0 software, SPSS Inc., Chicago, IL, USA. Means and standard deviation values, or medians and a 95% confidence intervals were used in data presentation in study I and studies III-IV, respectively.

6.6 CLUSTER ANALYSIS

6.6.1 K-means cluster analysis (Study I)

In study I, k-means clustering was used to classify SB samples into age groups. The number of clusters was set to three, corresponding to the number of constructed groups of samples. The squared Euclidean distance was used as a measure of dissimilarity between spectra. The k-means algorithm was run 50 times with different random initialization, and the solution with the smallest MSE value was chosen. Spectra of the two ROIs from the two loading sites were analyzed separately. Clustering was performed both on mean spectra and pixel-by-pixel for each sample. In the first approach, spectra from all samples were pooled together into one data matrix, and

normalized second derivative spectra were clustered altogether using k-means. The spectral region between 1200 and 1720 cm^{-1} was selected for clustering.

6.6.2 Performance of different clustering algorithms (Study II)

In study II, average normalized second derivative spectra of each sample were used as the input for different clustering algorithms: k-means, FCM and HCA clustering. The maximum number of iterations and number of repetitions for FCM and k-means were set to 1000 and 100, respectively. Each method was used to obtain three clusters (newborn, immature and adult) to represent the different stages of biological bone maturation. Each method utilized the whole spectral range 720-2000 cm^{-1}, as well as the spectral regions of AI, phosphate (900–1200 cm^{-1}) and carbonate (850–890 cm^{-1}) peaks.

The validity of clustering results was evaluated by the Rand index value and the overall MSE level. Two to eight clusters were tested for each spectral region, and MSE analysis was used to determine the actual number of distinct groups inside the data.

Discriminant analysis (DA) is a multivariate classification technique that can classify objects into two or more known groups on the basis of several variables [85, 106]. The goal of the analysis is to find the discriminant function (DF) which can differentiate between the groups, i.e., maximize the difference between the mean of the groups. With more than two groups, one can obtain more than one DF. The first DF is the one which maximally separates the groups (produces the largest ratio of among-groups to within-groups sum of squares on the resulting D scores). The second DF, orthogonal to the first value, maximally separates the groups based on the variance that was not yet explained by the first DF.

DA was used in study II to combine clustering results from three spectral regions and to define the contribution of each spectral region to the overall discrimination results. For that purpose, final cluster memberships were linearly combined into three groups of observations using DA [107]. Two DF were calculated for each of the three cluster methods. The performance of a DA was evaluated by estimating error rates (probabilities of misclassification) using leave-one-out cross-validation. DA was performed using SSPS software [85] (v.15, Chicago, IL, USA).

6.6.3 Fuzzy c-means cluster analysis of FTIR-MSP in cartilage (studies III and IV)

Several spectral regions were investigated in studies III and IV: complete amide region (1200-1720 cm^{-1}, referred as A), AI region (1585-1720 cm^{-1}), AII region (1510-1584 cm^{-1}), and CHO region (968-1140 cm^{-1}). When the complete amide region was used, the spectral region of 1300-1490 cm^{-1} was excluded. This was done to eliminate possible overlapping with the remaining spectra from the embedding medium [108].

In study III, two species, i.e., rabbits and bovine AC were analyzed independently. FCM clustering was performed pixel-by-pixel on normalized raw FTIR-MSP images. Three clusters were obtained for each spectral region (A, AI, AII and CHO), considered to represent three main histological zones of AC (SZ, MZ and DZ) within each sample.

In study IV, spectral images of the repaired and control rabbit AC samples were clustered: 1) independently using three clusters, or 2) together from each rabbit using four clusters. In both studies, the maximum number of iterations and number of repetitions of FCM were set to 1000 and 100, respectively. The clustering results were examined for each spectral region.

First, performance of clustering was evaluated by calculating the percentage of the correct clustering assignments for repaired and intact clusters. Histological images were used as a reference for the tissue type. The overall performance of clustering was expressed as a number of correct pixel assignments divided by the total number of pixels. Performance was compared for two types of repair and spectral regions used for clustering.

Second, the differences between the first and second largest membership degree values for each pixel were calculated to evaluate the uncertainty of clustering. Average differences were compared for intact and repaired clusters. A difference value close to 0 means that the first and the second largest membership degree values are almost equal. In this case, the pixel can be assigned with almost identical probability to both clusters represented by those values. Otherwise, a value close to 1 means that the first largest membership degree value is much larger than the others and clustering is very distinctive.

Qualitative differences between clusters were analyzed using raw average spectra of clusters. The second derivatives were calculated using the Savitzky-Golay algorithm with nine smoothing points and were used to enhance resolution and to locate the differences in positions of the peaks.

6.7 REFERENCE DATA

In studies III and IV, the results from the analysis of the same samples using different analysis techniques in other studies were used as reference for qualitative comparison. The structural integrity of the intact and repaired AC was evaluated by using histology [99], immuno-histological staining [99] and PLM [43]. Bovine AC sections were stained with safranin O to reveal the spatial PG concentrations [109]. In rabbit samples, toluidine blue was used to stain PGs and type I collagen

antibody was used for immunohistochemical staining of type I collagen. PLM was used to determine the organization of collagen network. By analyzing the histological images, the borders between the repaired areas in the SZ and MZ and the more hyaline-like AC tissue in the deeper zones within each repaired sample were manually determined. The modified O'Driscoll score for cartilage repair was used to evaluate the quality of the repair [13]. In PLM analysis, the parallelism index was calculated to evaluate the degree of parallelism of the collagen fibers, in other words, the organization of the collagen network. The histological and PLM methods and results have been published in detail by Pulkkinen et al. [99] and Viren et al. [13].

7 Results

This chapter summarizes the most relevant results of studies I-IV. The complete results are presented in the original articles I-IV.

7.1 DISCRIMINATION OF BONE AT DIFFERENT MATURATION STAGES WITH CLUSTER ANALYSIS

Studies I and II highlighted the ability of FTIR-MSP coupled with cluster analysis to detect changes in the collagen matrix of decalcified SB in horses and rabbit cortical bone during maturation and growth.

The overall clustering performance results for equine subchondral and rabbit cortical bone are shown in Figure 7.1. In SB at each studied site and region, newborn and adult samples were classified into two different groups using the amide region, while immature samples were assigned partially to each of the three age groups. The accuracy of clustering slightly differed for the SB, which was subjected to different types of loads and also differed for the location of the sample relative to the closeness to AC. Second derivatives revealed that the AI peak had its highest value at different wavenumbers in different clusters. The shape of the peak was also slightly different.

For rabbit cortical bone, both the organic matrix and mineral component were included in the cluster analysis. All cluster algorithms classified samples rather accurately into different age groups (Figure 7.1). The newborn samples were well separated from the other age groups by carbonate and phosphate spectral regions. The adult samples were well separated from the other age groups by carbonate and amide I, while some samples were

assigned to another group if phosphate was used in the cluster analysis.

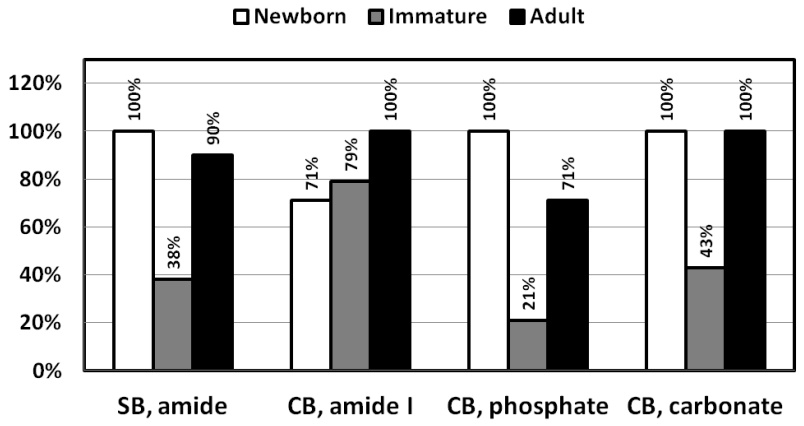

***Figure 7.1**: Performance of cluster analysis applied to FTIR-MSP spectra. Amide infrared region was used for equine subchondral bone (SB) (results from Site 2, region 2). Amide I, phosphate and carbonate spectral regions were used in analysis of rabbit cortical bone (CB). The bars show percentage of correct assignments of samples to their age groups.*

When conducting pixel-by-pixel clustering of FTIR-MSP images of samples, the results of clustering were similar to those obtained by clustering the average spectra (Figures 7.2-7.3).

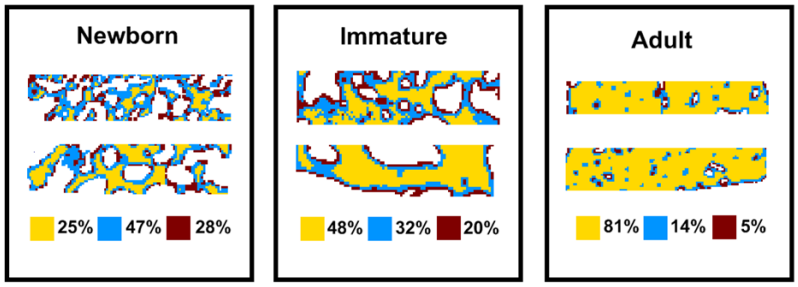

***Figure 7.2**: False color code clustering maps of representative samples from three age groups (newborn, immature and adult) of equine subchondral bone. Clustering is performed pixel-by-pixel using three clusters. The percentages represent the number of pixels assigned to the particular cluster divided by the total number of pixels in the sample.*

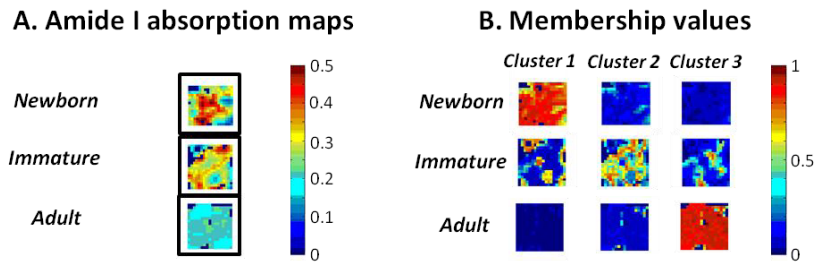

***Figure 7.3**: (A) Amide I absorbance images of one representative area in one sample of each age group, i.e. newborn, immature and adult bone sample. Spectra excluded by the quality tests appear dark blue; (B) Color representation of membership values obtained by FCM algorithm (amide I region, 3 clusters). A strong association of samples with their primary clusters can be seen.*

7.2 VALIDATION OF THE MOST ACCURATE CLUSTERING ALGORITHM

In study II, attempts to determine optimal clustering algorithm for differentiation of maturing cortical bone were made. Although all algorithms (k-means, FCM, HCA) produced good discrimination (Figure 7.4), for all algorithms, the highest variance in the results was observed in the AI region. Cluster analysis highlighted several spectral differences between the spectra of clusters (Figure 7.5).

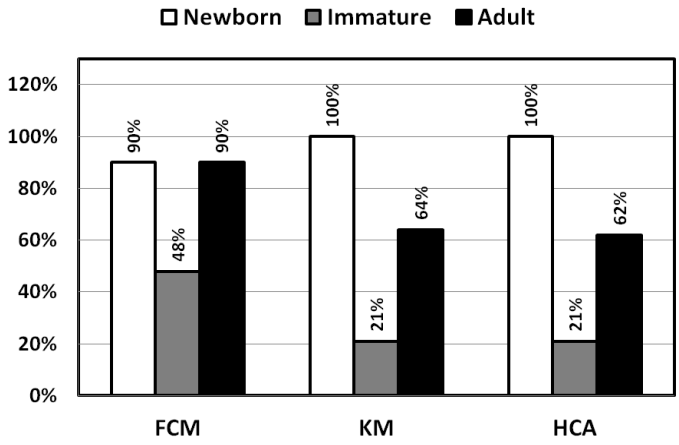

***Figure 7.4**: Classification performance of fuzzy c-means (FCM), k-means (KM) and hierarchical (HCA) clustering algorithms. Clustering performance is indicated by the average value obtained using carbonate, phosphate and amide I spectral regions individually.*

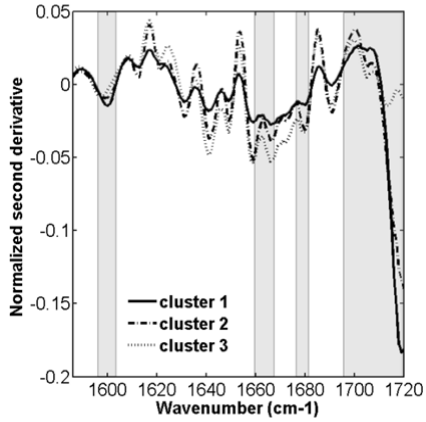

***Figure 7.5**: Average second derivative spectra for three clusters, as obtained using FCM clustering (amide I region). Spectral differences are highlighted.*

Using any spectral region FCM showed the best performance, as compared to HCA and k-means algorithms. The results suggested that three clusters were the best choice for all clustering methods. The highest Rand index values were obtained for the carbonate region followed by the AI region and then the phosphate region.

DA analysis showed that age groups differed significantly in all regions ($p < 0.001$). The adult samples were separated by carbonate region from the other two groups. By combining all clustering results, the ability to predict the correct age group of the samples increased. With the use of the discriminant function, 100% of the NB and adult samples and 43% of the immature samples were classified correctly in KM and HCA clustering. For FCM, correct classifications were obtained for 71% of the NB samples, 93% of the immature samples and 100% of the adult samples. Thus, in total, 77% (KM and HCA) and 91% (FCM) of the samples were classified correctly. In general, leave-one-out cross-validation showed that DA performed well for all clustering methods.

7.3 DETERMINATION OF HISTOLOGICAL LAYERS IN INTACT ARTICULAR CARTILAGE

In study III, FCM clustering revealed layered structure for intact AC of bovine and rabbits. Three distinct layers were found in rabbits AC. For bovine samples, the cluster in the SZ also appeared as a layer in the deeper cartilage, producing more variation in the clustering structure. Moreover, careful investigation of fuzzy maps revealed a distinct clustering of the rabbit AC, with a strong association of each pixel to its major cluster (Figure 7.6A). In contrast, pixels in the SZ of bovine AC were almost 50% associated with two bottom clusters (Figure 7.6B). In general, for all samples in both groups, clustering using the amide region produced more distinct results than those using the CHO region.

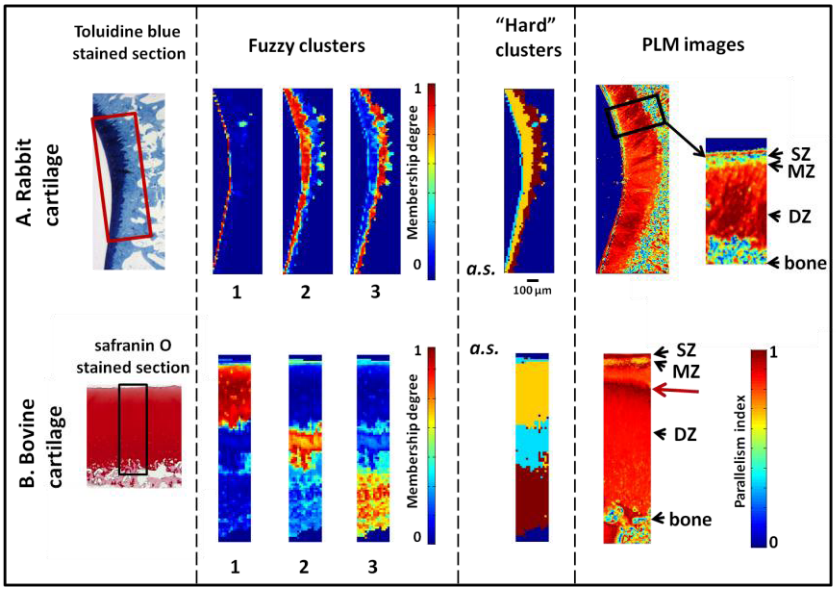

***Figure 7.6**: Figure shows fuzzy clustering maps of a (A) rabbit and (B) bovine articular cartilage samples obtained using the amide spectral region. Degrees of membership at each pixel, in each of the three clusters are given, and the gradual transition between clusters is revealed. The "hard" clustering maps were calculated based on the fuzzy maps, such that each pixel is assigned to the cluster with the maximum membership value. The parallelism index of collagen fibers calculated from polarized light microscopy (PLM) images. The red arrow indicates an extra lamina in the bovine sample. The magnified image from the rabbit sample displays the histological zones determined by PLM. The toluidine blue and safranin O stained histological images, sensitive to cartilage PGs, are shown as references.*

When clustering results were compared with PLM, clearly similar features were revealed including three distinct layers visible in both PLM and clustering images of the rabbit cartilage. In the bovine sample, an extra lamina is highlighted both by PLM and clustering (Figure 7.6B). However, the thickness of the layers and location of their boundaries do not match exactly between PLM images and cluster images.

Spectral differences between the clusters were investigated using the average spectra of clusters. The major infrared peaks

for AC were observed in all samples, regardless of species. However, minor non-regular shifts (2-4 cm^{-1}) were observed from cluster to cluster and from sample to sample. Other minor peaks appeared only in few samples. A shift of the AI peak position in both directions (around 1662 cm^{-1}) was observed between three clusters. The shift was greater in the bovine samples, varying in location around 1666-1669 cm^{-1} in SZ cluster, 1660-1662 cm^{-1} in the MZ cluster and 1662-1664 cm^{-1} in the DZ cluster. Moreover, although the location of AI, AII and AIII peaks changed almost randomly, the differences in the shapes of the peaks' shoulders were identified after careful visual examination.

7.4 DISCRIMINATION BETWEEN NORMAL AND REPAIRED ARTICULAR CARTILAGE

In study IV, the surgically and spontaneously repaired rabbit AC was separated from the intact AC using FTIR-MSP with FCM cluster analysis. When clustering was conducted independently, intact samples revealed a layered clustering structure from the superficial to deep zone, similar to that found in study III. For spontaneously and collagen II gel repaired AC, superficial fibrous and more hyaline-like deeper cartilage sites were separated in different clusters.

When the intact and repaired sections from each rabbit were clustered together at the same time, the discrimination between the intact and repaired AC was clear when using each spectral region. However, careful manual assessment indicated that the best separation could be achieved by using AI and CHO regions. FCM results for typical spontaneously and collagen gel repaired samples, paired with the corresponding intact AC samples, are shown in Figure 7.7 and Figure 7.8, respectively. The differences in clustering structure using AI and CHO regions could be visually observed.

In all samples, generally, repaired and intact tissues were separated into different clusters. They were represented by two clusters each, marked with a different color on the cluster maps (green and blue for intact, orange for repaired regions in the superficial and middle zones and red for the cartilage-like repaired region in the deeper zone). Two separate regions of repair were detected in each sample with repaired AC: the "repaired 1 cluster" located in the superficial and middle zones and the "repaired 2 cluster" located in the deeper zone.

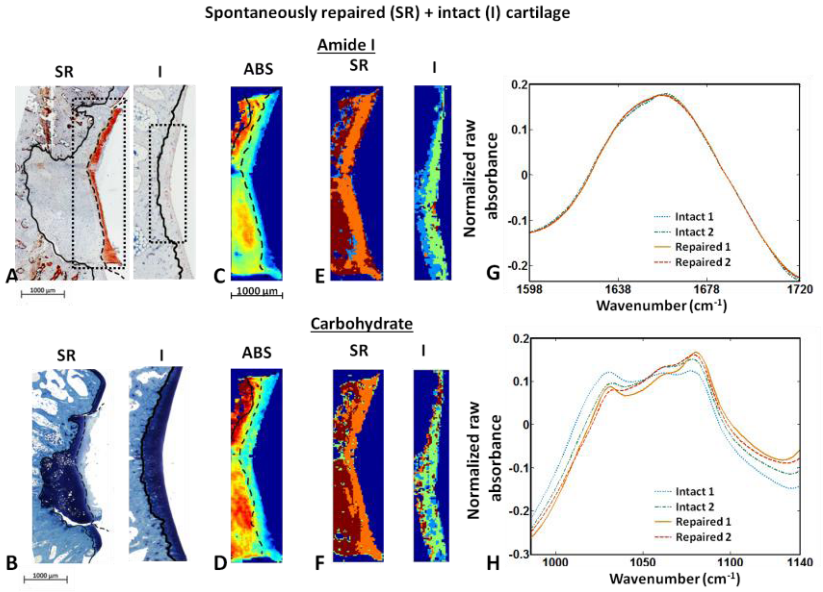

***Figure 7.7**: Fuzzy c-means clustering for a pair of spontaneously repaired and intact samples in a rabbit. (A,B) Collagen type I and toluidine blue histological images of the corresponding samples with the indicated areas of repair: the solid line is the border between cartilage and bone, dashed line separates the region of superficial repair; (C,D) Infrared absorption images based on amide I and carbohydrate spectral regions; (E,F) False-color cluster maps and (G,F) corresponding average spectra of clusters.*

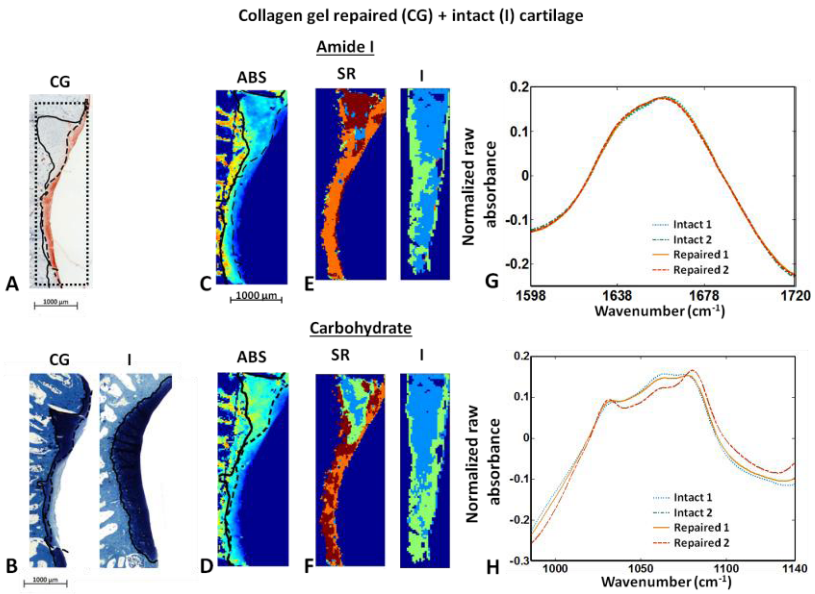

***Figure 7.8**: Fuzzy c-means clustering for a pair of gel repaired and intact samples in a rabbit. (A,B) Collagen type I and toluidine blue histological images of the corresponding samples with the indicated areas of repair: the solid line is the border between cartilage and bone, dashed line separates the region of superficial repair; (C,D) Infrared absorption images based on amide I and carbohydrate spectral regions; (E,F) False-color cluster maps and (G,F) the corresponding average spectra of clusters.*

Based on the performance analysis from both spectral regions, the repair tissue in spontaneously repaired samples was more accurately identified than that in the samples repaired using collagen gel. The intact tissue was more clearly identified using the AI region. Overall clustering performance was the same for both tissue types using the AI region (81%). However, somewhat a higher uncertainty in discrimination was noted when using AI region for the spontaneously repaired samples. The most distinct clustering was achieved by using CHO region for spontaneously repaired samples (82% compared to 75% for collagen gel repaired samples). These results indicate that the spontaneously repaired tissue was better distinguished from the intact AC compared to the collagen gel repaired tissue.

For the qualitative comparison of differences, the mean spectra of one intact and two repaired clusters were calculated as average spectra of all samples in each group. At the AI region, both repair clusters in collagen gel repaired tissue were different from the other clusters, by having slightly shifted sub-peaks at 1654-1656 cm^{-1} and 1680-1682 cm^{-1} and opposite curvatures of spectra of the region 1708-1720 cm^{-1}. At the CH region, repaired cluster 1 showed lower absorption values before and higher absorption values after 1078 cm^{-1}, as compared to other clusters. For the cartilage repaired using collagen gel, intact cluster spectra were highly similar to the repaired cluster 2 spectra, while all spectra in the spontaneously repaired cartilage were distinguishable.

The accuracy of clustering, when compared with histology, varied slightly from sample to sample. Different clustering structures as obtained with AI and CHO regions corresponded rather well to different structures of repaired tissue, as evaluated using the type I collagen and PG histological staining.

Average O'Driscoll score was the highest for the spontaneously repaired cartilage. The parallelism index calculated from PLM data was lower in the repaired AC in both repair groups, indicating mostly a random orientation of collagen fibers throughout the cartilage depth [13].

Quantitatively, in the spontaneously repaired AC, cluster 1 had lower PG concentration than cluster 2 and intact AC as estimated by CHO/AI ratio. In the AC with collagen gel repair, estimated collagen and PG contents were at their lowest in the repaired cluster 1, as compared to the intact cartilage. Moreover, estimated PG content in the repaired cluster 1 of the spontaneously repaired samples was significantly higher than that in the collagen gel repaired samples. However, no other differences were observed.

8 Discussion

Clustering has been applied in various fields of research and engineering. However, the first FTIR study on articular cartilage employing cluster analysis was conducted by Rieppo et al in 2007 [110]. Those preliminary results revealed a good potential of this technique to characterize qualitative differences in material properties of articular cartilage. The authors stated that this technique could open a new era for FTIR-MSP of cartilage. For the best of our knowledge, no clustering analysis of FTIR-MSP of bone tissue has earlier been conducted. The present work introduced a new approach to study depth-wise structure and composition of AC, as well as tissue changes in cartilage repair. This appears to be also the first work employing clustering techniques, based on its collagen and mineral contents, to differentiate bone of varying age.

8.1 CLUSTER ANALYSIS CAN IDENTIFY SUBTLE CHANGES IN FTIR-MSP SPECTRA

Each study showed that cluster analysis can identify minor differences in FTIR-MSP spectra, i.e. changes that are not apparent under visual analysis of the spectra. Moreover, one needs no expert assessment on tissue quality or type beforehand, since clustering algorithms work with minimal human interference. Although the correct age classifications for bone samples in studies I-II were known, this information was not available for the clustering algorithm. Nonetheless, it was able to distinguish effectively the newborn from the adult bone samples. Clustering also revealed the inhomogeneous structure of the immature group by assigning some samples to either the immature or adult group. The spectral differences between the clusters were characterized by peak shifts and shape changes in

the spectra. The shifts of collagen bands, as observed by some authors [31], are believed to be sensitive to the helicoidal arrangement of proteins [58, 111] and collagen crosslinks [59]. These qualitative changes could reflect the change in the collagen protein secondary structure.

In study III, FCM cluster analysis qualitatively demonstrated the layered structure of intact rabbit and bovine AC tissue. However, the observed laminar structure did not match exactly with that obtained using the PLM. This extra lamina in the DZ of bovine samples, as found using FCM cluster analysis [Figure 7.7D], has been also observed in several previous studies of bovine AC [98, 112, 113] and samples from beagle dogs [114]. Clustering indicated the high similarity of this lamina to the SZ cluster. Moreover, it was possible to identify a shift of the AI peak from 1600 cm^{-1} to higher wavenumbers of 1666-1668 cm^{-1} in this cluster. This shift could reflect changes in the collagen quality (renaturation) and organization [96].

In study IV, FCM detected two types of repair tissue in each AC sample and discriminated effectively the repair tissue from the intact tissue. In samples with collagen gel repair slight shifts of two sub-peaks in AI region were noted in clusters for neotissue. These sub-peaks are attributable to secondary protein vibrations [56, 58]. The observed variations in the curvature of spectra in the region 1708-1720 cm^{-1} may originate from the non-hydrogen bonded β-turns proteins. FTIR bands in the CHO region assigned to C–O stretching vibrations of the carbohydrate residues are present in collagen and in PGs and could be related to AI of PGs [23]. The change in the location of spectra at CHO region after 1078 cm^{-1} may be evidence of differences in the collagen contribution.

These findings, indeed, highlight the great potential of clustering for detecting small differences in infrared spectra. However, no definite conclusion could be made based on these

differences, except that they were sufficiently clear to enable successful clustering.

8.2 INTERPRETATION OF DISCRIMINATION RESULTS

It is most challenging to explain the present clustering results from the biological point of view. In order to appreciate the real potential of clustering, the specific changes in the infrared spectra need to be associated with the corresponding biological and molecular changes. Fortunately, most of the IR spectral bands can be assigned to a particular part of a molecule. However, several factors complicate designation of the particular wavelengths in the group frequencies [52]. Thus, it is difficult to observe minor changes in the shape and location of the infrared bands between classes.

In the process of visual examination of average spectra in different clusters, specific differences that distinguished one cluster from the others were discovered. In order to enhance those differences, second derivatives have been used [115]. In clustering, major differences in the amide region could be related to secondary protein vibrations. Shifts and relative changes in the intensity of infrared sub-peaks point to changes in the concentration of proteins present in the different clusters. However, assignment of particular absorbance to specific secondary protein structure needs to be approached with caution due to the overlapping between vibrations and dependence on hydrogen bonding [56, 58].

These difficulties in interpretation and treatment of FTIR data have been claimed to be the main drawback in utilizing FTIR-MSP in clinical practice [116].

8.3 THE MOST ACCURATE CLUSTER ANALYSIS APPROACH BASED ON DISCRIMINATION OF BONE

Different clustering algorithms have their own pros and cons and exhibit different scalability, sensitivity to outliers, or ability to handle clusters with specified shapes. Hence, no single clustering method can be optimal for all data sets. For instance, if the data exhibits a hierarchical structure, the hierarchical clustering methods will probably be appropriate, while partition algorithms, such as k-means, will not be effective with this type of information [117]. K-means and FCM clustering algorithms also require setting a specific number of clusters in advance, which is a non-trivial question.

Lasch et al. compared HCA, KM and, the less used FCM clustering techniques on the IR spectra of cancerous tissue [118]. In that study, HCA could identify different tissue types and achieved the best agreement with histopathology. However, the major drawback of HCA is its extreme computational complexity, especially with large-scale datasets [86, 118, 119]. It is also known that HCA is not capable of correcting possible misclassification [117], and thus, the algorithm is more sensitive to noise and outliers [86].

In study II, k-means, FCM and HCA could separate rabbit bone samples into different age groups based on their infrared spectra. The better results produced by FCM clustering can be explained by the fact that the ''hard'' clustering methods (k-means, HCA) assign a sample to only one group with 100% probability, while FCM allows a sample to belong to several clusters with different degrees of membership. If a spectrum at a pixel contains contributions from two or more constituents, e.g., due to different maturation stages within the tissue, it does not completely match any particular constituent or tissue type. Thus, these spectra cannot belong exclusively to any single cluster. The fuzzy membership approaches can ideally handle this situation and allow more specific and robust classification

[97, 120]. Therefore, FCM clustering may be more reliable, as compared to the ''hard'' algorithms, in many applications related to medical diagnostics.

In the current set of bone samples (study II), three samples appeared not to associate strongly with any cluster. Cluster membership values for these samples were close to 0.5 and suggest continuing maturation. This could not be captured by clustering using ''hard'' methods. Using other methods, including microCT and nanoindentation analysis [121], different BMD and mechanical properties were measured for these specific samples. Therefore, these deviations in the clustering are obviously related to true biological variation between samples.

Based on these results, only FCM clustering algorithm was used for discrimination in the following studies. It successfully identified histological layers in intact AC and discriminated intact and repaired AC.

8.5 DEVELOPING AN ALGORITHM FOR CLUSTER ANALYSIS OF FTIR-MSP

The aim for the present studies was to identify the best approach for clustering FTIR-MSP data of biological tissues. Further, it was hoped to maximize the discrimination ability by optimal preparation and analysis of the data, as well as interpretation of the clustering results. In each study, an initial algorithm was further developed by testing different approaches and settings, selecting only those which improved classification and enabled faster processing of extensive amount of data. The final algorithm comprises the following procedures (explained in detail in Chapter 6):

1. Pre-processing
    a. Quality tests
    b. Mathematical removing of background and secondary data from samples
    c. Removing contribution of the embedding medium
    d. Correction for scattering effect
    e. Correction for water vapor and CO_2
    f. Derivation and normalization of spectra
2. Cluster analysis
    a. Select suitable algorithm
    b. Set up a number of clusters and parameters for algorithm
    c. Run analysis
    d. Visualize results
    e. Validate results
3. Interpret results

The important step prior to analysis of spectral data is to standardize representation of the data and check for any inconsistencies [23, 60]. The pre-processing procedures include evaluation of data quality, definition of areas and spectral regions of interest, baseline correction, normalization and derivations, if needed [52, 60, 122]. The pre-processing is necessary in order to emphasize the biological differences and to remove variations caused by measurement conditions [23].

The overall high quality of the spectra is essential when spectra are being analyzed using multivariate techniques. Low-quality spectra must typically be removed [115]. A spatial resolution of 25μm was used for the FTIR-MSP. This resolution provides spectra of higher quality than when using a smaller pixel size, and thus, a superior SNR. Moreover, it is advisable to process only those spectra, which are informative for the study. By removing all background and other tissue spectra, a reduction in the time needed to run the cluster analysis decreases, and also a better focus on the data of interest can be achieved.

The possible contribution of the embedding substrate to the infrared absorption of the tissue must be removed. Therefore, those spectral regions where this kind of removal is impossible must be excluded from the analysis. One can also consider further correction for the samples thickness. However, this effect can be eliminated by the use of normalized second derivatives, instead of raw spectra. Importantly, clustering applied to the raw spectra mainly distinguishes the concentrations of constituents while derivation of the spectra prior to clustering differentiates the structural difefrences. In addition, normalization of spectra prior to the clustering removes quantitative information and makes possible to study only qualitative differences in the shape of spectra.

It is important to standardize measurement conditions in terms of temperature, composition and humidity of air inside the measurement chamber in FTIR-MSP. However, all remaining effects must be eliminated mathematically. Especially, one needs to pay attention to the water vapor contribution since water is a very good absorber of infrared light in the amide region. This should be performed very carefully to avoid distortion of initial spectra and, thus, changing its unique features.

Standardization, or normalization, scales the data so that the units used for measurements do not affect the similarities among the objects and this permits them to contribute equally to these similarities. In practice, the use of second derivative spectra is essential, as minor changes across the specimen are difficult to observe from the raw spectra due to the overlap of the different constituent bands. Further, as the second derivative spectra increase the resolution, small or subtle features can also be identified. It is also known that the second derivative of a spectrum removes the contributions of offset and slope from the original spectrum, as well as reducing the contribution related to a slowly varying baseline in a spectrum [123]. Although the use of derivatives will not eliminate the scattering component from the spectrum, it does discriminate against the scattering

component and reduces its effect on quantification. However, derivation should be used with caution when spectra of insufficient quality are used, since it may increase the noise.

When spectra are ready, they are passed to the cluster analysis. It was discovered that FCM was the most optimal algorithm for discrimination of non-hierarchical biological data. However, based on the problem, one can choose another clustering algorithm. Using nonhierarchical methods one should set the optimal number of clusters to identify differences in the data. Alternatively, one can search in advance for the optimal number of clusters for the particular data set. For instance, this can be done with the unsupervised cluster validity measures MSE or other validity measures [80, 81].

Validation of the clustering results is crucial when the suitability of a technique is being evaluated, or different clustering techniques need to be compared [80]. Various numerical measures, or indices, have been developed for cluster validation. Those indices which use no external information, like class labels, are called unsupervised or internal indices, and those, which measure how well the clustering results match with some external information, are called supervised or external indices. Study I evaluated the percentage of the correct cluster assignments compared to known class labels. In study II, the supervised Rand index was applied to compare the performance of the clustering algorithms. In studies III and IV, clustering maps were visually compared to histological images and PLM images to qualitatively evaluate the agreement of clusters with tissue types or zones.

In this study, custom made Matlab scripts were developed for all analyses. However, one can also use commercial software, e.g., CytoSpec[2] or iSys[3] spectral imaging software, to conduct similar analysis. The built-in Matlab functions with small

[2] CytoSpec (2014) , retrieved from http://www.cytospec.com/
[3] Malvern Instruments Ltd (2014), http://www.malvern.com/en/

enhancement could also be used, as well as some commercial software packages such as "Essential FTIR"[4] which have possibilities to create add-on scripts on built-in functions.

8.6 POTENTIAL OF CLUSTER ANALYSIS OF FTIR-MSP IN DIAGNOSTICS

The FTIR-MSP is a potentially useful diagnostic tool for biomedicine [19, 53-55, 124, 125]. It can capture both the spatial distribution and the concentration of biological components in tissue and it provides a great amount of data for quantitative assessment of tissue composition. For example, the present studies revealed that the collagen content as estimated from FTIR-MSP spectra of SB is altered during aging, and this was consistent to changes detected by traditional biochemical methods [32]. The estimated composition of clustering-derived repaired regions of AC corresponded well to the histological appearance of the samples, i.e. to differences in PG and collagen I staining.

In reality, spatial analysis can show only variations in certain parameters, e.g., collagen or PG composition across the tissue. However, to enable optimal comparisons and analyses, one must know what to particularly seek in the FTIR image. Cluster analysis determines region of similar spectra without human intervention, and discrimination of every pixel is directly seen from the clustering maps. Moreover, probabilities for assignment of each image pixel to the clusters can be observed. Then, one can undertake an inspection of spatial variations in tissue properties across the sample. The mean spectra of clusters can be used to allocate spectral differences or to evaluate tissue composition within each cluster. The clustering maps could be visually compared to histological images or to other image modalities (PLM, X-rays, etc) in order to match the clustering

[4] Essential FTIR (2014), http://www.essentialftir.com/

results with histopathological evidence. One can even visualize 3D images of tissue sections as was conducted by Wood et al. using HCA [126].

Since cluster analysis requires no a priori knowledge it can reveal both novel and as well as expected features in the data. Thus, in study III, the observed structure of histological zones in intact AC did not exactly match the known zonal structure of AC, and differed from the PLM-derived structure. However, it should be acknowledged that clustering can also capture other properties in addition to variations in collagen fibril orientation or relative composition of the matrix. FTIR-MSP can therefore provide complementary information on the different histological zones in intact AC. The "hard" clustering methods could not identify these features. One could also carefully investigate transitions between zones by studying membership maps.

In study IV, differences were observed between the histological and the clustering results. This inconsistency suggests that cluster analysis utilizes some other features for discrimination between tissues. Moreover, as two types of repaired AC could be discriminated, the non-optimal repair outcome was obvious. The less accurate discrimination of collagen gel repaired AC might indicate better overall integrity of the repair tissue or/and more variation in the healing response. On the other hand, the better discrimination of the spontaneously repaired tissue suggests that the overall quality of generated cartilage is rather different from the intact tissue.

Although some inconsistencies exist in interpretation of the present clustering results, the methodology for the FTIR data analysis is presented to discriminate accurately biological tissues without human interaction. This study is a step along the road towards further development of medical FTIR applications, primarily for research and use in the diagnostics of bone and AC.

9 Summary and conclusions

This is the first study to assess the feasibility of exploiting cluster analysis to discriminate bone of different ages, normal and repaired articular cartilage and histological zones in intact cartilage measured with FTIR-MSP.

The main results and conclusions of this thesis can be summarized as follows:

- Cluster analysis could discriminate tissues based on the subtle differences in the FTIR-MSP spectra. These differences could be enhanced by calculation of second derivatives. The results of classification were comparable to histological inspection and FTIR-based quantitative evaluation.
- Fuzzy clustering produced more accurate results than either the "hard" or hierarchical clustering methods. It makes multiple assignments and thus, allows investigation of transitions and uncertainties in its classification.
- Certain difficulties exist in the interpretation of spectral differences between detected classes due to the complexity of FTIR-MSP spectra.
- Cluster analysis is non-subjective, rapid and easy to implement. Potentially, it could be applied for use in real time clinical analysis.
- An algorithm for clustering of FTIR-MSP data was thoroughly investigated and programmed using Matlab for future reference.

In summary, fuzzy cluster analysis of FTIR spectra of bone and cartilage represents an effective tool for tissue analysis, and could be potentially useful in the diagnostics of tissue

pathologies. However, more research will be needed to link the specific changes of the spectrum with the corresponding biological and molecular changes.

In principle, the present clustering approaches may be used also for diagnostics of other tissues. All cluster analyses need good quality FTIR-MSP spectra, containing signatures of the tissue of interest. However, one will need to consider adjustments of the devised algorithm in order to apply methodology for other research subjects.

References

[1] *E. P. Paschalis, R. Mendelsohn, and A. L. Boskey, "Infrared assessment of bone quality: a review." Clin Orthop Relat Res, 469(8), 2170-2178 (2011).*

[2] *M. Huber, S. Trattnig, and F. Lintner, "Anatomy, biochemistry, and physiology of articular cartilage." Invest Radiol, 35(10), 573-580 (2000).*

[3] *J. A. Buckwalter, M. J. Glimcher, R. R. Cooper, and R. Recker, "Bone biology. I: Structure, blood supply, cells, matrix, and mineralization." Instr Course Lect, 45, 371-386 (1996).*

[4] *J. A. Buckwalter, M. J. Glimcher, R. R. Cooper, and R. Recker, "Bone biology. II: Formation, form, modeling, remodeling, and regulation of cell function." Instr Course Lect, 45, 387-399 (1996).*

[5] *V. C. Mow, A. Ratcliffe, and S. L. Woo, Biomechanics of Diarthrodial Joints. Vol. I, New York: Springer-Verlag (1990).*

[6] *D. W. Jackson, P. A. Lalor, H. M. Aberman, and T. M. Simon, "Spontaneous repair of full-thickness defects of articular cartilage in a goat model. A preliminary study." J Bone Joint Surg Am, 83-A(1), 53-64 (2001).*

[7] *A. P. Newman, "Articular cartilage repair." Am J Sports Med, 26(2), 309-324 (1998).*

[8] *D. B. Burr, "The contribution of the organic matrix to bone's material properties." Bone, 31(1), 8-11 (2002).*

[9] *H. Isaksson, M. J. Turunen, L. Rieppo, S. Saarakkala, I. S. Tamminen, J. Rieppo, H. Kroger, and J. S. Jurvelin, "Infrared spectroscopy indicates altered bone turnover and remodeling activity in renal osteodystrophy." J Bone Miner Res, 25(6), 1360-1366 (2010).*

[10] *M. Kim, L. F. Foo, C. Uggen, S. Lyman, J. T. Ryaby, D. P. Moynihan, D. A. Grande, H. G. Potter, and N. Pleshko, "Evaluation of early osteochondral defect repair in a rabbit model utilizing fourier transform-infrared imaging spectroscopy, magnetic resonance imaging, and quantitative T2 mapping." Tissue Eng Part C Methods, 16(3), 355-364 (2010).*

[11] *H. J. Nieminen, J. Toyras, J. Rieppo, M. T. Nieminen, J. Hirvonen, R. Korhonen, and J. S. Jurvelin, "Real-time ultrasound analysis of articular cartilage degradation in vitro." Ultrasound Med Biol, 28(4), 519-525 (2002).*

[12] *M. T. Nieminen, J. Toyras, J. Rieppo, J. M. Hakumaki, J. Silvennoinen, H. J. Helminen, and J. S. Jurvelin, "Quantitative MR microscopy of enzymatically degraded articular cartilage." Magn Reson Med, 43(5), 676-681 (2000).*

[13] *T. Viren, S. Saarakkala, J. S. Jurvelin, H. J. Pulkkinen, V. Tiitu, P. Valonen, I. Kiviranta, M. J. Lammi, and J. Toyras, "Quantitative evaluation of spontaneously and surgically repaired rabbit articular cartilage using intra-articular ultrasound method in situ." Ultrasound Med Biol, 36(5), 833-839 (2010).*

[14] *A. Boskey and R. Mendelsohn, "Infrared analysis of bone in health and disease." J Biomed Opt, 10(3), 031102 (2005).*

[15] *N. P. Camacho, P. West, P. A. Torzilli, and R. Mendelsohn, "FTIR microscopic imaging of collagen and proteoglycan in bovine cartilage." Biopolymers, 62(1), 1-8 (2001).*

[16] *E. P. Paschalis, "Fourier transform infrared analysis and bone." Osteoporosis International, 20(6), 1043-1047 (2009).*

[17] *K. Potter, L. H. Kidder, I. W. Levin, E. N. Lewis, and R. G. Spencer, "Imaging of collagen and proteoglycan in cartilage sections using Fourier transform infrared spectral imaging." Arthritis Rheum, 44(4), 846-855 (2001).*

[18] *N. Ramakrishnan, Y. Xia, and A. Bidthanapally, "Polarized IR microscopic imaging of articular cartilage." Phys Med Biol, 52(15), 4601-4614 (2007).*

[19] *C. Krafft and V. Sergo, "Biomedical applications of Raman and infrared spectroscopy to diagnose tissues." Spectroscopy, 20(5-6), 195-218 (2006).*

[20] *S. Saarakkala, L. Rieppo, J. Rieppo, and J. Jurvelin, "Fourier Transform Infrared (FTIR) Microspectroscopy of Immature, Mature and Degenerated Articular Cartilage. ." Microscopy Science Technology and Education, 1, 403-414 (2010).*

[21] *L. Wang and B. Mizaikoff, "Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy." Anal Bioanal Chem, 391(5), 1641-1654 (2008).*

[22] *A. Boskey and N. Pleshko Camacho, "FT-IR imaging of native and tissue-engineered bone and cartilage." Biomaterials, 28(15), 2465-2478 (2007).*

[23] *A. Kohler, D. Bertrand, H. Martens, K. Hannesson, C. Kirschner, and R. Ofstad, "Multivariate image analysis of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information." Anal Bioanal Chem, 389(4), 1143-1153 (2007).*

[24] *C. R. Ethier and C. A. Simmons, Introductory Biomechanics - From Cells to Organisms, ed. M. Saltzman, New York: Cambridge University Press (2007).*

[25] *X. Wang, R. A. Bank, J. M. TeKoppele, and C. M. Agrawal, "The role of collagen in determining bone mechanical properties." J Orthop Res, 19(6), 1021-1026 (2001).*

[26] *A. J. Bailey, T. J. Sims, E. N. Ebbesen, J. P. Mansell, J. S. Thomsen, and L. Mosekilde, "Age-related changes in the biochemical properties of human cancellous bone collagen: relationship to bone strength." Calcif Tissue Int, 65(3), 203-210 (1999).*

[27] *S. K. Saladin, Anatomy and Physiology : The Unity of Form and Function. 5 ed: McGraw-Hill Science/Engineering/Math. 1400 (2006).*

[28] *L. T. Kuhn, M. D. Grynpas, C. C. Rey, Y. Wu, J. L. Ackerman, and M. J. Glimcher, "A comparison of the physical and chemical differences between cancellous and cortical bovine bone mineral at two ages." Calcif Tissue Int, 83(2), 146-154 (2008).*

[29] *J. P. Mansell and A. J. Bailey, "Abnormal cancellous bone collagen metabolism in osteoarthritis." J Clin Invest, 101(8), 1596-1603 (1998).*

[30] *J. T. Holopainen, P. A. Brama, E. Halmesmaki, T. Harjula, J. Tuukkanen, P. R. van Weeren, H. J. Helminen, and M. M. Hyttinen, "Changes in subchondral bone mineral density and collagen matrix organization in growing horses." Bone, 43(6), 1108-1114 (2008).*

[31] *J. M. Very, R. Gibert, B. Guilhot, M. Debout, and C. Alexandre, "Effect of aging on the amide group of bone matrix, measured by FTIR spectrophotometry, in adult subjects deceased as a result of violent death." Calcif Tissue Int, 60(3), 271-275 (1997).*

[32] *P. A. Brama, J. M. TeKoppele, R. A. Bank, A. Barneveld, and P. R. van Weeren, "Biochemical development of subchondral bone from birth*

until age eleven months and the influence of physical activity." Equine Vet J, 34(2), 143-149 (2002).

[33] H. Isaksson, T. Harjula, A. Koistinen, J. Iivarinen, K. Seppanen, J. P. Arokoski, P. A. Brama, J. S. Jurvelin, and H. J. Helminen, "Collagen and mineral deposition in rabbit cortical bone during maturation and growth: effects on tissue properties." J Orthop Res, 28(12), 1626-1633 (2010).

[34] M. J. Turunen, S. Saarakkala, H. J. Helminen, J. S. Jurvelin, and H. Isaksson, "Age-related changes in organization and content of the collagen matrix in rabbit cortical bone." J Orthop Res, 30(3), 435-442 (2012).

[35] H. Isaksson, V. Tolvanen, M. A. Finnila, J. Iivarinen, J. Tuukkanen, K. Seppanen, J. P. Arokoski, P. A. Brama, J. S. Jurvelin, and H. J. Helminen, "Physical exercise improves properties of bone and its collagen network in growing and maturing mice." Calcif Tissue Int, 85(3), 247-256 (2009).

[36] H. Chiang and C. C. Jiang, "Repair of articular cartilage defects: review and perspectives." J Formos Med Assoc, 108(2), 87-101 (2009).

[37] J. A. Buckwalter, V. C. Mow, and A. Ratcliffe, "Restoration of Injured or Degenerated Articular Cartilage." J Am Acad Orthop Surg, 2(4), 192-201 (1994).

[38] I. M. Khan, S. J. Gilbert, S. K. Singhrao, V. C. Duance, and C. W. Archer, "Cartilage integration: evaluation of the reasons for failure of integration during cartilage repair. A review." Eur Cell Mater, 16, 26-39 (2008).

[39] M. Falah, G. Nierenberg, M. Soudry, M. Hayden, and G. Volpin, "Treatment of articular cartilage lesions of the knee." Int Orthop, 34(5), 621-630 (2010).

[40] H. A. Breinan, T. Minas, H. P. Hsu, S. Nehrer, S. Shortkroff, and M. Spector, "Autologous chondrocyte implantation in a canine model: change in composition of reparative tissue with time." J Orthop Res, 19(3), 482-492 (2001).

[41] G. Knutsen, L. Engebretsen, T. C. Ludvigsen, J. O. Drogset, T. Grontvedt, E. Solheim, T. Strand, S. Roberts, V. Isaksen, and O. Johansen, "Autologous chondrocyte implantation compared with microfracture in the knee. A randomized trial." J Bone Joint Surg Am, 86-A(3), 455-464 (2004).

[42] *F. Shapiro, S. Koide, and M. J. Glimcher, "Cell origin and differentiation in the repair of full-thickness defects of articular cartilage." J Bone Joint Surg Am, 75(4), 532-553 (1993).*

[43] *C. Hoemann, R. Kandel, S. Roberts, D. B. F. Saris, L. Creemers, P. Mainil-Varlet, S. Méthot, A. P. Hollander, and M. D. Buschmann, "International Cartilage Repair Society (ICRS) Recommended Guidelines for Histological Endpoints for Cartilage Repair Studies in Animal Models and Clinical Trials." Cartilage, 2(2), 153-172 (2011).*

[44] *Y. Xia, N. Ramakrishnan, and A. Bidthanapally, "The depth-dependent anisotropy of articular cartilage by Fourier-transform infrared imaging (FTIRI)." Osteoarthritis Cartilage, 15(7), 780-788 (2007).*

[45] *G. H. Welsch, T. C. Mamisch, S. Quirbach, L. Zak, S. Marlovits, and S. Trattnig, "Evaluation and comparison of cartilage repair tissue of the patella and medial femoral condyle by using morphological MRI and biochemical zonal T2 mapping." Eur Radiol, 19(5), 1253-1262 (2009).*

[46] *C. W. Han, C. R. Chu, N. Adachi, A. Usas, F. H. Fu, J. Huard, and Y. Pan, "Analysis of rabbit articular cartilage repair after chondrocyte implantation using optical coherence tomography." Osteoarthritis Cartilage, 11(2), 111-121 (2003).*

[47] *G. Li, M. Thomson, E. Dicarlo, X. Yang, B. Nestor, M. P. Bostrom, and N. P. Camacho, "A chemometric analysis for evaluation of early-stage cartilage degradation by infrared fiber-optic probe spectroscopy." Appl Spectrosc, 59(12), 1527-1533 (2005).*

[48] *P. A. West, M. P. Bostrom, P. A. Torzilli, and N. P. Camacho, "Fourier transform infrared spectral analysis of degenerative cartilage: an infrared fiber optic probe and imaging study." Appl Spectrosc, 58(4), 376-381 (2004).*

[49] *X. Bi, X. Yang, M. P. Bostrom, D. Bartusik, S. Ramaswamy, K. W. Fishbein, R. G. Spencer, and N. P. Camacho, "Fourier transform infrared imaging and MR microscopy studies detect compositional and structural changes in cartilage in a rabbit model of osteoarthritis." Anal Bioanal Chem, 387(5), 1601-1612 (2007).*

[50] *X. Bi, X. Yang, M. P. Bostrom, and N. P. Camacho, "Fourier transform infrared imaging spectroscopy investigations in the pathogenesis and repair of cartilage." Biochim Biophys Acta, 1758(7), 934-941 (2006).*

[51] *X. Bi, G. Li, S. B. Doty, and N. P. Camacho, "A novel method for determination of collagen orientation in cartilage by Fourier transform infrared imaging spectroscopy (FT-IRIS)." Osteoarthritis Cartilage, 13(12), 1050-1058 (2005).*

[52] *B. H. Stuart, Infrared Spectroscopy: Fundamentals and Applications. Analytical Techniques in the Sciences (AnTS), ed. D. J. Ando: Published online. 224 (2005).*

[53] *S.-Y. Lin, M.-J. Li, and W.-T. Cheng, "FT-IR and Raman vibrational microspectroscopies used for spectral biodiagnosis of human tissues." Spectroscopy, 21(1), 1-30 (2007).*

[54] *Z. Movasaghi, S. Rehman, and D. I. ur Rehman, "Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues." Applied Spectroscopy Reviews, 43(2), 134-179 (2008).*

[55] *M. Beekes, P. Lasch, and D. Naumann, "Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research." Vet Microbiol, 123(4), 305-319 (2007).*

[56] *M. Jackson and H. H. Mantsch, "The Use and Misuse of FTIR Spectroscopy in the Determination of Protein Structure." Critical Reviews in Biochemistry and Molecular Biology, 30(2) (1995).*

[57] *R. Servaty, J. Schiller, H. Binder, and K. Arnold, "Hydration of polymeric components of cartilage--an infrared spectroscopic study on hyaluronic acid and chondroitin sulfate." Int J Biol Macromol, 28(2), 121-127 (2001).*

[58] *J. Kong and S. Yu, "Fourier transform infrared spectroscopic analysis of protein secondary structures." Acta Biochim Biophys Sin (Shanghai), 39(8), 549-559 (2007).*

[59] *E. P. Paschalis, K. Verdelis, S. B. Doty, A. L. Boskey, R. Mendelsohn, and M. Yamauchi, "Spectroscopic characterization of collagen cross-links in bone." J Bone Miner Res, 16(10), 1821-1828 (2001).*

[60] *J. Rieppo, L. Rieppo, S. Saarakkala, and J. S. Jurvelin, "Fourier transform infrared imaging spectroscopy in biomedicine - Important things to consider when planning a new experiment ", in Fourier Transforms - New analytical approaches and FTIR strategies N. Goran, Editor, In Tech p. 3-14 (2011).*

[61] *C. H. Petter, N. Heigl, M. Rainer, R. Bakry, J. Pallua, G. K. Bonn, and C. W. Huck, "Development and application of Fourier-transform infrared chemical imaging of tumour in human tissue." Curr Med Chem, 16(3), 318-326 (2009).*

[62] *S. Aparicio, S. B. Doty, N. P. Camacho, E. P. Paschalis, L. Spevak, R. Mendelsohn, and A. L. Boskey, "Optimal methods for processing mineralized tissues for Fourier transform infrared microspectroscopy." Calcif Tissue Int, 70(5), 422-429 (2002).*

[63] *E. Donnelly, "Methods for assessing bone quality: a review." Clin Orthop Relat Res, 469(8), 2128-2138 (2011).*

[64] *C. Krafft, G. Steiner, C. Beleites, and R. Salzer, "Disease recognition by infrared and Raman spectroscopy." J Biophotonics, 2(1-2), 13-28 (2009).*

[65] *A. L. Boskey, E. DiCarlo, E. Paschalis, P. West, and R. Mendelsohn, "Comparison of mineral quality and quantity in iliac crest biopsies from high- and low-turnover osteoporosis: an FT-IR microspectroscopic investigation." Osteoporos Int, 16(12), 2031-2038 (2005).*

[66] *Y. Kobrina, H. Isaksson, M. Sinisaari, L. Rieppo, P. A. Brama, R. van Weeren, H. J. Helminen, J. S. Jurvelin, and S. Saarakkala, "Infrared spectroscopy reveals both qualitative and quantitative differences in equine subchondral bone during maturation." J Biomed Opt, 15(6), 067003 (2010).*

[67] *E. P. Paschalis, R. Recker, E. DiCarlo, S. B. Doty, E. Atti, and A. L. Boskey, "Distribution of collagen cross-links in normal human trabecular bone." J Bone Miner Res, 18(11), 1942-1946 (2003).*

[68] *M. J. Turunen, S. Saarakkala, L. Rieppo, H. J. Helminen, J. S. Jurvelin, and H. Isaksson. A comparison between Infrared and Raman spectroscopic parameters on maturing rabbit bone. in Annual meeting of the Orthopaedic Reserach Society. 2010. New Orleans: ORS.*

[69] *J. Rieppo, M. M. Hyttinen, E. Halmesmaki, H. Ruotsalainen, A. Vasara, I. Kiviranta, J. S. Jurvelin, and H. J. Helminen, "Changes in spatial collagen content and collagen network architecture in porcine articular cartilage during growth and maturation." Osteoarthritis Cartilage, 17(4), 448-455 (2009).*

[70] *J. Rieppo, J. Toyras, M. T. Nieminen, V. Kovanen, M. M. Hyttinen, R. K. Korhonen, J. S. Jurvelin, and H. J. Helminen, "Structure-function relationships in enzymatically modified articular cartilage." Cells Tissues Organs, 175(3), 121-132 (2003).*

[71] *L. Rieppo, S. Saarakkala, T. Narhi, J. Holopainen, M. Lammi, H. J. Helminen, J. S. Jurvelin, and J. Rieppo, "Quantitative analysis of spatial proteoglycan content in articular cartilage with Fourier*

transform infrared imaging spectroscopy: Critical evaluation of analysis methods and specificity of the parameters." Microsc Res Tech, 73(5), 503-512 (2010).

[72] J. Yin and Y. Xia, "Macromolecular concentrations in bovine nasal cartilage by Fourier transform infrared imaging and principal component regression." Applied Spectroscopy, 64(11), 1199-1208 (2010).

[73] E. David-Vaudey, A. Burghardt, K. Keshari, A. Brouchet, M. Ries, and S. Majumdar, "Fourier Transform Infrared Imaging of focal lesions in human osteoarthritic cartilage." Eur Cell Mater, 10, 51-60; discussion 60 (2005).

[74] Y. Xia, "Averaged and depth-dependent anisotropy of articular cartilage by microscopic imaging." Semin Arthritis Rheum, 37(5), 317-327 (2008).

[75] N. Ramakrishnan, Y. Xia, and A. Bidthanapally, "Fourier-transform infrared anisotropy in cross and parallel sections of tendon and articular cartilage." J Orthop Surg Res, 3, 48 (2008).

[76] N. P. Camacho, P. West, M. H. Griffith, R. F. Warren, and C. Hidaka, "FT-IR imaging spectroscopy of genetically modified bovine chondrocytes." Materials Science and Engineering: C, 17(1-2), 3-9 (2001).

[77] C. Krafft, L. Shapoval, S. B. Sobottka, G. Schackert, and R. Salzer, "Identification of primary tumors of brain metastases by infrared spectroscopic imaging and linear discriminant analysis." Technol Cancer Res Treat, 5(3), 291-298 (2006).

[78] C. Krafft, S. B. Sobottka, K. D. Geiger, G. Schackert, and R. Salzer, "Classification of malignant gliomas by infrared spectroscopic imaging and linear discriminant analysis." Anal Bioanal Chem, 387(5), 1669-1677 (2007).

[79] 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." Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy, 88, 90-96 (2012).

[80] A. K. Jain, M. N. Murty, and Flynn, "Data Clustering: A Review." ACM Computing Surveys, 31(3) (1999).

[81] L. Rokach, "A survey of Clustering Algorithms", in Data Mining and Knowledge Discovery Handbook, L. Rokach and O. Maimon, Editors, Springer Science+Business Media, LLC (2010).

[82] *B. Everitt, Cluster analysis. 5th ed. Wiley series in probability and statistics, Chichester, West Sussex, U.K.: Wiley. xii, 330 p. (2011).*

[83] *P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to data mining. 1st ed, Boston: Pearson Addison Wesley. xxi, 769 p. (2006).*

[84] *J. H. J. Ward, "Hierarchical Grouping to Optimize an Objective Function." Journal of the American Statistical Association, 58(301), 235-244 (1963).*

[85] *R. B. Burns and R. A. Burns, Business research methods and statistics using SPSS, Los Angeles ; London: SAGE. xii, 544 p. (2008).*

[86] *R. Xu and D. Wunsch, 2nd, "Survey of clustering algorithms." IEEE Trans Neural Netw, 16(3), 645-678 (2005).*

[87] *W. M. Rand, "Objective criteria for the evaluation of clustering methods." Journal of the American Statistical Association, 66, 846-850 (1971).*

[88] *P. Yu, "Application of advanced synchrotron radiation-based Fourier transform infrared (SR-FTIR) microspectroscopy to animal nutrition and feed science: a novel approach." Br J Nutr, 92(6), 869-885 (2004).*

[89] *B. Dzuiba, A. Babuchowki, D. Nalecz, and M. Niklewicz, "Identification of lactic acid bacteria using FTIR spectroscopy and cluster analysis." International Dairy Journal, 17(3), 183-189 (2007).*

[90] *A. Oust, T. Moretro, C. Kirschner, J. A. Narvhus, and A. Kohler, "FT-IR spectroscopy for identification of closely related lactobacilli." J Microbiol Methods, 59(2), 149-162 (2004).*

[91] *C. Krafft, M. Kirsch, C. Beleites, G. Schackert, and R. Salzer, "Methodology for fiber-optic Raman mapping and FTIR imaging of metastases in mouse brains." Anal Bioanal Chem, 389(4), 1133-1142 (2007).*

[92] *E. Ly, O. Piot, R. Wolthuis, A. Durlach, P. Bernard, and M. Manfait, "Combination of FTIR spectral imaging and chemometrics for tumour detection from paraffin-embedded biopsies." Analyst, 133(2), 197-205 (2008).*

[93] *L. Zhang, G. W. Small, A. S. Haka, L. H. Kidder, and E. N. Lewis, "Classification of Fourier transform infrared microscopic imaging data of human breast cells by cluster analysis and artificial neural networks." Appl Spectrosc, 57(1), 14-22 (2003).*

[94] *J. Mordehai, J. Ramesh, M. Huleihel, Z. Cohen, O. Kleiner, M. Talyshinsky, V. Erukhimovitch, A. Cahana, A. Salman, R. K. Sahu, H. Guterman, and S. Mordechai, "Studies on acute human infections*

using FTIR microspectroscopy and cluster analysis." Biopolymers, 73(4), 494-502 (2004).

[95] P. C. Lin, D. A. Reiter, and R. G. Spencer, "Classification of degraded cartilage through multiparametric MRI analysis." J Magn Reson, 201(1), 61-71 (2009).

[96] A. M. Croxford, D. Crombie, D. McNaughton, R. Holmdahl, K. S. Nandakumar, and M. J. Rowley, "Specific antibody protection of the extracellular cartilage matrix against collagen antibody-induced damage." Arthritis Rheum, 62(11), 3374-3384 (2010).

[97] A. Bonifacio, C. Beleites, F. Vittur, E. Marsich, S. Semeraro, S. Paoletti, and V. Sergo, "Chemical imaging of articular cartilage sections with Raman mapping, employing uni- and multi-variate methods for data analysis." Analyst, 135(12), 3193-3204 (2010).

[98] J. Rieppo, J. Hallikainen, J. S. Jurvelin, I. Kiviranta, H. J. Helminen, and M. M. Hyttinen, "Practical considerations in the use of polarized light microscopy in the analysis of the collagen network in articular cartilage." Microsc Res Tech, 71(4), 279-287 (2008).

[99] H. J. Pulkkinen, V. Tiitu, P. Valonen, T. Silvast, J. Jurvelin, J. Toyras, M. Lammi, and I. Kiviranta, "Repair of osteochondral defects with recombinant human type II collagen gel and autologous chondrocytes in rabbit." Osteoarthritis Cartilage, 21(3), 481-490 (2013).

[100] P. A. Brama, D. Karssenberg, A. Barneveld, and P. R. van Weeren, "Contact areas and pressure distribution on the proximal articular surface of the proximal phalanx under sagittal plane loading." Equine Vet J, 33(1), 26-32 (2001).

[101] H. J. Pulkkinen, V. Tiitu, P. Valonen, E. R. Hamalainen, M. J. Lammi, and I. Kiviranta, "Recombinant human type II collagen as a material for cartilage tissue engineering." Int J Artif Organs, 31(11), 960-969 (2008).

[102] J. Rieppo, M. M. Hyttinen, J. S. Jurvelin, and H. J. Helminen, "Reference sample method reduces the error caused by variable cryosection thickness in Fourier transform infrared imaging." Appl Spectrosc, 58(1), 137-140 (2004).

[103] P. Bassan, A. Kohler, H. Martens, J. Lee, H. J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke, and P. Gardner, "Resonant Mie scattering (RMieS) correction of infrared spectra from highly scattering biological samples." Analyst, 135(2), 268-277 (2010).

[104] *S. W. Bruun, A. Kohler, I. Adt, G. D. Sockalingum, M. Manfait, and H. Martens, "Correcting attenuated total reflection-Fourier transform infrared spectra for water vapor and carbon dioxide." Appl Spectrosc, 60(9), 1029-1039 (2006).*

[105] *A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." Anal. Chem., 36(8), 1627-1639 (1964).*

[106] *R. Gnanadesikan, "Discriminant Analysis and Clustering." Panel on Discriminant Analysis, Classification, and Clustering, Committee on Applied and Theoretical Statistics Board on Mathematical Sciences, Commission on Physical Sciences, Mathematics, and Resources, National Research Council, N. A. PRESS, Washington, D.C., 105 (1988).*

[107] *J. K. Galbraith and J. Lu, "Cluster and Discriminant Analysis on Time-Series as a Research Tool." UTIP Working Paper No. 6., (1999).*

[108] *A. Tfayli, O. Piot, A. Durlach, P. Bernard, and M. Manfait, "Discriminating nevus and melanoma on paraffin-embedded skin biopsies using FTIR microspectroscopy." Biochim Biophys Acta, 1724(3), 262-269 (2005).*

[109] *K. Kiraly, T. Lapvetelainen, J. Arokoski, K. Torronen, L. Modis, I. Kiviranta, and H. J. Helminen, "Application of selected cationic dyes for the semiquantitative estimation of glycosaminoglycans in histological sections of articular cartilage by microspectrophotometry." Histochem J, 28(8), 577-590 (1996).*

[110] *J. Rieppo, L. Rieppo, M. M. Hyttinen, J. Jurvelin, I. Kiviranta, and H. J. Helminen. Characterization of Articular Cartilage Repair Tissue with Fourier Transform Infrared Imaging Spectroscopy and Cluster Analysis. in 53rd Annual Meeting of the Orthopaedic Research Society. 2007. San Diego, CA*

[111] *C. Chadefaux, A.-S. L. Hô, B. Bellot-Gurlet, and I. Reiche, "Curve fitting Micro-ATR-FTIR studies of the amide I and II bands of type I collagen in archaeological bone materials." e-Preservation Science, 6, 129-137 (2009).*

[112] *M. T. Nieminen, J. Rieppo, J. Toyras, J. M. Hakumaki, J. Silvennoinen, M. M. Hyttinen, H. J. Helminen, and J. S. Jurvelin, "T2 relaxation reveals spatial collagen architecture in articular cartilage: a comparative quantitative MRI and polarized light microscopic study." Magn Reson Med, 46(3), 487-493 (2001).*

[113] *M. J. Nissi, J. Rieppo, J. Toyras, M. S. Laasanen, I. Kiviranta, J. S. Jurvelin, and M. T. Nieminen, "T(2) relaxation time mapping reveals age- and species-related diversity of collagen network architecture in articular cartilage." Osteoarthritis Cartilage, 14(12), 1265-1271 (2006).*

[114] *Y. Xia, "Heterogeneity of cartilage laminae in MR imaging." J Magn Reson Imaging, 11(6), 686-693 (2000).*

[115] *A. J. Owen, "Uses of Derivative Spectroscopy." (1995).*

[116] *C. Petibois and B. Desbat, "Clinical application of FTIR imaging: new reasons for hope." Trends Biotechnol, 28(10), 495-500 (2010).*

[117] *G. Chen, S. Jaradat, N. Banerjee, T. Tanaka, M. Ko, and M. Zhang, "Evaluation and comparison of clustering algorithms in analyzing ES cell gene expression data." Statistica Sinica, 12, 241-262 (2002).*

[118] *P. Lasch, W. Haensch, D. Naumann, and M. Diem, "Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis." Biochim Biophys Acta, 1688(2), 176-186 (2004).*

[119] *A. A. Osama, "Comparison between data clustering algorithms." The International Arab Journal of Information Technology, 5(3), 320-325 (2008).*

[120] *J. R. Mansfield, M. G. Sowa, G. B. Scarth, R. L. Somorjai, and H. H. Mantsch, "Analysis of Spectroscopic Imaging Data by Fuzzy C-Means Clustering." Analytical Chemistry, 69, 3370-3374 (1997).*

[121] *H. Isaksson, M. Malkiewicz, R. Nowak, H. J. Helminen, and J. S. Jurvelin, "Rabbit cortical bone tissue increases its elastic stiffness but becomes less viscoelastic with age." Bone, 47(6), 1030-1038 (2010).*

[122] *Severcan F. and H. P.I., Vibrational spectroscopy in diagnosis and screening. Advances in biomedical spectroscopy. Vol. 6, Washington, DC: IOS Press. 432 (2012).*

[123] *F. Holler, H. D. Burns, and J. B. Callis, "Direct use of second derivatives in curve-fitting procedures." Applied Spectroscopy, 43(5), 877-882 (1989).*

[124] *H. H. Mantsch, L. Choo-Smith, and R. A. Shaw, "Vibrational spectroscopy and medicine: an alliance in the making." Vibrational Spectroscopy, 30(1), 31-41 (2002).*

[125] *I. W. Levin and R. Bhargava, "Fourier transform infrared vibrational spectroscopic imaging: integrating microscopy and molecular recognition." Annu Rev Phys Chem, 56, 429-474 (2005).*

[126] *B. R. Wood, K. R. Bambery, C. J. Evans, M. A. Quinn, and D. McNaughton, "A three-dimensional multivariate image processing technique for the analysis of FTIR spectroscopic images of multiple tissue sections." BMC Med Imaging, 6, 12 (2006).*

Yevgeniya Kobrina

Infrared Microspectroscopic Cluster Analysis of Bone and Cartilage

The composition and structure of bone and articular cartilage (AC) change with age, and in response to exercise or pathology. This study employed Fourier-transform infrared microspectroscopy and multivariate cluster analysis to study these changes. Without human supervision, clustering discriminated between age groups of maturing bone, histological zones in intact AC, and normal and repaired AC with good accuracy. The developed algorithm for processing infrared spectra using cluster analysis will be useful in future research and may have a potential in diagnostics of musculoskeletal diseases.

UNIVERSITY OF EASTERN FINLAND

Publications of the University of Eastern Finland
Dissertations in Forestry and Natural Sciences

ISBN 978-952-61-1457-6