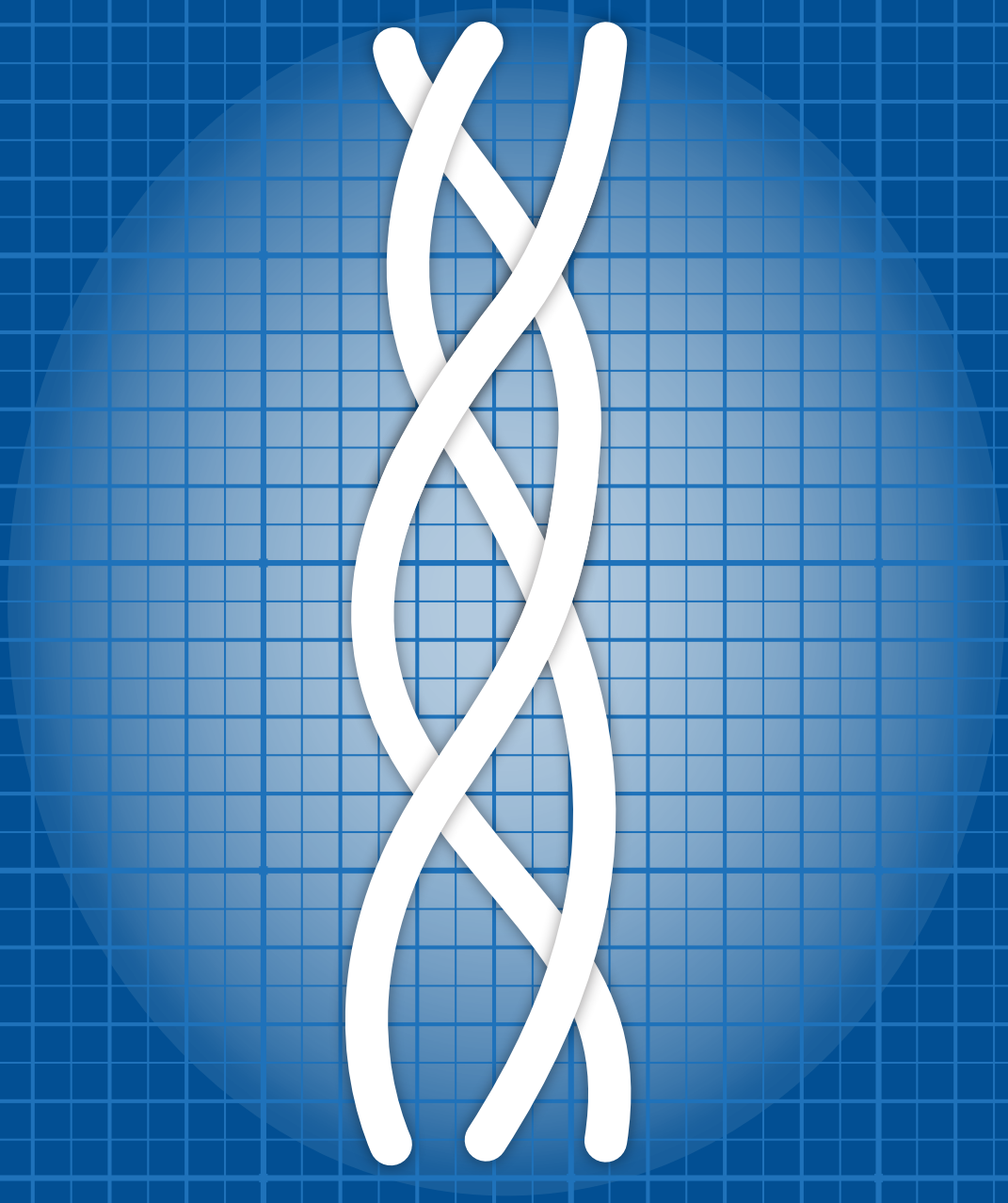
COLLAGEN IN COLORECTAL CANCER

– a mass spectrometry analysis –



Nick A. van Huizen

Collagen in Colorectal Cancer

– a mass spectrometry analysis –

Nick A. van Huizen

ISBN: 978-94-6323-939-4
Cover design: Ilse Modder, www.ilsemodder.nl
Layout: Ilse Modder, www.ilsemodder.nl
Printed by: Gildeprint Enschede, www.gildeprint.nl

MIX
Paper from responsible sources
FSC® C012854

Financial support for the printing of this thesis was kindly provided by:
Erasmus MC University medical center.

© Nick A. van Huizen, 2020.
For all articles published, the copyright has been transferred to the respective publisher. No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without written permission from the author or, when appropriate, from the publisher.

Collagen in Colorectal Cancer

– a mass spectrometry analysis –

Collageen in colorectale kanker

– een massaspectrometrie analyse –

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

Dinsdag 14 januari 2020 om 15:30 uur

door

Nick Arnold van Huizen
geboren te Rotterdam

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotoren

Prof.dr. J.N.M. IJzermans
Prof.dr. P.A.E. Sillevis Smitt

Overige Leden

Prof.dr. J.M. Kros
Prof.dr. R.A. Bank
Prof.dr. G.J.V.M. van Osch

Co-promotor

Dr. T.M. Luider

TABLE OF CONTENTS

Chapter 1	Introduction	9
Chapter 2	Collagen analysis with mass spectrometry	13
Chapter 3	Identification of a collagen marker in urine improves the detection of colorectal liver metastases	55
Chapter 4	Up-regulation of collagen proteins in colorectal liver metastasis compared with normal liver tissue	71
Chapter 5	Down-regulation of collagen hydroxylation in colorectal liver metastasis	89
Chapter 6	Identification of 4-hydroxyproline at the Xaa position in collagen by mass spectrometry	107
Chapter 7	General discussion	121
	Summary/Samenvatting	128
References		131
Appendices	Acknowledgements	153
	List of publications	156
	PhD Portfolio	158
	Biography	160

CHAPTER 1

Introduction

INTRODUCTION

Colorectal cancer (CRC) is the third most often diagnosed cancer in Europe, and is listed third of all cancer-related deaths.[1] People who suffer from CRC have a 20-40% chance of developing liver metastasis (colorectal liver metastasis; CRLM).[2-5] Both CRC and CRLM can be detected with (a combination of) various techniques: CT-scan, MRI, ultra sound, PET, fine-needle aspiration, serum carcinoembryonic antigen (CEA), colonoscopy, laparoscopy, stool analysis.[5-8] After detection of CRC and/or CRLM, the primary tumor will be surgically resected – if technically possible. After resection of the primary tumor, patients are offered an intense 5-year follow-up program consisting of multiple scans and CEA measurements. The 5-year survival rate is 40-50% [9, 10], and the 5-year disease-free survival rate is 20-30% [10].

A possible alternative to the above-mentioned techniques is a targeted mass spectrometry method analyzing natural occurring peptides (NOPs) of collagen in urine, described by Broker et al. and Lalmahomed et al.[11, 12] Urine was selected as a biofluid since large volumes can easily be collected non-invasively. Furthermore, it is much more convenient for a patient to hand-in urine, then to visit a hospital for a CT-scan. Analysis of a combination of collagen NOPs in urine and serum CEA resulted in a sensitivity of 85% and a specificity of 84% for the detection of colorectal liver metastases[12], which figures are comparable to those obtained with the currently used techniques [5, 13, 14], although not yet sufficient for clinical use. In this thesis we describe possibilities to find markers for metastasis of primary colorectal cancer.

The interest in this thesis is focused on the family of collagen proteins as potential markers for CRLM. Members of the collagen family have the ability to form a triple helix, and contain many posttranslational modifications. With the availability of state-of-the-art mass spectrometry collagen can be studied in relation to CRLM. An in-depth introduction to collagen and the analysis of collagen with mass spectrometry is provided in **Chapter 2**. The aim of this thesis is to improve the combination method of analyzing collagen NOPs in urine and serum CEA to detect CRLM by expansion of the collagen NOP panel. The discovery and validation of additional collagen NOPs in urine to improve the detection of CRLM is described in **Chapter 3**. It is difficult to prove that the collagen NOPs measured in urine directly originate from the CRLM. Therefore, we further investigated collagen in colon, CRC, liver, and CRLM tissues to gain better insight in collagen pathology. In **Chapter 4**, the upregulation, at the protein level, of collagen in CRLM and the similarity in expression levels to colon, CRC, and liver tissue has been described. **Chapter 5** describes the collagen hydroxylation pattern between colon, CRC, liver, and CRLM tissue. **Chapter 6** places a focus on the identification of an unknown proline hydroxylation. In the general discussion in **Chapter 7**, the results of the studies described in this thesis are discussed and summarized.

CHAPTER 2

Collagen analysis with mass spectrometry

Nick A. Van Huizen,[1,2] Jan N.M. IJzermans,[2] Peter C. Burgers,[1] and Theo M. Luider[1]*

[1] Department of Neurology, Erasmus Medical Center, [2] Department of Surgery, Erasmus University Medical Center, 3015 CN, Rotterdam, The Netherlands

Mass Spectrometry Reviews. 2019 Sep 9. [Epub ahead of print]

ABSTRACT

Mass spectrometry-based techniques can be applied to investigate collagen with respect to identification, quantification, supramolecular organization, and various post-translational modifications. The continuous interest in collagen research has led to a shift from techniques to analyze the physical characteristics of collagen to methods to study collagen abundance and modifications. In this review we illustrate the potential of mass spectrometry for in-depth analyses of collagen.

Reprinted with permission from Mass Spectrometry reviews. Copyright 2019 John Wiley and Sons.

I. INTRODUCTION

Proteins that belong to the collagen family are the most abundant proteins in the animal kingdom [15]. Approximately 30% of the protein content of the human body consists of collagen [15], and structures like bone (the organic part) and tendon might even consist of more than 90% of collagen [16]. Collagen has not been identified in only a few tissues, such as nail plates, hair shafts, and the eye lens [17-20].

So far, twenty-eight different collagen types have been identified – each with its specific characteristics. A collagen type consists of one to six of 45 different alpha chains, see below. In each collagen type, three alpha chains twisted around each other form a triple helix. On the basis of their supramolecular structure, eight subgroups of collagen types can be distinguished [21, 22]. Multiple triple helices form collagen supramolecular structures via covalent and non-covalent bonding, and these supramolecular structures can consist of multiple collagen types.

Collagen is a major component of the extracellular matrix (ECM). Frantz et al. have defined the ECM as: "The non-cellular component present within all tissues and organs, and provides not only essential physical scaffolding for the cellular constituents but also initiates crucial biochemical and biomechanical cues that are required for tissue morphogenesis differentiation and homeostasis" [23]. Among other functions, collagen is involved in handling physical stress and supporting tissue structures. Dysregulation of specific collagen types can have devastating effects. Mutations in collagen-related genes may cause various diseases, such as Alport syndrome, Bethlem syndrome, Ehlers-Danlos syndrome, and osteogenesis imperfecta (OI). A change in collagen turnover caused by external influences (e.g., scurvy or fibrosis) also results in a diseased state. Collagen also plays a role in tumor development and metastasis [24-26]. For the reader interested in the role of collagen in fibrosis we recommend the review of Karsdal et al. [27].

Although much is already known about collagen, a literature search will reveal a substantial lack of knowledge about, among other things, posttranslational modifications and collagen function. Mass spectrometry has great potential to contribute to our knowledge on collagen. To identify areas that lack knowledge, we searched for articles that match the search criteria "collagen" and "mass spectrometry" at "Web of Science", and have been cited over 30 times from 2010 up to 2019. In addition, we included older or newer articles or articles with fewer citations that contain valuable information about collagen. As a result, 41% of the references in this review date from 2010-2019, and 79% date from 2000-2019.

In section 1, we discuss collagen nomenclature and the production of collagen from DNA to the formation of supramolecular collagen structures. Section 2 describes the structural and functional characteristics of collagen as a background to this review, although most of the research referred to in this section was performed without mass spectrometry. Several gaps in collagen knowledge are pointed out and addressed. Mass spectrometry has the potential to fill up these gaps. Section III describes use of mass spectrometry in the analysis of collagen, and introduces circular dichroism and (immunohistochemistry) staining applied in collagen research.

I-A. COLLAGEN NOMENCLATURE

An overview of the nomenclature of the collagen triple helix is shown in figure 1. A collagen triple helix consists of three alpha chains, which after folding form procollagen. A collagen molecule, containing the triple helix, is formed by enzymatic cleavage of the terminal N- and terminal C-propeptides. If during the translation of an alpha chain the addition of post-translational modifications (PTMs) is interrupted, then the collagen molecule formed would be called a protocollagen. PTMs are modifications of the protein that are not encoded in the DNA and can alter the protein function. In literature, the plural "collagens" might be used; however we recommend using "collagen" because it is a collective noun.

In addition, different nomenclature styles for collagen types and alpha chains exist. The specific collagen type is annotated, for instance, as type 3, or (III), or III. The numbering follows the order of discovery [21, 22]. The specific alpha chain is represented by addition of an alpha numbering; for instance, alpha-1(III), or α1(III). Only the abbreviations for the collagen genes are used consistently in literature, and are also often used as the protein name abbreviation. The gene abbreviation is, for example, COL3A1, which stands for the protein collagen alpha-1(III). The stoichiometry of a collagen triple helix is annotated as $[\alpha 1(I)]_2\ \alpha 2(I)$ – a collagen type I that consists of two alpha-1 chains and one alpha-2 chain.

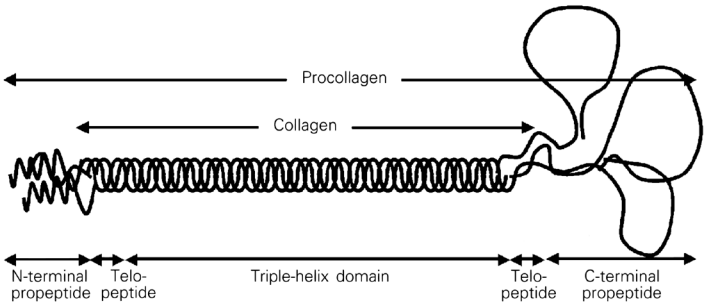


Figure 1. Nomenclature of the primary and secondary structures of the collagen triple helix.[28] Reprinted with permission from the author.

I-B. FROM DNA TO SUPRAMOLECULAR STRUCTURES

A schematic overview of collagen and fiber formation is shown in figure 2. The actual picture can be slightly different for the various collagen types and the supramolecular structures formed. Like any other protein, the primary amino acid structure of collagen is defined by DNA. Collagen genes are transcribed from DNA into RNA in the cell core and translated into a protein on ribosomes mostly present in the endoplasmic reticulum [29]. The protein formed is the building block of collagen and is called an alpha chain. Alpha-chain formation and the production of a complete procollagen molecule takes place in approximately 6 min (translational speed of ≈209 residues per min) [30].

Collagen contains various PTMs, such as hydroxylation of proline and lysine, oxidation of lysine, glycosylation of hydroxylated lysine, and cross-linking of oxidized lysine. Hydroxylation of proline and lysine, and glycosylation already start during the translation of the alpha chain and finish shortly after translation.

After translation and the addition of PTMs, three alpha chains of the same collagen type are linked together by sulfur bridges formed between specific cysteine's. Correct alignment of alpha chains is crucial for correct folding of a triple helix. The triple helix-forming regions in collagen contain the repetition of three amino acids (Gly-Xaa-Yaa; also written as Xaa-Yaa-Gly) in the primary amino acid sequence. The following pattern: Gly-Xaa-Yaa-$[\text{Gly-Xaa-Yaa}]_n$-Gly-Xaa-Yaa is present in the triple helical forming regions. In this pattern, glycine (Gly) points to the inside of the triple-helix. The triple helix is formed by a zipper-like folding mechanism. The rate-limiting step of triple helix folding is the requirement of all proline moieties to be in a trans conformation. In most amino acids, the trans conformation is energetically the more stable confirmation; however, proline contains a ring that allows the cis- and trans-conformation [31]. Proline is actively transformed from the cis- to the trans- conformation by peptidyl-prolyl cis-trans isomerase (PPIase) [32]. Triple-helical folding is sensitive to mutations of the primary alpha chain that inhibit further folding. Amino acid substitution of especially glycine inhibits helix formation due to steric hindrance. The rate-limiting step of collagen folding is the transformation of cis-proline into trans-proline [32].

The protein formed after triple helix-folding is called procollagen. The triple-helical region of procollagen is twisted in a right-handed way (clockwise); the alpha chains themselves are twisted in a left-handed way (counterclockwise).

All processes described so far take place inside a cell; at this stage, however, collagen is excreted into the ECM via the Golgi apparatus [33, 34]. During this event, triple-helix unwinding is prevented by so-called chaperone proteins (e.g., heat shock protein 47), which stabilize the

triple helix [35]. Furthermore, C- and N-propeptides are both cleaved off by enzymatic activity. The removal of the C- and N-propeptides stimulates the self-assembly of supramolecular structures [36]. The collagen supramolecular structures [29] are further stabilized by formation of covalent cross-linking and non-covalent bindings. Covalent cross-linking of collagen is mainly initiated by oxidation of lysine by lysyl oxidase. The oxidation of lysine occurs in the ECM [21, 37].

The produced collagen triple helix is built into large fibers. The type of fiber formed depends on the collagen types involved [21]. However, collagen types of similar composition can form different fiber types. For example, in the skin, collagen is mainly located in the dermis. The papillary layer contains collagen fibers of 0.3-3.0 µm thick, which are made of much thinner fibers with a diameter of 20-40 nm. These fibers form a loose network with no particular orientation [38]. The reticular layer is located below the papillary layer, which contains fibers with diameters between 10 and 40 µm made of collagen type I. These fibers form more compact and better arranged structures [38]. [38] obtained several images made by scanning electron microscopy that illustrate the above described networks in the skin and also show how a blood vessel is connected to the surrounding network by collagen fibers. Tateya et al. identified collagen type I in the vocal folds with immuno-scanning electron microscopy [39]. This technique enables to distinguish collagen type I and type III. Thus, Kadler et al. visualized collagen fibers in tissue, and produced a movie of consecutive tissue sections showing how the collagen fiber 'moves' through tissue [40].

Not only the fiber thickness varies, but also the ratio between, for example, collagen type I and II in skin. The latter ratio decreases during aging [41]. Besides the heterotypic fiber in the skin consisting of collagen type I and III, other heterotypical fibers exist. The 10+4 fiber is a supramolecular organization whereby four microfibrils are surrounded by a ring of ten microfibrils. The four core fibrils consist of two collagen type II and two collagen type XI microfibrils. Individual collagen type IX helices are assumed to be present at the surface of collagen type XI at the N-terminus, preventing further addition of microfibrils and thereby regulating fiber thickness [42]. An exception is the so-called super-twisted microfiber, which is formed of five collagen type I triple helices twisted around each other. These super-twisted microfibers interact with each other, thereby forming a hexagonal fiber purely made of collagen type I [43]. Fibers in cartilage are made of collagen types I and III, or collagen types II, IX, and XI, or collagen types II and III [44].

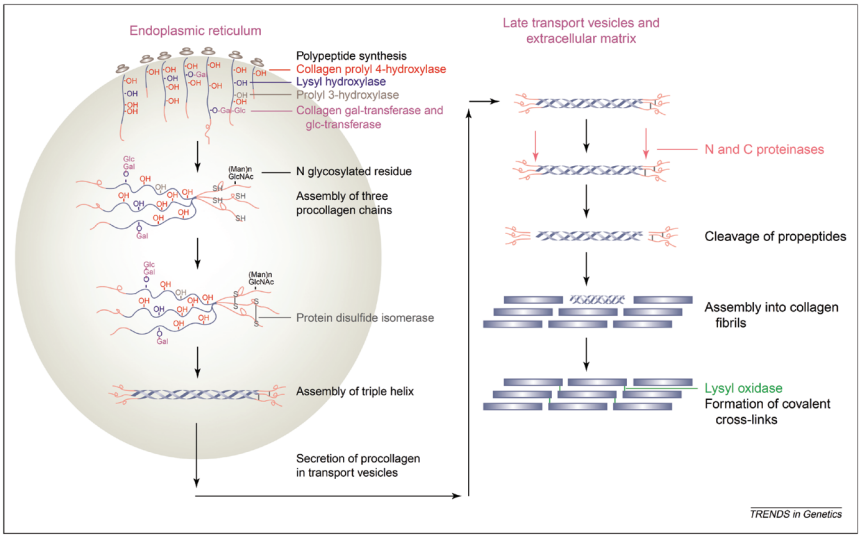


Figure 2. Schematic overview of collagen type I fiber formation from translation to fibers.[21] Reprinted with permission from the authors, 2004 Elsevier.

II. STRUCTURAL AND FUNCTIONAL CHARACTERISTICS OF COLLAGEN

II-A. PRIMARY STRUCTURE

II-A-1. Amino acid sequence

The most important property of collagen is its ability to form a triple helix. The ability to form a quaternary structure can be readily observed from the repetitive primary structure specific for a specific domain in the amino acid sequence, the helical domain. This domain consists of a tripeptide polymer with a distinct pattern, $[\text{Gly-Xaa-Yaa}]_n$. The amino acid glycine, which has the smallest side chain (-H), is located at the inside of the triple helix. A mutation of glycine into any other amino acid will result in steric hindrance – and consequently an imperfectly folded triple helix – or even inhibition of triple helix formation. [45-47]

In theory, four hundred different tripeptides patterns (Gly-Xaa-Yaa) can be expected just by chance (Xaa and Yaa are variable amino acids that yield a random variation of 20 times 20). In reality, not all amino acids are present at the Xaa or Yaa position, probably due to steric hindrance and helix stability. Ramshaw et al. have provided an overview of the appearance frequency given by for a limited number (n=14) of alpha chains from 12 collagen types [48]. The 10 most frequent tripeptides make up 39% of the helical domain, of which 10.5 % is GPP. Proline itself is in 55.7 % of cases located at the position Xaa, Yaa, or both. Other amino acids such as

cysteine, phenylalanine, glycine, histidine, tryptophan, and tyrosine are hardly present at the Xaa or Yaa position. Also, some amino acids are more often found at the Xaa than at the Yaa position or vice versa, such as lysine and leucine [49]. These amino acids might be present due to effects on the melting temperature (Tm); the Tm will be discussed in more detail in II-A-2.

The helical domain is not the only functional group present in the primary structure. At the N- and C-terminus of the triple helix, telopeptides and propeptides are located. At the N-terminus, a propeptide is located that serves as a signal peptide. Signal peptides "tell" the cell the destination of a protein. Furthermore, propeptides and telopeptides are involved in collagen cross-linking and the alignment of three alpha chains for triple helix formation.

Furthermore, in the primary structure of the non-helical regions of collagen, information can be present for angiogenesis inhibition. Three of the six collagen type IV alpha chains, collagen type XV, and collagen type XVIII contain angiogenesis inhibitor information (see table 1).

Table 1. Angiogenesis inhibitors present in the primary structure of the non-helical domains of collagen.

Collagen type	Angiogenesis inhibitor
COL4A1	Arresten [50, 51]
COL4A2	Canstatin [51-53]
COL4A3	Tumstatin [54]
COL15A1	Restin (Endostatin-like) [51, 55]
COL18A1	Endostatin [56]

Mass spectrometry has played an important role in the analysis of collagen angiogenesis inhibitors. Angiogenesis in healthy tissue occurs only in tissue repair, physical exercise, and in the female reproduction cycle [51, 57]. However, angiogenesis is activated by several diseases, such as psoriasis, rheumatoid arthritis, diabetic retinopathy, and cancer [51]. Endostatin, which functions as an anti-angiogenic cytokine, was first detected in human blood with the use of mass spectrometry (MALDI-MS and ESI-MS) and N-terminal sequencing [58]. Standker et al. showed the presence of sulfur bridges between cysteine residues 1-3 and 2-4. Later analysis with mass spectrometry (MALDI-TOF and ESI-Ion trap) showed that cysteine residues 1-4 and 2-3 were connected in endogenous endostatin [59]. Furthermore, John et al. showed that recombinant endostatin, used as drug, also has sulfur bridges between cysteine residues 1-4 and 2-3, which are similar to those in endogenous endostatin. A few years later, the same group identified new proteolytic forms of endostatin and restin, with two different forms of O-glycosylation, with the use of chromatographic purification, followed by characterization and sequencing with mass spectrometry and Edman degradation [60].

II-A-2. Triple-helix stability and melting temperature

The amino acid sequence of collagen alpha chains contains specific motifs that allow the formation of a triple helix and the assembly of other supramolecular structures such as fibers and basement membranes. A single mutation in the primary amino acid sequence can prevent triple-helix formation and can, for that reason, result in a diseased state (e.g., osteogenesis imperfecta), which can be lethal [46]. Local variations in stability gives rise to micro-unfolding that can lead to, for example, enzyme activity, cell attachment, and remodeling of connective tissue.

The thermal stability of the collagen triple helix is defined by its Tm. In general, if the temperature equals a protein's Tm, half of these proteins are folded, and half are unfolded; if the temperature exceeds the Tm, then these proteins unfold. The Tm of a triple helix is assumed to be a few degrees above body temperature [61]. The Tm of collagen type I, however, is a few degrees below body temperature, and this probably holds for other collagen types as well [61]. At a too low Tm, the triple helix formed would unfold before a supramolecular structure could be formed. The denaturation time is similar to the time required to build a triple helix into a supramolecular structure. In addition, Leikina et al. state: "Our data support an earlier hypothesis that in fibers collagen helices may melt and refold locally when needed, giving fibers their strength and elasticity".

When a fiber is formed from multiple triple helices, the Tm increases strongly by approximately 27 °C, to 72 °C [62]. The large increase in Tm is probably the result of additional stabilization due to increased interactions between the various triple helices and the lack of water in the formed fiber, making the fiber more compact.

Apart from the primary amino acid sequence, PTMs also help increase the triple-helix stability as a result of additional cross-links and extra hydrogen bridges [63]. The amount of PTMs increases by culturing fibroblasts at an elevated temperature, probably due to a slower folding rate, which will result in a longer reaction time for the enzymes responsible for PTM formation to modify amino acids [64]. The Tm of the triple helix will probably not exceed the elevated temperature unless more PTMs are added.

II-B. POST-TRANSLATIONAL MODIFICATIONS

II-B-1. Overview of post-translational modifications and enzymes involved

Besides the characteristic tripeptides in collagen, other characteristics of great importance for triple-helix stability and cross-linking can be found in the primary structure, namely PTMs. The most common PTMs can be divided into three groups: hydroxylation/oxidation, glycosylation, and cross-linking.

The two amino acids most often involved in collagen PTMs are proline and lysine. The most important PTMs are shown in figure 3. The most important enzymes involved in PTM formation are given in table 2. Molecular oxygen is important in the hydroxylation/oxidation of proline and lysine. A lack of oxygen reduces the amount of hydroxylated proline and lysine; hydroxylation of proline becomes the rate-limiting step of triple-helix formation, leading to fewer cross-links, reduction of fiber organization and density, and loss of tensile strength [65-68].

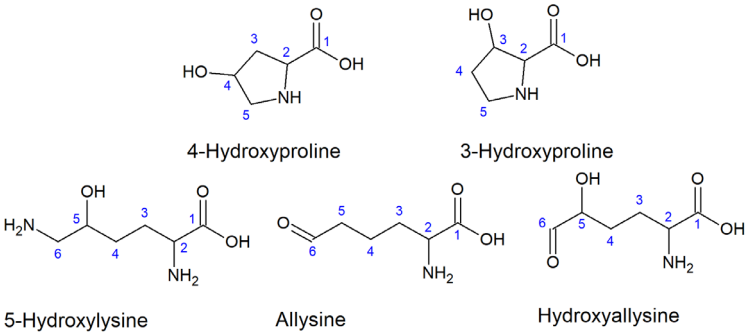


Figure 3. PTMs of proline and lysine.

Table 2. Most important enzymes involved in collagen PTM formation.

PTM	Enzymes involved	Uniprot code
4-hydroxyproline	P4HB	P07237
	P4HA1	P13674
	P4HA2	O15460
	P4HA3	Q7Z4N8
3-hydroxyproline	P3H1	Q32P28
	P3H2	Q8IVL5
	P3H3	Q8IVL6
	CTRAP	O75718
	CYPB	P23284
5-hydroxylysine	PLOD1	Q02809
	PLOD2	O00469
	PLOD3	O60568
Allysine	LOX	P28300
	LOXL1	Q08397
	LOXL2	Q9Y4K0
	LOXL3	P58215
	LOXL4	Q96JB6
5-hydroxyallysine	Combination of 5-hydroxylysine and allysine enzymes	
O-glycosylation	PLOD3	O60568
	COLGALT1	Q8NBJ5
	COLGALT2	Q8IYK4

Mass spectrometry can be used to analyze collagen PTMs. Collagen PTMs have been mapped with mass spectrometry for a number of collagen types (COL5A1 Bos Taurus, COL2A1 Bos Taurus, COL4A1 Mus musculus, and COL4A1 human) [69-73]. Mapping all PTMs requires, among other things, different enzymes and complementary fragmentation techniques. Uncertainties will remain, even then. This approach has also resulted in the interesting finding of a hydroxylated proline at the Xaa position.[69-73]. Its hydroxylation form is ambiguous, because hydroxyproline at the Xaa position does not match known enzyme activity; see II-B-2 and II-B-3. With the use of mass spectrometry it proved possible to establish that the PTM was either 3-hydroxyproline or 4-hydroxyproline, as shown by Kassel et al. in mussel adhesive proteins [74]. However, van Huizen et al. recently obtained, with mass spectrometry, the proof of principle that hydroxyproline at the Xaa position (amino acid position 584) in COL1A2 is 4-hydroxyproline [75]. The authors suggested naming this new PTM '4xHyp'. The 'x' differentiates between the Xaa and Yaa position. The identification of 4xHyp was achieved by applying ETD-HCD (MS/MS/MS) fragmentation on GLHGEFGLP(4Hyp)GP(?xHyp, pos. 584)AGPR, which contains the 4xHyp PTM of interest. Synthetic peptides containing 3Hyp, 4Hyp, Pro at the position of the 4xHyp were measured. The peptide containing 3Hyp and the peptide containing 4Hyp had different retention times and fragmented differently, whereby m/z 400 and 454 were distinctive between 3Hyp and 4Hyp, see the mass spectra in figure 4. The exact function of 4xHyp is unknown.

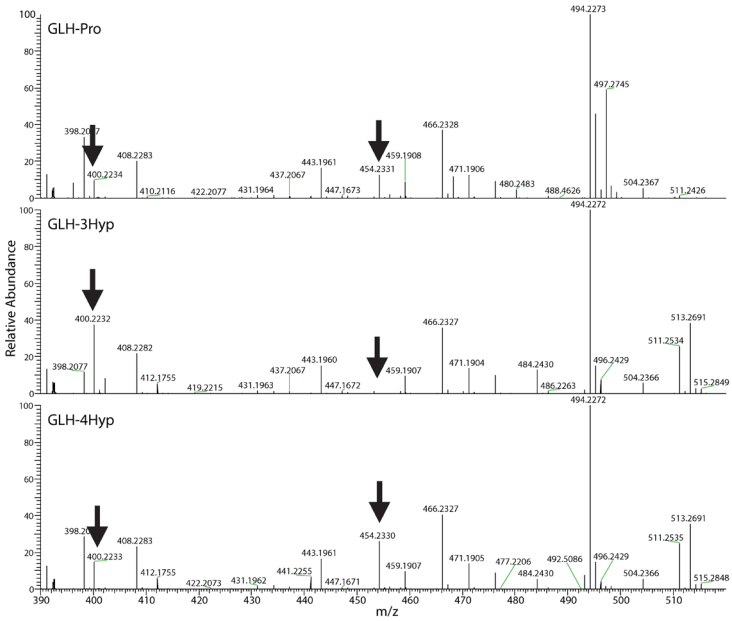


Figure 4. Zoom-in of the ETD-HCD mass spectra of the synthetic peptides acquired by direct infusion. GLH-3Hyp and GLH-4Hyp are distinguishable by singly charged fragments at m/z 400 and 454.[75]. Reprinted with permission from the authors, 2019 ACS Publications.

II-B-2. (2S,4R)-4-Hydroxyproline

Below, the normal 4-hydroxyproline is discussed, not the newly identified 4xHyp.

(2S,4R)-4-Hydroxyproline (in the following section referred to as 4-hydroxyproline or 4Hyp as three-letter amino acid code) is the most common PTM in collagen. Its function is well known: building hydrogen bridges -via one to three water molecules- with other amino acids from one of the other three alpha chains, therefore increasing stability.

Proline is hydroxylated into 4Hyp via enzymatic activity of prolyl 4-hydroxylase (P4H). Prolyl 4-hydroxylase consists of two alpha-subunits and two beta-subunits (protein disulfide isomerase) [76, 77]. Three alpha subunits are known for prolyl 4-hydroxylase: P4HA1, P4HA2, and P4HA3. The chemical reaction is shown in figure 5.

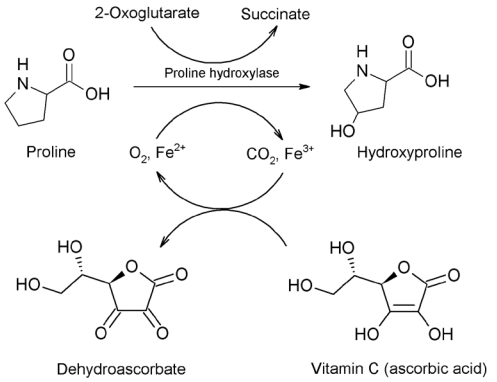


Figure 5. Chemical reaction of the hydroxylation of proline by proline 4-hydroxylase.

The Gly-Xaa-Pro substrate is required for hydroxylation of proline with P4H [78]. Amino acids at the Xaa position strongly influence the degree of proline hydroxylation [79].

A minimum number of hydroxylations is required to obtain a Tm that is sufficient to prevent unfolding before the triple helix can be built into supramolecular structures. Utting et al. found with amino acid analysis that, if hydroxylation of proline became the rate-limiting step of collagen formation, then the amount of hydroxyproline decreased from 43.9% to 42.2% [68]. A study by Rapaka et al. indicates a certain randomness in the hydroxyproline pattern [79]. The hydroxylation pattern should be further assessed to identify possible prevalent hydroxylation positions. To this aim, it would be interesting to compare the collagen hydroxylation pattern on the individual level. Mass spectrometry would be the ideal technique to this aim. The

randomness of hydroxyproline can be determined in tissues and cell cultures with standard bottom-up proteomics and with proteomics databases (PRIDE archive). Montgomery et al. showed with mass spectrometry the frequency of occurrence of proline hydroxylation in COL1A1 [73]. However, information regarding the variation between organs, diseases, and organisms is still lacking.

Three alpha chains self-associate at a specific point by sulfur bridges between several cysteine's; from that point onwards the triple helical folds in the way a zipper is closed [80, 81]. The rate-determining step of collagen folding is cis/trans isomerization of proline by peptidyl-prolyl cis-trans isomerase (PPIase), whereby trans proline is required for folding. During enzymatic hydroxylation of proline into 4Hyp by P4H, only 4R-hydroxyproline is formed. The possible influence of 4S-hydroxyproline on the cis/trans isomerization was tested by Bretscher et al. [82]; it was found that a hydroxyl group present in the S-position reduced the cis/trans ratio and lowered the Tm dramatically.

Furthermore, an error in enzyme P4H, or aberrant levels of the necessary cofactors, can influence the enzyme's function. Non-optimally functioning of the enzyme likely results in under-hydroxylated collagen. As described in II-A-2, under-hydroxylated collagen is less likely to be built into supramolecular structures. Giunta et al. showed with HPLC that under-hydroxylated collagen is present in the urine of patients with a specific form of Ehlers-Danlos syndrome [83]. Zn^{2+} is present in the cell at a higher level, and possibly interferes with the cofactor Fe^{2+}, thereby inhibiting proline hydroxylation.

Levels of 4Hyp in a sample can be determined with LC-MS and GC-MS. For LC-MS analysis, hydrolysis with hydrochloric acid is required [84-87]; GC-MS analysis requires an additional derivatization step to measure 4Hyp [88]. For LC-MS analysis, glycylphenylalanine [86] and N-methyl-L-proline [84, 85, 87] may serve as internal standards. For mass spectrometry analysis, however, it is recommended to use a stable isotope-labelled version of the analyte of interest [89, 90]. The analysis of 4Hyp will give information on the change in collagen levels. The original articles regarding the LC separation of hydroxyproline mention the separation of cis- and trans-4-hydroxyproline, and the separation of 3-hydroxyproline and 4Hyp [91, 92]. However, in later publications the separation of hydroxyproline isomers is not discussed. Probably, 3-hydroxyproline and 4-hydroxyproline levels were obtained simultaneously, which gives biased results.

II-B-3. (2S,3S)-3-Hydroxyproline

Apart from 4Hyp, (2S,3S)-3-hydroxyproline also exists as a proline PTM (3-hydroxyproline or 3Hyp as three-letter amino acid code). The mass of 3Hyp (113.05 Da) is equal to that of

its isomer 4Hyp and similar to those of its isobars leucine/isoleucine (113.08). The mass of the tripeptide Gly-Ala-Hyp is identical to that of its isomer Gly-Ser-Pro. The differentiation between Gly-Ala-Hyp and Gly-Ser-Pro can, with mass spectrometry, be based on b- and y-ions. Kassel et al. demonstrated that 3Hyp, 4Hyp, leucine, and isoleucine have different w- and a-ions, by which they can be distinguished [74]. W- and a-ions are formed upon fragmentation of amino acids side chains.

3Hyp is formed by an enzymatic reaction of prolyl 3-hydroxylase with proline. Until now, three different subunits are known for prolyl 3-hydroxylases (P3H): P3H1, P3H2 and P3H3.[93]. P3H1 forms a complex with cartilage-associated protein (CTRAP) and prolyl cis-trans isomerase cyclophilin B (CYPB) [94]. The reaction scheme of prolyl 3-hydroxylase is similar to that of prolyl 4-hydroxylase [95]. The discovery that mutations of CTRAP could cause a recessive form of osteogenesis imperfecta has increased the interest in 3Hyp [94, 96].

3Hyp is not present in all collagen types; it is present only in collagen types I, II, III, IV, V, and XI [97]. The amount of 3Hyp per collagen alpha chain is much lower than the amount of 4Hyp; the hydroxylation of proline into 4Hyp is sometimes even assumed to be 100% [98]. While 3Hyp occurs only at 1 or 2 positions in collagen types I and II, it occurs at 3-6 positions in collagen types V and XI. COL1A1 contains 47x Gly-Pro-Pro, which indicates that many possible substrates exist for P3H. However, in collagen type IV, which is part of the basement membrane, 10% of the total number of hydroxyprolines can be a substrate for 3Hyp. Furthermore, with mass spectrometric analysis it was shown that the 3Hyp frequency is tissue type-dependent [99]. Mass spectrometry analysis of periodontal ligament in human and mouse showed an absence of 3Hyp in collagen type I [100].

Eyre et al. and Weis et al. identified 3Hyp with mass spectrometry in several collagen types; subsequent analysis of the surrounding amino acid sequences was performed for sequence commonalities [99, 101]. An overview of the identified 3Hyp substrates is shown in table 3.

Table 3. In literature mentioned 3Hyp positions [99, 101]

Collagen type	Position[1]	Site[2]	Sequence[3]
V α-1	434	B3	GET-GFQ-GKT-GPP-GPP-GVV
XI α-1	434	B3	GET-GFQ-GKT-GPP-GPG-GVV
XI α-2	434	B3	GEV-GFQ-GKT-GPP-GPP-GVV
V α-2	470	A4	GFQ-GLP-GPP-GPP-GEG
V α-1	665	B2	GDE-GPR-GFP-GPP-GPV-GLQ
XI α-1	665	B2	GDE-GAR-GFP-GPP-GPI-GLQ
XI α-2	665	B2	GDE-GTR-GFN-GPP-GIV-GLQ

Table 3 continued.

Collagen type	Position[1]	Site[2]	Sequence[3]
V α-1	692	B1	GDV-GQM-GPP-GPP-GPR-GPS
XI α-1	692	B1	GDV-GPM-GPP-GPP-GPR-GPQ_
XI α-2	692	B1	GDV-GPM-GPP-GPP-GPR-GPA
I α-2	707	A3	GFP-GAA-GRT-GPP-GPS
V α-2	707	A3	GFP-GSA-GRV-GPP-GPA
II α-1	944	A2	GFT-GLQ-GLP-GPP-GPA
V α-2	944	A2	GFT-GLQ-GLP-GPP-GPN
I α-1	986	A1	GLN-GLP-GPI-GPP-GPR
II α-1	986	A1	GAN-GIP-GPI-GPP-GPR
V α-2	986	A1	GNP-GPL-GPI-GPP-GVR

[1]Position (positional counting) includes signal peptides and other parts which are nominally present.

[2] Modification that occurs at a certain site in the alpha-chain.

[3]P is 3Hyp, P reported as 4Hyp.

From table 3, we can conclude that the common substrate for site A1 is G?N-G"I/L"?-GPI-GPP-GPR; this motif is not present in the other sites. The "?" in the motif represents the possibility of implementation of different amino acids. In the first tripeptide part (GLN, GAN, or GNP), the position of asparagine is variable. In the second tripeptide part, the positons of proline and (iso)leucine are variable. This variability is probably not of much influence. For sites A2, A3, A4, and B2, the only shared motif is a phenylalanine (F), nine positions or less to the left of 3Hyp (P). In B1, phenylalanine is not present. It should be noted that the distance between sites A2 and A3 and that between A3 and A4 is 237 amino acids, and that the distance between B2 and B3 is 231 amino acids. These distances are remarkably close to the D-spacing (230-240 AA (≈60 nm)), which varies per tissue type [102]. The D-spacing or D-period is the distance between the N- and C- terminus of two triple helices in a fiber [103, 104]. This distance indicates that 3Hyp might play a role in fiber formation and other supramolecular structures in collagen. P3H1 is assumed to be responsible for 3Hyp at position 986. Sites A2, A4, B1, and B2 are assumed to be formed by P3H2. More research is necessary to more fully understand the mechanism for the formation of 3Hyp and its influence on the functionality in collagen. Indeed, the influence of 3Hyp on the collagen Tm is of interest to understand the possible function of 3Hyp. Jenkins et al. analyzed the influence of 3Hyp on the Tm of a guest peptide [105]. Substitution of a 3Hyp at the Yaa position resulted in a decreased Tm (-10°C), whereas substitution of a 4Hyp increased the Tm (+4°C). Surprisingly, after substitution of more than one 3Hyp, the Tm started to increase again [106].

II-B-4. Lysine

Lysine, too, can have PTMs. Lysine can be modified into two different derivatives: 1) allysine, whereby the ε amine of lysine is transformed into an aldehyde; and 2) 5-hydroxylysine,

whereby the 5th carbon is hydroxylated, comparable to hydroxyproline. The amine group of hydroxylysine can be further oxidized into an aldehyde (hydroxyallysine). While 5-hydroxylysine is formed inside the cell, allysine and hydroxyallysine are formed in the ECM. The chemical structures of the lysine PTMs are given in figure 3.

All lysine modifications are a result of enzymatic activity and can be intermediate steps for further modifications, such as cross-linking and glycosylation. The abundance of allysine and that of hydroxylysine in the C- and N-non-helical domains are tissue-dependent: allysine is dominant in soft connective tissue (e.g., skin); hydroxyallysine is dominant in skeletal connective tissue (e.g., bone, tendon) [107, 108]. In this review, only the formation of 5-hydroxylysine, allysine, and hydroxyallysine and their functions are discussed; other PTMs based on these lysine modifications are discussed later (glycosylation, cross-linking).

II-B-4a. 5-Hydroxylysine

An intermediate step of glycosylation of collagen, 5-Hydroxylysine, is formed by an enzymatic reaction of lysyl hydroxylase (LH) with lysine; the hydroxylation pathway is comparable to prolyl 4-hydroxylase. Lysyl hydroxylase also goes by the name of procollagen-lysine, 2-oxogluterate 5-dioxygenase (PLOD). The required substrate is Xaa-Lys-Gly in the helical domain and Xaa-Lys-Ala or Xaa-Lys-Ser in the telopeptides (both C and N termini) [109]. So far, three slightly different isoforms of lysyl hydroxylase have been identified in humans [110].

Lysine in the helical region is hydroxylated by lysyl hydroxylase 1. It has been suggested that lysyl hydroxylase 2 (LH2) modifies lysine in the telopeptides [111]. Lysyl hydroxylase 3 (LH3) is assumed to hydroxylate lysine, galactosylation, and glycosylation of 5-hydroxylysine [111].

II-B-4b. Allysine

Lysine can also be modified into allysine by lysyl oxidase (protein-lysine 6-oxidase, LOX), in which reaction the amine group is transformed to an aldehyde group. The reactive aldehyde group is important for collagen cross-linking. In total, five lysyl oxidase enzymes are known – four belong to the family of lysyl oxidase like proteins (LOXL). Lysyl oxidase is strongly bound to copper(II) [112]. As opposed to lysyl hydroxylation and prolyl hydroxylation, which are intracellular processes, lysyl oxidation is an extracellular process [113]. The latter's reaction scheme is different too: lysine + O_2 + H_2O ➔ (LOX) ➔ allysine + NH_3 + H_2O_2 [114].

LOXL-2 is involved in collagen type IV lysyl oxidation, whereas other LOX and LOXL enzymes are involved in lysyl oxidation of fibril-forming collagen types. The enzymatic activity of LOX is not limited to collagen – elastin is also oxidized by LOX [115]. The substrate specificity of LOX has been studied in a similar manner to P4H. The kinetics of LOX was determined

with fluorescence measurements of the formed H_2O_2 [116]. In general, LOX oxidizes lysine in the presence of a variety of other adjacent amino acids; the reaction rate is influenced by these other amino acids. For example, the hydroxylation rate decreases with increasing size of the side chain (Ala>Val>Leu>Phe); and a negative charge located next to lysine (e.g., glutamic acid) also negatively affects the hydroxylation rate. On the other hand, a positive charge (e.g., lysine or arginine) located next to lysine exerts a positive effect on the hydroxylation rate [116].

II-B-4c. Hydroxyallysine

Hydroxyallysine is formed by a combined enzymatic activity of lysyl hydroxylase and lysyl oxidase. The formation of 3-hydroxyallysine occurs by hydroxylation of lysine, followed by oxidation of the amine into an aldehyde. Due to the highly reactive character of the aldehyde, 3-hydroxyallysine will rapidly form cross-links (see also section II-B-7).

II-B-5. Collagen glycosylation

Protein glycosylation is known for its complexity. During glycosylation, sugar branches (glycans) are added to proteins. Glycans can be attached to oxygen or nitrogen of the functional sidechain of an amino acid. Both O- and N-glycosylation are present in collagen [117, 118]. We refer to a review by Perdivara et al. in which the current status of collagen O-glycosylation analysis with mass spectrometry is discussed [118]. Furthermore, the O-glycosylation of collagen type IV has been mapped with mass spectrometry by Basak et al. [71].

In comparison to O-glycosylation, hardly anything is known about N-glycosylation of collagen.[119, 120] With mass spectrometry, the glycans attached by N-glycosylation to collagen can be identified [121, 122].

II-B-5-a. O-linked collagen glycosylation

The two forms of O-linked glycosylation that occur in collagen are a galactose connected to 5-hydroxylysine (galactosyl-hydroxylysine, G-Hyl), or a galactose-glucose connected to 5-hydroxylysine (glucosylgalactosyl-hydroxylysine, GG-Hyl). Glycosylation of 5-hydroxylysine in collagen is achieved by hydroxylysyl galactosyltransferases, galactosylhydroxylysyl glucosyltransferases, and lysyl hydroxylase 3. The chemical structure of GG-Hyl is shown in figure 6. Schegg et al. showed the possibility to identify new galactosyltransferases with a combination of affinity chromatography and mass spectrometry [123]. The function of O-linked glycosylation is largely unknown, but it has been suggested that collagen glycosylation is involved in collagen fibrillogenesis and cross-linking.

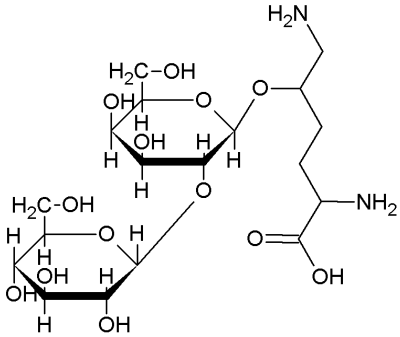


Figure 6. 2-O-α-D-glucopyranosyl-O-β-D-galactopyranosylhydroxylysine (GG-Lys).

II-B-5b. Fibrillogenesis

Experiments have shown that GG-Hyl influences collagen-fiber formation. The diameter of a glycosylated collagen fiber decreases with increasing glycosylation. The high hydrophilic character of the glycosylation could induce formation of a water layer around the glycans. The glucose-galactose unit is assumed to be oriented parallel to the triple helix to shield three to four amino acids from interactions with other triple helices. These phenomena are expected to influence collagen-fiber formation [64, 124].

II-B-5c. Cross-linking

Initial analysis of collagen cross-links in combination with glycosylation has resulted in two hypotheses. 1) Immature cross-links are often connected to GG-Hyl, whereas 2) mature cross-links are more often connected to G-Hyl or non-glycosylated allysine. Cross-linking will be discussed in more detail in II-B-7. Mass spectrometry demonstrated that the relation between mature cross-links and glycosylation was a tissue-specific one [125, 126]. The relation between collagen glycosylation and cross-linking has not been proven irrefutably, and current experiments point to a more tissue-specific correlation [126].

II-B-6. Advanced glycation end-products (AGE)

Apart from enzymatic glycosylation, also non-enzymatic glycosylation occurs (glycation), whereby carbohydrates react with proteins, lipids, or nucleic acids [127]. Cross-linking occurs by further reaction of glycation moieties (advanced glycation end products (AGE)). Together with certain types of cross-linking, AGE is a factor involved in collagen maturation. In general, AGE only affects proteins with turnover longer than a few weeks.

AGE formation starts by the reaction of saccharides (fructose, glucose etc.) with the amine groups of proteins to form a Schiff base, and the saccharides further into ketoamines

(Amadori products). Irreversible cross-links are formed by a further reaction known as Maillard reaction. An overview of this reaction scheme is shown in figure 7. The rate of AGE formation depends on the saccharides involved; for example, fructose is approximately seven times more reactive than glucose [128].

AGE affects the mechanical characteristics of collagen such that it becomes stiffer and less elastic [129]. These changes can be harmful, depending on the location of collagen. For example, loss of elasticity in vascular walls could lead to an increased blood pressure [130], tendons could become less viscoelastic and stiffer [131-133]; bone tissue could lose plasticity and toughness [134, 135], and cartilage could become more fragile [136]. The amount of pentosidine (an AGE product) increases linearly with age in dura mater, and is also increased in diaphysial femurs; in both tissue types the amount of pentosidine increases 4-5 fold from the age of 10 to 80 [137, 138].

Collagen AGE can be analyzed with mass spectrometry [139, 140]. Holte et al. took biopsies of the tissue of interest, and cleaved proteins into amino acids [139]. Eight different AGEs were measured with a triple quadrupole mass spectrometer. These AGEs were present in the pmol/mg range. Mikulikova et al. added AGEs to collagen in vitro by incubation of collagen with different sugars [140]. After incubation, collagen was reduced with mercaptoethanol, and cleaved into peptides by addition of CNBr followed by addition of trypsin. Samples were analyzed with CE-MS/MS and HPLC-MS/MS. Collagen peptides with AGE were identified with CE-MS/MS and with HPLC-MS/MS, whereby HPLC-MS/MS was more sensitive.

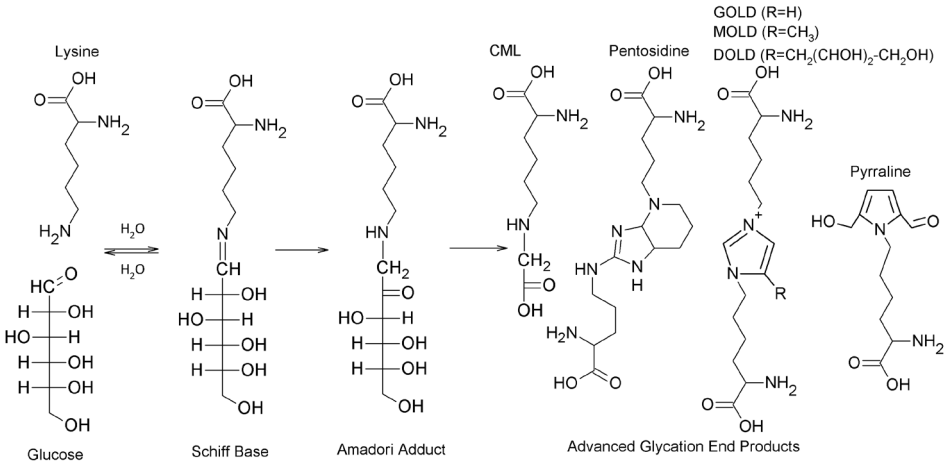


Figure 7. Overview of AGE formation in proteins. GOLD, MOLD, DOLD: glyoxal-, methylglyoxal-, and 3-deoxyglucosone-derived lysine dimers.

II-B-7. Cross-linking

Enzymatic cross-linking is initiated by oxidation of lysine into allysine or hydroxyallysine. The aldehyde group is highly reactive, and reacts easily with nearby collagen lysines. In contrast to the hydroxylation of proline and lysine, this modification occurs outside the cell in the ECM. A flow scheme of collagen cross-linking is shown in figure 8. The formation of cross-linking consists of several intermediate steps. Initially formed cross-links are immature and reversible. Further reaction with a third amino acid (trivalent cross-linking with a lysine or histidine) results in mature cross-links, which are irreversible. These cross-links are illustrated in the boxes at the right-hand side of figure 8 inside a double-edged box. The only reversible trivalent cross-link is present in dehydro histidinohydroxymerodesmosine (deH-HHMD, single-edged box).

The type of cross-link formed is tissue-dependent. Skin and tendon contain mainly lysine-derived cross-links, whereas bone, cartilage, and dentin contain mainly hydroxylysine-derived cross-links. The amount of trivalent cross-links per mol of collagen differs among various tissue types. In skin it is 280 mmol/mol; in bone tissue 495 mmol/mol; in patellar tendon: 870 mmol/mol; and in articular cartilage 1800 mmol/mol [67].

O-linked glycosation and cross-linking can be present together on the same lysine. Lysines containing an immature cross-link have more often glucosylgalactosyl-hydroxylysine attached [126]. Mature cross-links more often contain galactosyl-hydroxylysine. It is assumed that the larger carbohydrates hinder cross-link maturation.

Collagen cross-links can be directly studied with, for example, collagen digestion and analysis of peptides that contain the cross-link with mass spectrometry. Eyre et al. showed the possibilities of such a method in the study of collagen cross-linkers [141]. The chemical structure of a specific collagen cross-link had remained unknown for two decades. With the aid of mass spectrometry, this cross-link was identified as sulfilimine, and it is present as S-lysyl-methionine and as S-hydroxylysyl-methionine [142].

Also, it has been shown indirectly that the collagen turnover in bone tissue can be monitored by measuring urine concentrations of pyridinoline (Pyr) and deoxypyridinoline (d-Pyr). Patients with bone fractures or osteoporosis showed a significant increase in the excretion of Pyr and d-Pyr [143]. The constant ratio between Pyr and d-Pyr indicates that bone tissue was the source of origin [143]. Pyridinoline-derived cross-links can be measured easily with a fluorescence detector, because pyridinoline is a natural fluorescent compound [144].

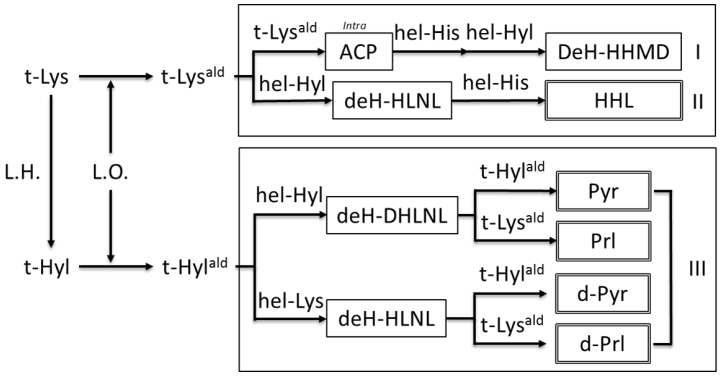


Figure 8. A flow chart of the major cross-links in collagen type I. Double edged boxes indicate irreversible cross-linking while single edged boxes indicate reversible cross-linking. The roman numbers on the right-hand side indicate: I, dominant in normal tissue; II, dominant in collagen type I, located in the skin and cornea; III, dominant in skeletal collagen. L.H., lysyl hydroxylase; L.O. lysyl oxidase; t, telopeptide; hel, triple helix; ACP, aldol condensation product; deH, dehydro; HLNL, hydroxylysinonorleucine; DHLNL, dihydroxylysinonorleucine; HHMD histidinohydroxymerodesmosine; HHL, histidinohydroxylysinonorleucine; d-, deoxy; Pyr, pyridinoline; Prl, pyrrole.

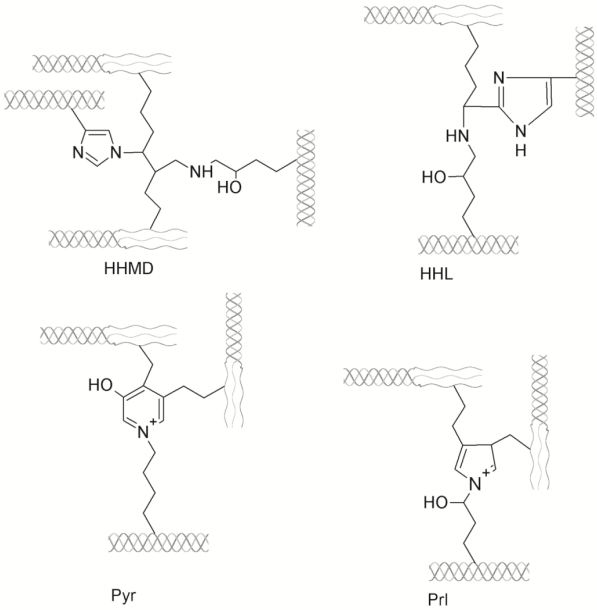


Figure 9. The chemical structures of the final collagen cross-link form.

Couppe et al. have given a short overview on the effect of aging on the amount of collagen and cross-links present in tendons [145]. These authors measured collagen cross-links levels with HPLC-fluorescence and pointed out that the literature is inconsistent with respect to the effects of aging in animals and humans on the amount of collagen and collagen cross-links present in tendon. It is not clear why these effects might differ; however, prior life-long training history might play a role.

II-C. FUNCTIONAL CHARACTERISTICS OF COLLAGEN

II-C-1. The collagen family

The previous section focused mainly at collagen PTM structures and cross-linking in relation to structural chemistry. In this section, the focus will be on the normal biological function of the various supramolecular structures inside the human body and in pathology. A complete overview of all collagen types is given in table 4; an overview of different types of supramolecular structures is shown in figure 10.

All collagen types in the collagen family contain a triple helix. The size of the triple-helix domain with respect to the total collagen length can vary from 96% (collagen type I) to less than 10% (collagen type XII). The length of a collagen domain varies from 75 nm (collagen type XII) up to 425 nm (collagen type VII) [146]. However, the triple-helix structure is not restricted to collagen; non-collagen proteins (e.g. Emilin, Ficolins 1, 2, and 3) show also collagen-like domains (Gly-Xaa-Yaa pattern; see [147]). For visualization of the fibers, see van der Rest et al. [22].

Table 4. Molecular composition of collagen triple helices in various tissues.

Type	Class	Triple helix composition	Tissue distribution
I	Fibril-forming	$[\alpha1(I)]_2\ \alpha2(I)$	Abundant and widespread; bone, dermis, tendon, ligaments, cornea
II	Fibril-forming	$[\alpha1(II)]_3$	Cartilage, vitreous body, nucleus pulposus
III	Fibril-forming	$[\alpha1(III)]_3$	Skin, vessel wall, reticular fibers of most tissue (lungs, liver, spleen, etc.)
IV	Basement membrane	$[\alpha1(IV)]_2\ \alpha2(IV)$ $\alpha3(IV)\ \alpha4(IV)\ \alpha5(IV)$ $[\alpha5(IV)]_2\ \alpha6(IV)$	Basement membranes
V	Fibril-forming	$[\alpha1(V)]_3$ $[\alpha1(V)]_2\ \alpha2(V)$ $\alpha1(V), \alpha2(V), \alpha3(V)$	Widespread: lung, cornea, bone, placenta, fetal membranes; together with type I collagen
VI	Microfibrillar	$\alpha1(VI), \alpha2(VI), \alpha3(VI)$ $\alpha1(VI), \alpha2(VI), \alpha4(VI)$	Widespread; dermis, cartilage, placenta, lungs, vessel wall, intervertebral disc

Table 4 continued.

Type	Class	Triple helix composition	Tissue distribution
VII	Anchoring fibrils	$[\alpha1(VII)]_3$ $[\alpha1(VII)]_2\ \alpha2(VII)$	Skin, bladder, dermal-epidermal junctions; oral mucosa, cervix
VIII	Hexagonal network-forming	$[\alpha1(VIII)]_3$ $[\alpha2(VIII)]_3$ $[\alpha1(VIII)]_2\ \alpha2(VIII)$	Widespread; dermis, brains, heart, kidney, endothelial cells, Descemet's membrane
IX	FACIT	$\alpha1(IX)\ \alpha2(IX)\ \alpha3(IX)$	Cartilage, vitreous humor, cornea
X	Hexagonal network-forming	$[\alpha1(X)]_3$ $[\alpha3(X)]_3$	Hypertrophic cartilage
XI	Fibril-forming	$\alpha1(XI)\ \alpha2(XI)\ \alpha3(XI)$	Cartilage, vitreous body, intervertebral disc
XII	FACIT	$[\alpha1(XII)]_3$	Perichondrium, ligaments, tendon, dermis
XIII	MACIT	$[\alpha1(XIII)]_3$	Epidermis, hair follicle, endomysium, intestine, chondrocytes, lungs, liver, eye, heart
XIV	FACIT	$[\alpha1(XIV)]_3$	Widespread; bone, dermis, tendon, vessel wall, placenta, lungs, liver, cartilage
XV	Multiplexins	$[\alpha1(XV)]_3$	Fibroblasts, smooth muscle cells, kidney, pancreas, testis, capillaries
XVI	Multiplexins/FACIT	$[\alpha1(XVI)]_3$	Fibroblasts, amnion, keratinocytes, dermis, kidney
XVII	MACIT	$[\alpha1(XVII)]_3$	Dermal-epidermal junctions, hemidesmosomes in epithelia
XVIII	Multiplexins	$[\alpha1(XVIII)]_3$	Lungs, liver, basement membrane
XIX	FACIT	$[\alpha1(XIX)]_3$	Human rhabdomyosarcoma, basement membrane
XX	FACIT	$[\alpha1(XX)]_3$	Corneal epithelium, embryonic skin, sternal cartilage, tendon
XXI	FACIT	$[\alpha1(XXI)]_3$	Blood vessel wall, stomach, kidney
XXII	FACIT	N/A	Tissue junctions
XXIII	MACIT	N/A	Heart, retina
XXIV	Fibril-forming	N/A	Bone, Cornea
XXV	MACIT	N/A	Brain, heart, testis
XXVI	FACIT	N/A	Testis, ovary
XXVII	Fibril-forming	N/A	Cartilage
XXVIII	N/A	N/A	Dermis, sciatic nerve

FACIT, fibril-associated collagen with interrupted triple helices; MACIT, membrane-associated collagen with interrupted triple helices; Multiplexin, multiple triple-helix domains and Interruptions; N/A, not available [21, 148, 149].

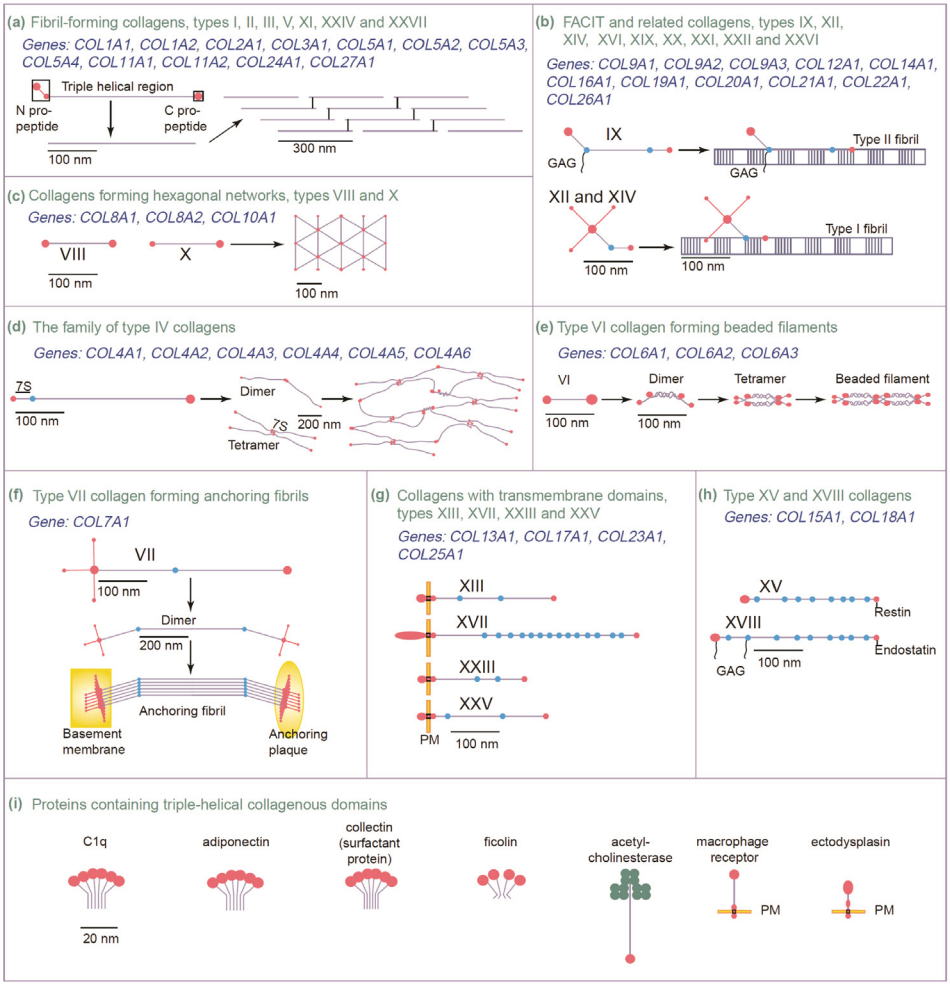


Figure 10. Eight different supramolecular structures which can be formed by members of the collagen family can be distinguished. a) Fibril forming collagen types. b) Fibril-associated collagen types with interrupted triple helices (FACITs), FACITs are located at the surface of fibrils. c) Hexagonal networks. d) Basement membrane formed by collagen type IV. e) Beaded filaments formed by collagen type VI. f) Anchoring fibrils for basement membrane formed by collagen type VII. G) Collagen types containing transmembrane domains. h) Collagen type XV and XVIII. I) proteins containing triple-helical collagenous domains.[21] Reprinted with permission from the authors, 2004 Elsevier.

II-C-2. Collagen-related diseases

Collagen is involved in many biological processes. Aberrant collagen turnover, genetic mutations, or aberrant PTMs can result in the collagen-related diseases shown in table

5. Although this table does not list tumors, collagen is also strongly involved in tumor progression and metastasis. In tumor development, there is probably an important role for the ECM. It has been proposed that the ECM can act as a tumor suppressor as long as it remains in its 'natural shape', and a tumor promoter activity if the 'natural morphology' is lost [150]. The 'natural morphology' of the ECM will change over time, for example, due to inflammation, growth factors, and accumulated damage [150]. In breast cancer, Ma et al. have shown upregulation of the gene expression of 23 collagen genes and of 94 other ECM genes [151]. Furthermore, Naba et al. and van Huizen et al. have shown that the collagen composition of colorectal liver metastasis is significantly different from that of healthy liver tissue [26, 152].

Table 5. Overview of collagen-related diseases.

Collagen type	Disease
I	Caffey disease
	EDS type I, II, VIIA, VIIB
	OI type I, II, III, IV
	Osteoporosis
II	Achondrogenesis, type II or hypochondrogenesis
	Avascular necrosis of the femoral head
	Legg-Calve-Perthes disease
	Osteoarthrosis
	SED congenita
	Several (chondro)dysplasias
	SMED Strudwick type
	Stickler syndrome, type I
III	Arterial aneurysms
	EDS type IV
IV	Alport syndrome
	Angiopathy, hereditary, with nephropathy, aneurysms, and muscle cramps
	Brain small vessel disease with or without ocular anomalies
	Hematuria
	Porencephaly type 1, 2
	Retinal arteries, tortuosity
	Schizencephaly
V	EDS type I and II
VI	Bethlem myopathy
	Ullrich congenital muscular dystrophy 1
	Dystonia 27
VII	Epidermolysis bullosa dystrophica, dystrophic forms
IX	Intervertebral disc disease
	Multiple epiphyseal dysplasia
	Osteoarthrosis
	Stickler syndrome, type IV, V

Table 5 conitnued.

Collagen type	Disease
X	Chondrodysplasia
	Schmid metaphyseal chondrodysplasia
XI	Deafness
	Fibrochondrogenesis type I
	Marshall syndrome
	Non-syndromic hearing loss
	Osteoarthrosis
	Several mild chondrodysplasias
	Stickler syndrome type II
XII	Bethlem myopathy 2
XIII	Myasthenic syndrome, congenital, 19
XVII	Epidermolysis bullosa
	Epithelial recurrent erosion dystrophy
XVIII	Knobloch syndrome
XXV	Fibrosis of extraocular muscles, congenital, 5
XXVII	Steel syndrome

OI, osteogenesis imperfecta; EDS, Ehlers-Danlos syndrome. [153] and omim.org (accessed 20-11-2017).

II-C-3. Collagen remodelling

The first step in collagen remodeling is the enzymatic cleavage by matrix metalloproteinases (MMPs); see Table 6 for an overview of all MMPs and the collagen types and other substrates they cleave. Remodeling is an important process in inflammation, wound healing, tumor proliferation, and exercising [154-156]. Besides collagen, MMPs can also degrade other ECM components. MMPs cleave collagen at specific points in the triple helix; i.e., at roughly three-quarters of the length of the helical domain in collagen types I, II, and III. More specifically, collagen type I alpha 1 is cleaved at site Gly^{775}-Ile^{776}, and type 1 alpha 2 is cleaved at site Gly^{775}-Leu^{776} [157]. Zhen et al. used mass spectrometry to identify novel cleavage products of various MMP types in cartilage [158]. The produced peptides could be of further interest for diseases in which collagen turnover takes place.

Table 6. The MMP family, the collagen types they cleave and other substrates [159-161].

Enzyme groups	MMP	Collagen type	Other substrates
Collagenases			
Collagenase-1	MMP-1	I, II, III, VII, VIII, X, XI	Gelatin
Collagenase-2	MMP-8	I, II, III, VII, VIII, X	Aggrecan, gelatin
Collagenase-3	MMP-13	I, II, III, IV, VI, X, XIV	Gelatin
Gelatinases			
Gelatinase A	MMP-2	I, II, III, IV, V, VII, X	Gelatin
Gelatinase B	MMP-9	IV, V, XI, XIV	Gelatin

Table 6 continued.

Enzyme groups	MMP	Collagen type	Other substrates
Stromelysins			
Stromelysin-1	MMP-3	II, III, IV, VII, IX, X, XI	Gelatin
Stromelysin-2	MMP-10	III, IV, V	Laminin, fibronectin, elastin
Stromelysin-3	MMP-11	IV	Laminin, fibronectin, aggrecan
Matrilysins			
Matrilysin-1	MMP-7	I, IV	Laminin, fibronectin, gelatin
Matrilysin-2	MMP-26	IV	Fibronectin, fibrinogen, gelatin
Membrane Type MMPs			
MT1-MMP	MMP-14	I, II, III,	Gelatin, fibronectin, laminin, proMMP-2
MT2-MMP	MMP-15		Gelatin, fibronectin, laminin, proMMP-2
MT3-MMP	MMP-16	III	Gelatin, fibronectin, laminin
MT4-MMP	MMP-17		Fibrinogen, fibrin
MT5-MMP	MMP-24		Gelatin, fibronectin, kaminin
MT6-MMP	MMP-25	IV	Gelatin
Others			
Macrophage metalloelastase	MMP-12	I, IV	Elastin, fibronectin
	MMP-19	I, IV	Aggrecan, elastin, fibrillin, gelatin
Enamelysin	MMP-20		Aggrecan
XMMP	MMP-21		Aggrecan
	MMP-23		Gelatin, casein, fibronectin
CMMP	MMP-27	Unknown	Unknown
Epilysin	MMP-28	Unknown	Unknown

After cleavage of a collagen triple helix, the triple helix will start to unfold and denaturate to increase accessibility for other enzymes to further cleave collagen [162]. For a more complete overview on the regulation of MMPs and the reaction mechanism, the review by Ala-aho et al. [162] is recommended. With hydrogen-deuterium exchange mass spectrometry, the sites where collagen type I binds with MMP-1 have been mapped [163, 164].

The triple helix of collagen is embedded in supramolecular structures that are cross-linked by enzymatic and non-enzymatic initiated cross links to thereby create a structure that is difficult to degrade. Thus, collagen in the human body has a long half-life, which in certain tissue types can, surprisingly, exceed a human life time. Yet, the half-life varies strongly between different organs: from 15 years in skin, 117 years in cartilage and to up to 95-215 years in vertebrate discs [165, 166]. Collagen half-life times in rats have been determined with isotope-ratio mass spectrometry: 45 days in muscle; 74 days in skin; and 244 days in gut [167]. In these rats, the half-life times of collagen type III were shorter than those of collagen type I [167]. Isotope-ratio mass spectrometry also revealed that the core of the Achilles tendon has almost no turnover after an individual's growth has stopped [168]. Furthermore, isotope ratio mass

spectrometry revealed that damaged cartilage from osteoarthritis did not show additional regeneration [169].

Apart from the long-term collagen turnover, also the short-term collagen turnover can be studied with mass spectrometry. Wilkinson et al. showed the possibility to detect day-to-day muscle protein synthesis and collagen turnover by administering deuterated water to eight healthy young men and have them perform one-legged resistance exercises. Deuterated water present during protein turnover will be incorporated in the newly formed proteins, forming heavy isotope labelled proteins. The one-legged resistance exercises stimulate extra muscle grow. The turnover of the growing muscle increases and more deuterated water will be incorporated in comparison to the non-exercised leg. The different amounts of heavy isotope labelled proteins between both legs can be analysed in biopsies with isotope ratio mass spectrometry [170]. Exercised legs showed a larger increase in deuterium-labelled collagen than did the non-exercised legs.

Using ELISA Karsdal et al. analysed collagen fragments in rat blood over the course of one year and found that the collagen turnover was strongly related to age [171]. The turnover in collagen types I and type II showed a strong decrease; that collagen type III increased over the first month and then stabilized; that of collagen type IV increased over time; and the turnover of collagen types V and VI remained fairly constant [171]. This study implicates that age can be a confounding factor when a collagen-related disease is studied.

III. TECHNIQUES TO ANALYZE COLLAGEN

Several techniques are available to study collagen. The three most commonly used techniques are mass spectrometry, circular dichroism spectroscopy, and staining, both chemical staining and immunological. These techniques will be discussed in this section, preceded by a short overview of other possible techniques to analyze collagen.

III-A OVERVIEW OF TECHNIQUES

Apart from mass spectrometry, circular dichroism, and staining, other techniques have been used to analyze collagen. X-ray based techniques, applied as early as the 1940s [172], determine the collagen triple helix structure [173]. Electron microscopy, also applied since the 1940s [174] visualizes the collagen fibril size and three-dimensional organization [175]. Likewise, second-harmonic generation (SHG) can also be used to visualize and study collagen fibers with a resolution of approximately 1 μm [176, 177]. Enzyme-linked immunosorbent assay (ELISA) is another way to quantify collagen in serum [178].

III-B. MASS SPECTROMETRY

The first publication on collagen and mass spectrometry dates back to 1970 [179]; this article describes the analysis of collagen cross-linking products with GC-MS. From then onwards, the number of publications on collagen and mass spectrometry steadily increased. In the period 2013-2018, between 100 and 150 articles were published annually. This rise in publication rate went hand in hand with the developments in the field of mass spectrometry, which made it possible to study (large) biomolecules such as peptides and proteins. The most important developments were the invention of electrospray ionization (ESI; Nobel Prize 2002, J.B. Fenn) [180] and matrix-assisted laser desorption/ionization (MALDI; Nobel Prize 2002, K. Tanaka) [181]. Other developments in the fields of bioinformatics and engineering have also been valuable for the analysis of biomolecules as they have increased the overall performance (e.g. sensitivity, speed, versatility) of mass spectrometry. For the reader interested in the development of mass spectrometry and proteomics we recommend the following articles: [182-186]. All these developments have made it possible that thousands of proteins and a multitude of peptides can be identified in a single measurement [187]. Below, an overview is given of the use of mass spectrometry for the study of collagen.

III-B-1. Sample processing for mass spectrometry analysis

Collagen can be analyzed in a wide variety of liquid (e.g. urine [11], serum [188], CSF [189]) and solid biomaterial (e.g. bone [190], tissue [152] with different mass spectrometry applications such as bottom-up proteomics [26], top-down proteomics [191], targeted analysis [12], MALDI-imaging [192], hydrogen-deuterium exchange [164], and isotope ratio mass spectrometry [170].

The most standard approach for proteomic analysis of collagen (bottom-up-, or shotgun proteomics) in tissue is the solubilization of proteins followed by reduction, alkylation, and digestion with an enzyme (often trypsin). Peptides produced by such a sample preparation method are online separated on a liquid-chromatography system (LC) followed by MS/MS identifications [193-195]. This standard approach is already suitable to study collagen, and van Huizen et al. identified with this approach 1,137 unique collagen peptides that belong to 22 different alpha chains in liver and colorectal liver metastasis [26].

Hard tissue such as bone and cartilage tissue, where collagen is highly abundant, requires a harsher extraction method than for soft tissue or cells. Addition of hydrochloric acid and heating is required to extract collagen from bones. This method works very well, and can, for example, be used to distinguish bone materials from different animal species with MALDI-TOF [196].

Another source of collagen is provided by stable isotope labeling by amino acids in cell

culture (SILAC). This method to label collagen produced by cells in culture enables to quantify specific collagen types, turn-over rates, total collagen levels, and collagen PTMs [197]. Another method is culturing of cells in Petri dishes followed by removal of cells and extraction of ECM proteins, including collagen [198]. Taylor et al. cultured mouse primary osteoblasts on coverslips [199]. After the culture period, the cells were removed (decellurization) and the produced ECM proteins remained behind on the coverslip. After a clean-up step of the ECM proteins, the coverslip with ECM proteins was inserted into a time of flight - secondary ion mass spectrometer (TOF-SIMS) for analysis. Collagen levels were indirectly analysed by measuring picrosirius red (staining with picrosirius red is discussed in more detail in III-D-1). The picrosirius red levels, and consequently collagen levels, differed under different culture conditions.

Enrichment of the sample for collagen leads to a higher number of identified peptides. There are several methods to enrich samples for collagen, of which two are especially suitable for tissue. One consists of decellularization), whereby ECM structures remain behind. The decellularized tissue can be further digested and analyzed with bottom-up mass spectrometry [200, 201]. In the other method, ECM proteins are extracted from tissue. Naba et al. showed the capability of such an extraction method to extract collagen from tissue and identify up to 32 different alpha chains in colon tissue [152, 202]. This extraction method is based on the extraction of proteins from different cell compartments (cytoplasm, nucleus, membrane, and cytoskeleton), whereby ECM proteins are left behind. These can be further digested and analyzed with mass spectrometry.

A disadvantage of protein digestion is the loss of spatial resolution, which can be prevented, however, with MALDI imaging mass spectrometry [203]. This approach will be discussed in more detail in III-D-3. Currently a resolution in the lower µm range can be achieved [204]. To image proteins in tissue sections with MALDI imaging, proteins are cleaved into peptides on a glass slide. These peptides can be ionized and subsequently visualized by a mass spectrometer.

Apart from enrichment of whole proteins, enrichment of peptides after protein digestion is also possible. Fractionation allows for a relative deeper view on the proteins present in a clinical sample. Roughly two ways of fractionation can be distinguished: online separation on an LC system where fractions are collected automatically and online measured with mass spectrometry, and off-line separation with, for example, disposable columns (e.g., ZipTips) or off-line automated affinity separation (e.g., separation of PTM-containing peptides).

Furthermore, the use of different enzymes to cleave proteins into peptides and the use of different fragmentation techniques of peptides in the mass spectrometer allow for the

mapping of hydroxyproline, hydroxylysine, and O-glycosylation in collagen [70-72]. Basak et al. used trypsin, Glu-C, and LysC to cleave proteins into peptides. In a part of the sample set, glycosylations were removed with PNGase F. Peptides were fragmented with both CID and ETD in an LTQ_Orbitrap Velos, and also on a Q_Exactive with HCD fragmentation. The use of different enzymes and fragmentation techniques resulted in a sequence coverage of 81.5% and 85% of mouse EHS tumor COL4A1 and human lens capsule COL4A1, respectively. Song et al. performed protein digestion by use of GluC and trypsin, and performed CID, HCD, and ETD fragmentation in an LTQ_Orbitrap Velos. This approach resulted in a sequence coverage of 90% (excluding signal-, pro-, and C-telopeptides) of pepsinized immunization-grade bovine COL2A1. Yang et al. performed digestion with GluC, chemotrypsin, trypsin, and a combination of GluC and AspN. Digestion was followed by LC-MS analysis, whereby peptides were fragmented with HCD and ETD, and a combination of MS^2 and MS^3 was applied. This protocol resulted in a 94% sequence overlap for collagen alpha-1(V) from bovine placenta and a 90% sequence overlap for human collagen alpha-1(V).

Collagen cross-linking can be analyzed with mass spectrometry in tissue [141, 205, 206]. Preparing samples for mass spectrometry analysis of collagen cross-links is commonly done by reduction of the sample with sodium borohydride followed by a clean-up step (e.g. filtration with a cellulose column). After the clean-up step or reduction step, the sample is hydrolyzed by addition of hydrochloric acid. After removal of the hydrochloric acid the sample is ready for LC-MS analysis [205, 206]. Eyre et al. give an overview of several methods to analyze collagen cross-links in bone and tendon. They also address the analysis of cross-link markers in urine, which markers can detect bone cancer with good accuracy. The methods described are based on mass spectrometry measurements, immunoassays, and fluorescence measurements [141].

Besides the digestion of collagen, naturally occurring collagen peptides can also be detected. Naturally occurring peptides can be purified with high molecular weight filters and a desalting step [207]. An alternative, more convenient way is the use of a C18 column for the clean-up of urine to perform qualitative (shotgun proteomics) and quantitative (targeted analysis) analysis of natural occurring collagen peptides [208, 209].This method allows detecting naturally occurring peptides with mass spectrometry also in serum [188]. Naturally occurring peptides can be analyzed in both a bottom-up (semi-quantitative) and in a targeted way (absolute quantification with stable-isotope labelled peptides).

Most of the above-described methods use an ESI-source connected to a LC. Less commonly used is capillary electrophoresis (CE) for chromatographic separation. Equipped with an UV detector, CE could separate collagen peptides and detect them directly [210-212] or via an online connection to a mass spectrometer. Many publications describe analyzing collagen in urine

with CE-MS [207, 213-216]; collagen was often part of a biomarker panel to predict the presence of a disease (e.g. diabetes, kidney diseases). Urine was prepared for CE-MS by addition of urea, ammonium hydroxide, and SDS.

Drube et al. added ureum, NH_4OH, SDS to urine, and removed the proteins with a 20 kDa filter [215]. The filtered samples were then applied to a desalting column to remove urea, electrolytes, and salt. The purified material was concentrated and injected into the CE-MS. This protocol is almost identical to that used by Broker et al. for the measurement of collagen NOPs in urine with LC-MS [11]. Molin et al. compared the performance of CE-MS and that of MALDI-MS to detect chronic kidney disease; collagen in urine was one the most significant markers of chronic kidney disease, irrespective of the system used [216].

III-B-2. Collagen stoichiometry

The stoichiometry of different alpha chains in a triple helix or the composition of supramolecular structures can be analyzed with mass spectrometry. Onnerfjord et al. showed that the collagen composition in cartilage collected from eight different locations in the body differed per location [217]. The collagen had been extracted with the use of hydrolysis induced by hydrochloric acid. Samples were reduced, alkylated, and digested with trypsin. Samples were combined using iTRAQ (isobaric tag for relative and absolute quantitation). Another example is the composition of collagen type IV in different locations, reported by Uechi et al. In-gel digestion of three different parts of the eye (blood vessels, inner limiting membranes, and lens capsules) followed by intensity-based absolute quantification (iBAQ) mass spectrometry analysis, revealed that the collagen type IV composition and alpha chain ratios differed between the tissues types [218]. Dreisewerd et al. showed that it is possible to ionize complete collagen alpha chains –and even dimers, trimers, up to hexamers– with an IR-MALDI-TOF and UV-MALDI-TOF [191]. The measurement of mono- (α), di- (β), tri- (γ, triple helix), tetra- (τ), and hexamers (η) leads to complex mass spectra. Examples shown by these authors indicate that overlapping masses exist for: $\alpha^+/\gamma^{3+}/\beta^{2+}$, $2\alpha^+/\beta^+$, and $3\alpha^+/\gamma^+$. To prevent overlapping molecular masses, size-exclusion chromatography was performed; however, overlapping masses were still present after size-exclusion chromatography. Also, the mass of a triple helix is not a single peak but a broad envelop with varying amounts of hydroxylations and glycosylations; the envelop was determined to be 283,300 ± 1,300 Da for collagen type I. High-resolution capabilities are required to separate all different isotopes, differently-charged species with a similar mass, and various numbers of hydroxylations and glycosylations. For example, an alpha chain is approximately 100 kDa when α^+/γ^{3+} are present, and a resolution of >33,000 is needed to separate the two masses. Such a resolution can be easily obtained in state-of-the-art high-resolution mass spectrometers.

III-B-3. The role of collagen in dating of bone

Collagen has a high in vivo half-life, sometimes even exceeding the life-time of the organism (see II-C-3 Collagen remodelling). Collagen in bone, for example, is stable up to 10,000 years and beyond [219], which implies that collagen can be used to date bones. Collagen bone dating can be performed in two different ways i.e. isotope ratio analysis and D/L racemization of amino acids [220]. For both analyses, collagen extraction is required; collagen can be extracted in different ways. A general procedure starts with cleaning of the bones, taking a sample of ca. 0.5 - 1 g, and dissolve this in HCl. Humics can be removed with addition of NaOH or rinsing the sample to a neutral pH, followed by gelatinization under acidic and heated conditions. Finally, the sample is filtered and analyzed with mass spectrometry. Jorkov et al., Piotrowska et al., and Cersoy et al. have compared several slightly different protocols that in essence are similar to the general procedure described above; the various protocols gave comparable results [221-223].

The prepared samples can be analyzed with acceleration mass spectroscopy (AMS). The strength of AMS is the ability to measure isotopes, and even distinguish between isobars, for example, nitrogen-14 and carbon-14. A review by Kutschera gives a good overview of the workings and applications of AMS [224]. Samples can also be measured with isotope-ratio mass spectrometry (IR-MS), which are sector instruments. The website www.planetisotopes.com provides an overview of the various instruments used in IRMS and their applications.

Quality of the bone tissue sample is determined by analyzing the total collagen content, the nitrogen content, carbon/nitrogen ratio, carbon-13, and nitrogen-15 levels. Carbon-14 levels relative to carbon-12 levels indicate the age of the original sample. For example, Hublin et al. showed that 40 bone samples from different locations in the Grotte du Renne (in France) were between 29,000 – 43,000 years old [225].

However, there are a few pitfalls in the analysis of isotope ratios in collagen. The most important one is the sample preparation; any contamination of younger organic material will affect the isotope ratios in the sample [226]. Because recent material will have a different isotope ratio than the sample of interest, it is key to clean the tissue prior to analysis as well as possible, and also during sample preparation take care not to contaminate the sample under investigation.

III-B-4. Collagen-based animal identification and classification

The field that identifies animal bones with mass spectrometry is called 'Zooarchaeology by mass spectrometry' (ZooMS). Collagen is an ideal protein to study with ZooMS, because it is highly stable over time [219] and is highly conserved between different species; a higher level of

conservation is observed in more closely related species. Welker et al. showed the clustering of South American animals based on COL1A1 and COL1A2 primary amino acid sequences [190]. Buckley et al. showed similar results for 47 different mammals and 4 types of birds [196]. The analysis described by Buckley et al. only requires 1 mg of bone tissue, which is demineralized with hydrochloric acid, and digested with trypsin. Digestion was followed by a clean-up step done with a C18 ZipTip. The purified peptides were analyzed with a MALDI-TOF and eventually 92 collagen peptides were found to sufficiently distinguish between the different animals. Another publication by Buckley and co-workers shows the possibility to distinguish between 44 different species and sub-species of sheep and goats [227]. Collagen peptides were collected by incubating bone with hydrochloric acid, followed by SPE separation and measurement with a MALDI-TOF.

This principle can also be used to identify the animal source(s) of, for example, gelatin [228, 229], leather [230], and meat. Zhang et al. demonstrated the possibility to detect unique collagen peptides in a trypsin digested gelatin mixture of bovine and porcine [228]. Zhang et al. demonstrated that difficulties can arise in the identification collagen primary structures because Gly-Ala-4Hyp is a structural isomer of Gly-Ser-Pro; therefore they recommend manual inspection of the MS^2 spectra. Grundy et al. compared the identification of species with mass spectrometry with current PCR and ELISA techniques and showed that mass spectrometry was comparable to ELISA; ELISA and mass spectrometry both outperformed PCR [229].

III-C. CIRCULAR DICHROISM SPECTROSCOPY

III-C-1. Circular Dichroism Spectroscopy

Collagen has characteristic properties with respect to refraction and absorption of visible light. First, collagen is birefringence, which means that the refraction index of light of the same wavelength depends on the polarization direction. Second, collagen shows circular dichroism, which means that the absorption of light of a certain wavelength is dependent on the polarization. Both properties are characteristic for the triple helix. Circular dichroism spectroscopy (CD) is used to study secondary structures of proteins by the detection of the difference in absorbance between left- and right-handed circular polarized light.

Furthermore, CD is commonly used to study the Tm of collagen triple helices. When the triple helical structure is lost, circular dichroism also disappears. Therefore, (un)folding of a triple-helix can be monitored over a temperature gradient. However, triple helix (un)folding does not occur instantaneously but takes place over a temperature range in solution. The (un)folding Tm can also be influenced by solvents, salts, collagen concentration, amino acid sequence, and PTMs. We recommend for more detailed information the article by Mizuno et al. on the effect on the Tm by different temperature gradients and collagen concentrations

[231] and that by Woodlock et al. regarding the effect on the Tm by the pH and different concentrations of various salts [232].

The Tm of specific tripeptides or the effect of PTMs of interest can be tested with 'guest peptides'. Guest peptides are short synthetic peptides that are used to test a hypothesis. Because the experimental design differs slightly among the various publications [48, 82, 233], only relative comparisons of the Tm can be made. Besides analyzing guest peptides, full triple helices can also be analyzed [61, 62].

Ramshaw et al. have analyzed the Tm of 34 different tripeptides [48], which might be just a small step in light of the 400 possible tripeptides. Nevertheless, these 34 different tripeptides cover roughly 50 % of the helical region. The most stable tripeptides contain amino acids capable of ion-ion interactions. Furthermore, the position (Xaa or Yaa) of an amino acid influences the ability to form hydrogen bounds and interaction with the peptide backbone [234-237]. This positional effect can result in Tm differences of 25°C and can cause locally micro-unfolding. Tripeptides with a high stability are, in general, more numerous than tripeptides with a low stability [48, 238, 239].

III-C-2. Possible mass spectrometry alternatives for CD

Circular dichroism spectroscopy is a well-established technique to determine the Tm of collagen as a whole but also of small guest peptides to study specific changes in the Tm as a function of change in the amino acid sequence. These experiments can also be performed with hydrogen/deuterium exchange mass spectrometry (HDX-MS). For the reader interested in HDX-MS we recommend the review by Masson et al. [240].

HDX-MS is also suitable to detect collagen micro-unfolding, which is supposed to occur at specific locations in collagen [241-243]. Collagen micro-unfoldings in the triple helix can occur as a consequence of a local low T_m in the helix. At those micro-unfoldings, hydrogens might be exchanged for deuterium; by subsequent lowering of the Tm, the micro-unfoldings are closed and the deuterium will be captured. The captured deuterium causes a mass shift which can be measured with mass spectroscopy.

III-D. STAINING

III-D-1. Picrosirius red

Apart from the difference in polarized light absorption, collagen triple helices are also birefringent. Birefringence is the optical property of a compound whereby the refractive index depends on the polarization direction. With polarization microscopy, collagen can be studied in tissue sections. The collagen becomes better visible with an increased fiber

thickness, especially for collagen types I and type III. Picrosirius red staining can enhance the birefringent properties of collagen fibers. This technique has been known since 1979 [244] and is often used in the pathological analysis of tissue. This staining technique is highly specific for collagen; the staining itself is not specific for collagen, but the synergistic effect of combining it with birefringence makes it so. The compound that enhances birefringence is 'direct red 80' (see figure 11). The acidic groups of this compound bind with basic groups of collagen and align along the fibers and this binding enhances the birefringence properties of collagen. 'Direct red 80' has a high extinction coefficient that allows detecting collagen levels as low as 5.76 µg collagen/mg protein in tissue [245]. The observed color depends mainly on the fiber thickness and ranges from green (< 0.8 µm) to yellow, orange and red (1.6 - 2.4 µm); also fiber packing density and alignment of the fibers influence the color (from green to red) [246].

In addition to polarized light, picrosirius red staining can also be visualized with fluorescence microscopy, which is more sensitive, but does not show differences in fiber thicknesses [247].

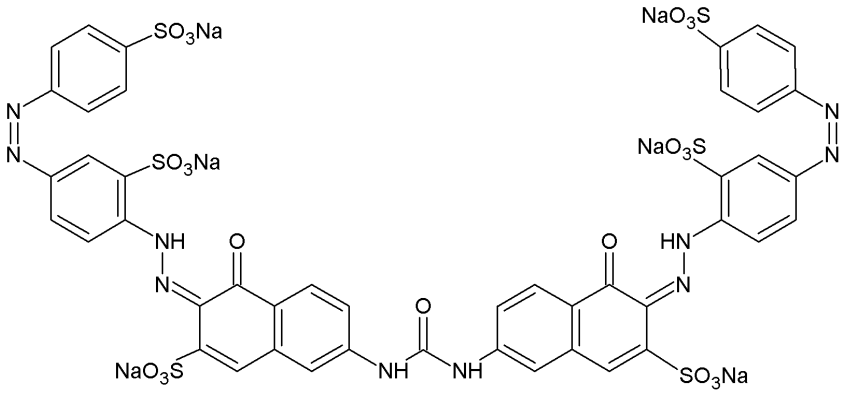


Figure 11. Direct red 80, active compound in picrosirius red.

III-D-2. Other staining techniques

Apart from picrosirius red staining, other techniques can be applied to stain collagen. For example, Masson's trichrome stains collagen blue, nuclei black, and cytoplasm red. Van Gieson's picrofuchsin stains several tissue types, muscles and cytoplasm yellow-brownish, collagen red, red blood cells yellow, and cartilage pink.

In theory, a specific and sensitive staining technique is immunohistochemistry (IHC), which makes use of an antibody that binds to the protein of interest. Karagiannis et al. determined the proteomic signature of the desmoplastic invasion front during colorectal cancer

metastasis [248]. They observed a large difference in the expression pattern of collagen type XII and that of type III (figure 12).

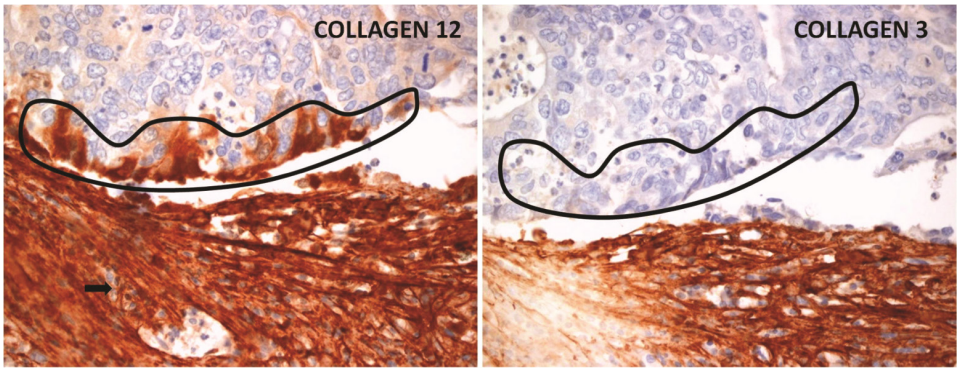


Figure 12. Immunohistochemically staining of collagen type XII (left) and collagen type III (right) at the invasion front of colorectal cancer. The black arrow in the left-hand panel indicates tumor budding; the demarcated cells line the invasion front.[248] Reprinted with permission from the authors, 2012 Impactjournals.

III-D-3. Mass spectrometry – an alternative for staining?

The advantage of (IHC) tissue staining over mass spectrometry analysis is preservation of the spatial distribution. This allows determining co-localization and observing tissue morphology changes beyond the protein(s) of interest. Still, specific sample processing methods and desorption/ionization methods make mass spectrometry competitive with (immunohistochemistry) staining.

Selected subsections of tissue can be collected with laser-capture microdissection (LCM). These subsections can be as small as individual cells. The analysis of these tissue subsections makes it possible to select specific cell types, or stroma components and to identify with mass spectrometry more than 3,400 proteins [249]. The sample processing after LCM is similar to normal bottom-up proteomics [250]. Clair et al. described a method whereby tissue is selected with LCM, followed by sonication, and online trypsin digestion followed by mass spectrometry analysis, which only takes 50 minutes, in comparison to almost two workdays when overnight trypsin digestion is performed. Prior to LCM, staining of tissue is still necessary to select the area of interest, and assessment by a pathologist or an experienced researcher is needed to select the cells or structures of interest.

In MALDI imaging mass spectrometry (MALDI-IMS), direct analysis of the tissue of interest is

feasible without the need to stain, although staining is compatible with mass spectrometric analysis of collagen [251, 252]. Collagen can be detected in both FFPE tissue [253] and fresh frozen tissue [254]. MALDI-IMS is strongly limited by the low number of analytes measured per pixel and by the absence of separation techniques (such as liquid chromatography). It has been suggested to use MALDI-IMS to characterize tissue (molecular histology) and to select interesting areas with LCM [255] for further in-depth identification and analysis. Dilillo et al. described the analysis of tissue by MALDI-IMS on a polyethylene naphthalate (PEN) slide followed by LCM from the same PEN slide. In the MALDI-IMS measurement, the spatial location is determined by the various masses of molecules, and the identification is based on in parallel performed LCM followed by LC-MS/MS measurements in consecutive sections. To detect collagen in fresh frozen tissue, a tissue slice is placed on a glass slide, followed by drying, lipid removal, and subsequent matrix deposition. Extra steps are required for the analysis of FFPE with MALDI-IMS. Cazares et al. describe a heat-induced antigen retrieval step followed by in situ tryptic digestion.

These limitations do not inhibit the analysis of collagen with MALDI-IMS; see table 7 for an overview of collagen peptides observed with MALDI-IMS. Decellularization of tissue even further increases the number of ECM proteins and especially collagen identifications [192]. Direct identification of peptides with MALDI-IMS is not common practice; often tissue is also analyzed in parallel with LC-MS/MS to provide a list of peptides with accurate masses that are matched to the masses detected with MALDI-IMS. For the reader interested in the application of MALDI-imaging in pathology, the following review is recommended [256].

Apart from collagen, also collagen turnover related proteins are visible with MALDI-imaging. An example is MMP-11, which was validated by analysis of the zinc location in the tissue with laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) [257] which allows elemental imaging. Luptakova et al. analyzed with LA-ICP-MS the Na^+/K^+ ratio in rat brains [258].

Table 7. Overview of MALDI-Imaging mass spectrometry analysis of collagen.

Collagen type	Number of peptides	Tissue	Reference
COL6A3	1	Stenotic aortic valve (Human)	[254]
COL1A1	30	Kidney (mouse)	[192]
COL1A2	25	Decellularized	
COL3A1	20		
COL4A1	7		
COL4A2	14		
COL4A3	2		
COL4A4	2		
COL4A5	2		
COL5A1	5		
COL5A2	2		
COL6A1	12		
COL6A2	9		
COL6A3	33		
COL12A1	2		
COL1A1	1	Lymph node (Human)	[259]
COL6A2	1		
COL6A3	Whole protein	Cardiac tissue	[260]

PubMed was systematically searched with the following queries:

1) ("MALDI imaging") AND "collagen" (n=4 publications) = ("MALDI imaging mass spectrometry") AND "collagen" = ("MALDI ims") AND "collagen"
2) ("collagen") AND "imaging mass spectrometry" (n=9 publications)
3) ("MALDI imaging") AND "extracellular matrix" (n=6 publications)

The different search queries retrieved a number of overlapping publications; furthermore, not all articles covered imaging of collagen.

IV. CONCLUSIONS AND FUTURE PERSPECTIVES

Mass spectrometry is a very versatile technique that allows the study of the collagen amino acid sequence, PTMs, interactions with other proteins, as well as the qualitative analysis of collagen. In addition, differences in collagen types can be studied, but also differences in stoichiometry, and quantitative analysis.

Mass spectrometry can be applied in a wide variety of tissues and body fluids. Analyzing the characteristics and localizations of collagen PTMs remains technically challenging. However, the continuing improvements in hardware (increase in sensitivity, and new fragmentation techniques) and developments in sample preparation (new enzymes) will undoubtedly improve the characterization and localization of PTMs.

The function of 4Hyp in alpha helices is well understood, but little is known about, for example, the in vivo hydroxylation rate and whether certain hydroxylation patterns are more prevalent. These knowledge gaps can potentially be filled through the use of mass

spectrometry. Likewise, more knowledge can be obtained about less-common PTMs, such as 3Hyp, which may be involved in the alignment of supramolecular structures, and which so far have hardly been studied with mass spectrometry. In addition, the substrate of prolyl 3-hydroxylase could be further studied by synthesis of peptides, similar to the experimental approaches used for prolyl 4-hydroxylase and lysyl hydroxylase. The degree of hydroxylation could be addressed with mass spectrometry using a targeted assay. 5-Hydroxylysine can be O-glycosylated with galactose, or glucose-galactose. Hardly anything is known about collagen N-glycosylation, and mass spectrometry is capable of identifying glycans linked to collagen. The function of collagen O-glycosylation is largely unknown; it is assumed to affect fiber thickness and collagen cross-linking. The amino group of the side chain of lysine and 5-hydroxylysine can be oxidized into a very reactive aldehyde functionality which initiates collagen cross-linking. Collagen PTM analysis is up to now based on a mixture of different triple helices. The PTM analysis of individual collagen triple helices is still a technical challenge.

The way in which a change in the primary amino acid sequence or the addition of PTMs effects the Tm can be analyzed with CD. Both variations have been shown to strongly influence the Tm; up to 15°C differences are described. The difference in Tm between different tripeptides allows the local unfolding of the triple helix, thereby making it accessible for enzymes. These experiments also showed that a single mutation in the amino acid sequence can prevent the folding or results in the misfolding of a triple helix, which is the case in e.g. osteogenesis imperfecta and Ehlers-Danlos. The location of the mutation influences the severity of the effect on the folding. The data obtained with CD should only be compared in a relative way, considering the different experimental settings (e.g., heating/cooling rate, guest peptide composition). As alternative to CD, which by now is a well-established technique and has proven its value, hydrogen/deuterium exchange in combination with mass spectrometry is a most interesting alternative.

Staining of collagen is an easy way to quickly assess collagen expression and possible pathological differences in tissues. Among other things, picrosirius red has been used for nonspecific staining of fibrous collagen types, of which mainly collagen types I and III will be stained. Specific collagen types have been stained with immunohistochemistry. Replacement of immunohistochemistry by MALDI-imaging mass spectrometry has been a topic of interest for the last two decades. The possibility to detect a large number of molecules simultaneously with good spatial resolution is intriguing. MALDI-imaging mass spectrometry has shown to be able to measure collagen peptides in tissue, which allows the imaging of, for example, fiber composition and distribution of hydroxylation patterns in tissue.

The role of collagen in healthy tissue is of great interest, especially to better understand diseases where collagen is defective or lacking. Collagen has various biological functions that are crucial for the normal functions of tissue, such as: handling of physical stress, cell-cell and cell-matrix interaction, cell signaling, cell migration, hydration of tissue, and extracellular matrix tension by regulation of collagen fiber thickness. Inhibition of the normal functioning of collagen will often lead to a disease. Besides the diseases described in table 5, collagen is also involved in tumor growth and metastasis. During tumor growth, the invasion of healthy tissue by tumor cells results in a remodeling of the locally produced ECM. Tumors produce MMPs which can cleave collagen and other ECM proteins. Especially collagen types I, III, and IV provide a barrier which could prevent tumor invasion. During the cleavage of collagen, degradation products are produced which could be of interest for the detection of tumors. A better understanding of the complex modifications of collagen will give more insight in the various functions of collagen in the normal and diseased states.

V. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Netherlands Organization for Scientific Research NWO (Domain Applied and Engineering Sciences; Perspectiefprogramma P12-04, project 13541) and the Dutch Biomarker Development Center (BDC; http://www.biomarkerdevelopmentcenter.nl/) for support of this work.

CHAPTER 3

Identification of a collagen marker in urine improves the detection of colorectal liver metastases

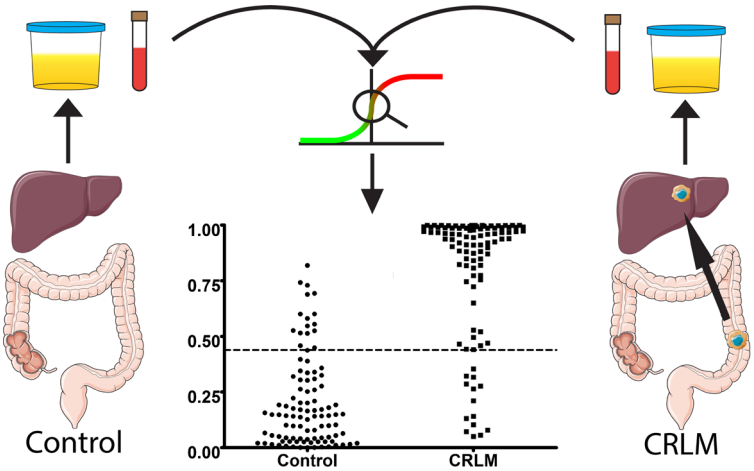
Nick A. van Huizen[1,2], Joost van Rosmalen[3], Lennard J.M. Dekker[2], Robert R.J. Coebergh van den Braak[1], Cornelis Verhoef[1], Jan N.M. IJzermans[1], and Theo M. Luider[2*]

[1]Department of Surgery, [2]Department of Neurology, [3]Department of Biostatistics, Erasmus University Medical Center, Rotterdam, the Netherlands. Mailing address for all authors: P.O. Box1738, 3015 GE Rotterdam, The Netherlands.

Under Review: Journal of Proteome Research (2019)

ABSTRACT

Previously, we reported a combination of an urine collagen alpha-1(I) natural occurring peptide (NOP) AGPP(-OH)GEAGKP(-OH)GEQGVP(-OH)GDLGAP(-OH)GP (AGP) and serum carcinoembryonic antigen (CEA) to have the potential to detect colorectal liver metastasis (CRLM). The combined method requires further adaption for better sensitivity and specificity prior to clinical implementation. This mass spectrometry study aimed to identify additional collagen NOPs in urine and determine the most discriminating NOP panel. We improved the combined method on the basis of analysis of urine samples from 100 healthy controls and 100 CRLM patients. Two additional NOPs were identified: **GPP**GEAGK(-OH)P(-OH)GEQGVP(-OH)GDLGAP(-OH)GP (**GPP**), collagen alpha-1(I), and **GND**GARGSDGQPGPP(-OH)GP(-OH)P(-OH)GTAGFP(-OH)GSP(-OH)GAK(-OH)GEVGP (**GND**), collagen alpha-1(III). A molecular model combining NOPs (AGP, GPP, and GND) and CEA was generated. Molecules that did not contribute significantly were removed, resulting in a model consisting of GND and CEA. With this model, 88% sensitivity and 88% specificity were reached in the discovery set, and 100% sensitivity and 75% specificity in the validation set (control, n= 12; CRLM, n = 10). The AUC of the ROC curve is significantly higher than the current model based on AGP and CEA (p=3.3*10-4The new model performs better than the currently used techniques in the clinic that have a 57-70% sensitivity and a 90-96% specificity.



INTRODUCTION

Colorectal cancer is the third most often diagnosed cancer in the Netherlands. Between 2010 and 2017, annually 10,000-16,000 new patients were diagnosed and 5000 patients died.[261] In the Western world, the probability that a patient will develop metastases in the liver (colorectal liver metastases; CRLM) [262] after curative surgery of the primary tumor is 20-40% [2-5]. After curative surgery, a patient is offered an intensive 5-year follow-up program, which consists of regular computed tomography (CT) scans, ultrasound studies, and carcinoembryonic antigen (CEA) serum measurements.[5, 7, 8] The availability of a robust alternative test could strongly reduce the number of procedures in the follow-up period.

We previously reported that a combination of serum CEA and a collagen natural occurring peptide (NOP) in urine is very promising in detecting CRLM (sensitivity 85%, specificity 84%). [12] Even though the sensitivity and specificity are better than in the currently used techniques, these are not yet sufficient for clinical purposes.[12] In the current study we performed validation testing on a cohort of 100 patients and 100 controls. Urine and blood samples are still required for this combined method. These samples can be collected at home or at the general practitioner's practice. This could reduce the number of hospital visits for CT-scans, etcetera.

The aim of this study was to increase the sensitivity and specificity of our previously published method by searching for the most discriminating NOP panel in urine. The main focus was on NOPs originating from collagen, since collagen is more strongly upregulated in CRLM tissue than it is in normal liver tissue.[26] A large proportion of the NOPs present in urine originates from collagen, [11] which feature makes urine an ideal non-invasive sample. In addition, an enzymatic step is not needed for this type of molecules and a reason why sample preparation is relatively easy and more accessible for clinical use, in contrast to proteins.

MATERIALS AND METHODS

EXPERIMENTAL DESIGN AND STATISTICAL RATIONALE

This study was approved by the Erasmus MC ethics review board (MEC-2008-062) and was performed according to the Declaration of Helsinki. Urine samples of healthy kidney donors (controls) and CRLM patients were measured alternately with mass spectrometry.

The identification of new collagen NOPs in urine was based on the identification of all NOPs in urine (discovery set 1: controls, n=40; CRLM, n=40). In a previous study, a sample

size of 25 samples per group proved sufficient to identify peptide based markers in bottom-up proteomics in tissue.[26] However, in urine the observed differences in NOP levels are smaller. The mean and standard deviations (SD) used for the power analysis (alpha = 0.05, beta = 0.20) were calculated from the overall data of log-transformed significant upregulated collagen peptides in urine samples of five CRLM patients and five control patients (control mean = 6.76, CRLM mean = 6.98, SD_{pooled} = 0.75). The power analysis resulted in a sample size of 184 samples per group. The bottom-up analysis (untargeted) is not based on a sufficiently powered data set, however, by performing targeted analysis on NOPs of interest, we will verify the observed significance.

3

Targeted analysis on NOPs of interest was performed on discovery set 1 and on an additional urine sample set (discovery set 2: control, n=60; CRLM, n=60). The discovery sets 1 and 2 are described in Lalmahomed et al.[12] Validation was performed on independently collected urine samples (control, n=12; CRLM, n=10).[11] To verify that the sample size of the validation set is sufficient, a power analysis was performed based on the targeted analysis of the full discovery set data of the log-odds (control mean = -2.39, CRLM mean = 3.83, SD_{pooled} = 2.98). The power analysis (alpha = 0.05, beta = 0.20) resulted in a sample size of 4 samples per group, indicating that the validation set (targeted approach) is well powered. A flow chart of the samples used is shown in Figure 1.

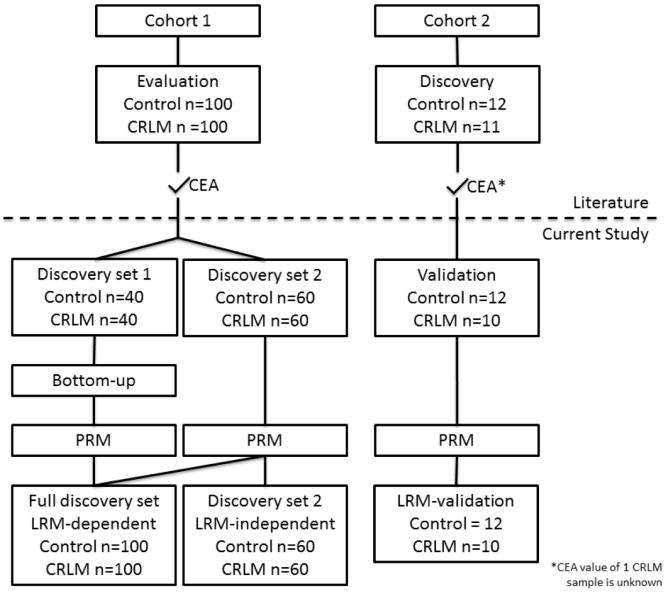


Figure 1. Flowchart of the samples used in this study, originating from Lalmahomed et al. (cohort 1) and Broker et al (cohort 2).[11, 12]

Bottom-up proteomics was used to identify new NOPs. Replicate measurements were not performed as the biological variation in these kind of samples is much larger than the technical variation.[263] Assessment of the number of significant NOPs by chance in the bottom-up proteomics data was determined by permutation testing.

The three most significant NOPs associated with the three most abundant collagen alpha chains that are also more strongly upregulated in CRLM tissue than in healthy liver tissue were selected.[26] As bottom-up proteomics is a semi-quantitative technique, a targeted quantitative mass spectrometry method (parallel reaction monitoring, PRM) was developed to validate these findings.

The developed PRM method is conform tier 3 (of 3 levels) of analytical assay validation [264], which implies that the assay is a targeted discovery assay. The PRM method was applied on both the full discovery set and the validation set. To determine the best model, a logistic regression model (LRM) was fit, containing the NOPs (referred to by the three letter code) and CEA. The optimal LRM was fit by backward elimination of predictors from the LRM that contained all molecular markers (AGP, GND, GPP, and CEA). The optimal model was validated on the validation set. The statistical analysis was performed on the discovery set 2, and on the combined discovery sets 1 and 2 (full discovery set). However, combining the discovery set 1 with discovery set 2 created a dependent data set, because discovery set 1 was already used for bottom-up proteomics. Yet, combining generates a higher statistical power. After measuring all samples, and optimizing the LRM as described above, we selected three samples with low, medium, and high levels of the predictors present in the optimal-LRM. Because SIL peptides were not available for all predictors, these three samples were processed five times to get an estimate of the reproducibility.

CHEMICALS

Ultra-high pressure liquid chromatography grade solvents were obtained from Biosolve (Valkenswaard, the Netherlands). A stable isotope labeled (SIL) peptide was obtained for AGPP(-OH)GEAGK(SIL)P(-OH)GEQGVP(-OH)GDLGAP(-OH)GP from Pepscan (Lelystad, the Netherlands), the lysine is labelled with $^{13}C_6{}^{15}N_2$. This SIL peptide was characterized using HPLC-UV and ESI-MS. These three urine NOPs will be abbreviated by the first three amino acids (AGP, GPP, and GND, respectively). All other chemicals were obtained from Sigma-Aldrich (Zwijndrecht, the Netherlands).

SAMPLE SELECTION

Samples of the cohorts described in the studies of Lalmahomed et al. [12] and Broker et al. [11] were reanalyzed (figure 1). Samples of cohorts 1 and 2, were, after collection, stored at -80°C in

polypropylene tubes. One CRLM sample was excluded from the validation set of the current study because the corresponding CEA value was not known. As CEA levels for the validation set of Broker et al. study are not known, this set of samples was excluded.

Age and BMI differences between controls and CRLM patients were calculated with a Wilcoxon rank sum test , and differences in gender and serum creatinine levels above 115 µM/L with a chi-square test. A two-sided p-value below 0.05/4 = 0.0125 (Bonferroni correction to correct for multiple testing) was considered as significant.

SAMPLE PREPARATION

NOPs for bottom-up proteomics and targeted mass spectrometry were isolated from urine as described by Lalmahomed et al.[12] In brief, NOPs were separated from small molecules, salts, and proteins with a mRP C-18 Hi-Recovery Protein Column (4.6 × 50 mm) (Agilent, Amstelveen, the Netherlands) installed in an Ultimate 300 LC system (Dionex, Amsterdam, the Netherlands) equipped with an online-fractionator. After separation, the NOP fraction was collected, dried, reconstituted, and analyzed with mass spectrometry. The same protocol was used to prepare urine samples for bottom-up proteomics as for targeted proteomics measurement.

BOTTOM-UP PROTEOMIC

For the identification of NOPs we applied a standard bottom-up LC-MS/MS method as described by van Huizen et al.[26] In short, an Ultimate 3000 nano RSLC system (Thermo Fischer Scientific, Germering, Germany) was coupled online to an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fischer Scientific, San Jose, CA, USA). Injected samples were trapped and washed on a trap column (C18 PepMap, 300 µm ID x 5 mm, 2 µm particles size, 100 Å pore size; Thermo Fisher Scientific, the Netherlands). After washing, the trap column was switched in line with an analytical column (PepMap C18, 75 µm ID x 250 mm, 2 µm particle size, 100 Å pore size; Thermo Fisher Scientific, the Netherlands) for peptide separation prior to mass spectrometry analysis. We deviated from the protocol described by van Huizen et al. in that prior to mass spectrometry analysis, neither the samples were analyzed on a test HPLC-system, nor the injection volume was normalized. For every sample, a volume of 2 µL was injected. Bottom-up proteomics data was uploaded to the PRIDE archive (PXD013533).

ANALYSIS OF BOTTOM-UP DATA

MGF peak list files were extracted from raw files by ProteoWizard (v3.0.9166). MGF peak list files were searched using the Mascot search engine (v2.3.2, Matrix Science Inc., London, UK) and the UniProt/SwissProt database (20194 entries). The following settings were used for the database search: enzyme was set to open because we analyzed NOPs; the mass tolerance was set to 10

ppm for peptide mass and 0.5 Da for fragment mass. As variable modification hydroxylation of proline, lysine, and oxidation of methionine was selected (+16 Da); no fixed modifications were added. MASCOT identifications were imported into Scaffold (v4.6.2, Portland, OR, USA). In Scaffold, protein confidence levels were set to 1% false discovery rate (FDR), at least 2 peptides per protein, and a 1% FDR at peptide level. FDRs were estimated by inclusion of a decoy database search generated by MASCOT. Raw files were aligned and combined with the identification list exported from Scaffold (supplement S-1) in Progenesis QI (v4, Nonlinear Dynamics, Newcastle-upon-Tyne, United Kingdom) followed by exporting the normalized abundance to Excel 2010 (Microsoft, Redmon, WA, USA). Duplicate feature intensities were summed. Data was further processed with Excel, GraphPad Prism (v5.01, La Jolla, CA, USA), and R (v3.3.1, Vienna, Austria). Prior to ^{10}log-transformation, a value of '10' was added, with the aim to include missing values for further data analysis. After ^{10}log-transformation the data was assumed to be normally distributed. Differences in NOPs between between controls and CRLM patients were tested with a Mann-Whitney test . Two-sided p-values below 0.05 were considered significant.

Only collagen alpha chains were taken into account, which also were found to differ between CRLM tissue and normal liver tissue.[26] A NOP molecular panel was constructed consisting of nine NPOs, i.e., the three most significant NOPs from the top three most abundant collagen alpha chains. In addition to these nine NOPs, the NOP named AGP, reported in ref [11, 12], was included in the targeted mass spectrometry method.

Permutation testing was performed according to the R-script published as supplemental file by van Huizen et al. [26]. In short, the data was randomly divided in two groups at the peptide level; significant differences between the two groups were determined using the Mann-Whitney test. The ^{10}log was taken per permutation iteration of the total number of significant differences (p-value < 0.05). The distribution of the ^{10}log number of significant p-values was assumed to be normal. The difference was assumed to be significant if the true dataset value was greater than the average value of the permutation test plus twice the SD ($p<0.05$).

TARGETED MASS SPECTROMETRY ANALYSIS

Targeted mass spectrometry measurements were performed on the same nanoLC-ESI-Orbitrap Lumos Fusion as used for the bottom-up proteomics. To measure the samples, a PRM method with optimized collision energies was developed. NOPs for which no optimal collision energy could be determined or which had too low signal intensities for identification were excluded. A table listing, among other characteristics, the optimized collision energies is available in the supplement S-2.

The full discovery set and the validation set were measured at different times to increase validity. The control groups of all data sets were normalized to the control group of the discovery set 1. The same correction factor was applied to the CRLM samples. This was only necessary for NOPs for which no SIL peptides were available (GPP and GND). For AGP a SIL peptide was present and a normalization was not applied. Targeted mass spectrometry data was uploaded to the PRIDE archive (PXD013705).

ANALYSIS OF TARGETED DATA

Raw files produced by the mass spectrometer were imported into Skyline.[265] Per peptide, we selected a maximum of five transitions with a high intensity, and no obvious interference of neighboring peaks. The GND and GPP peptide peak areas, the sum of all selected transitions, from Skyline were used and for AGP a ratio with the SIL peptide was used.

LOGISTIC REGRESSION MODEL

Statistical analyses were performed in R (version 3.3.1, Vienna, Austria).[266] The complete R-script is available in the supplement S-3. The data used is provided in supplement S-4. The predictor selection (backwards elimination) was applied separately on the discovery set 2 (independent data set) and the full discovery set (dependent data set). If the data set (full discovery set or discovery set 2) used to select predictors did not show a different predictor selection, then the analysis was performed with the full discovery set to prevent a loss of power. To select relevant predictors via a backward elimination process to fit the optimal logistic regression model, a significance level of 0.05 was used. The critical p-value was Bonferroni corrected for the number of predictors or comparisons tested.

The current molecular panel consists of AGP and CEA, which was extended with the newly identified NOPs (GPP and GND). To fit a new model with the molecular markers, these markers were tested on any relationship between patient characteristics, individual significance, and multicollinearity. The relationship between the patient characteristics 'age', 'gender', 'BMI', 'urine creatinine', 'serum creatinine > 115 µM/L' was determined by fitting a linear model that predicts an individual molecular marker per patient characteristic and the predictor 'group (healthy/sick)'. Molecular markers that were significantly correlated with a patient characteristic were excluded from further analysis. All remaining individual predictors were tested for significance by fitting a LRM with the individual predictors. Significance of an individual predictor was based on Wald statistics. The selected significant predictors were assessed for multicollinearity by calculating the variance inflation factor (VIF). Multicollinearity was assumed to be present with a VIF above 10; if necessary predictors were discarded to prevent multicollinearity.

The selected predictors were fit into a combined LRM (full-LRM). The optimal LRM (optimal-LRM) was formed by backward elimination of non-significant predictors from the full-LRM. The relation between the molecular markers in the optimal-LRM and the size of the largest tumor, as well as, the number of tumors were tested by fitting a linear model. Significance of an individual predictor was based on Wald statistics.

The Cook's distance test was used to inspect the data for outliers and/or leverage points. The threshold for a point to be suspected of being an outlier/leverage point was calculated with the formula 4/(n-k-1), whereby n= number of samples, k= number of predictors. Outliers and/or leverage points identified by manual inspection of the samples were removed from the data set.

Our previous logistic regression model (old-LRM) contained AGP and CEA. Prior to comparison of the old-LRM and the optimal-LRM, a Pearson correlation was calculated. To select the LRM with the highest predictive power, the performance of the optimal-LRM needed to be compared to that of the old-LRM. The predictive power was compared with the 'anova' function if the models were nested, otherwise the DeLong's test was used to compare AUCs.

RESULTS

PATIENT CHARACTERISTICS

Table 1 provides an overview of the basic patient characteristics. Age and gender were significantly different between the controls and CRLM patients. A serum creatinine level above 115 µM/L, indicating renal impairment, was measured in four patients. Urine creatinine levels did not significantly differ between urine of control and patients suffering from CRLM.

Table 1. Patient characteristics

Parameter	Full-Discovery			Validation		
	Control	CRLM	p-value	Control	CRLM	p-value
Age (years)	52 [43 - 63]	64 [57 - 70]	<0.001	59 [54 - 65]	66 [65 - 74]	0.04
Male gender, No. (%)	32 (32%)	68 (68%)	<0.001	0 (0%)	0 (0%)	-
BMI	26 [23 - 28]	26 [25 - 28]	0.08	26 [23 - 28]	25 [22 - 28]	0.20
No. of lesions	-	2 [1 - 4]	-	-	2 [1 - 2][1]	-
Size of largest lesion (cm)	-	2.7 [1.8 - 4]	-	-	3.5 [3.0 - 4.5][1]	-
Serum creatinine > 115µM/L	0 (0%)	4 (4%)	0.24	0 (0%)	1 (10%)	0.93
Urine creatinine	9.3 [5.9 - 13]	9.5 [6.1 - 15]	0.22	4.6 [2.8 - 8]	4.3 [2.4 - 6]	0.47

Data are presented as a median with the interquartile range (25th-75th percentile) or n (%)

[1] For one patient the no. of lesions, and size of the largest lesion were not available.

BOTTOM-UP MASS SPECTROMETRY

A total of 1683 NOPs were identified in the discovery set 1, belonging to 175 proteins. The three most common proteins are collagen type1(I) (n=183 NOPs), collagen type-1(III) (n=157 NOPs), and uromodulin (n=84 NOPs). Four hundred and fifty-three NOPs (27%) belong to 13 collagen alpha chains (table 2). Four hundred and six NOPs (24%) were significantly different between controls and CRLM patients, of which 118 belong to collagen (table 2).

Table 2. Number of NOPs identified per collagen alpha chain.

Protein [i]	Peptides	Up [ii] p <0.05 & FC > 1	Down [ii] p <0.05 & FC < 1
Collagen alpha-1(I) chain[iii]	183	30	10
Collagen alpha-2(I) chain[iii]	54	8	2
Collagen alpha-1(II) chain	4	0	0
Collagen alpha-1(III) chain[iii]	157	56	1
Collagen alpha-1(IV) chain	3	1	0
Collagen alpha-5(IV) chain	4	1	0
Collagen alpha-1(V) chain	2	1	0
Collagen alpha-2(V) chain	8	3	0
Collagen alpha-1(VI) chain	3	0	0
Collagen alpha-1(X) chain	3	0	0
Collagen alpha-1(XV) chain	11	0	0
Collagen alpha-1(XVIII) chain	20	1	4
Collagen alpha-1(XXII) chain	1	0	0

[i] Proteins marked in bold are shown to be upregulated in CRLM compared to normal liver tissue.[26]

[ii] The number NOPs up- or downregulated with a p-value below 0.05 and the fold change (FC) above or below 1.

[iii]The three most abundant collagen alpha chains.

TARGETED MASS SPECTROMETRY

The urine NOP panel was constructed by including AGP [11, 12] and the three most significantly different NOPs of the three most abundant collagen alpha chains. The optimal collision energy could not be determined for seven out of ten urine NOPs. The three remaining urine NOPs were **AGP**P(-OH)GEAGKP(-OH)GEQGVP(-OH)GDLGAP(-OH)GP, **GPP**GEAGK(-OH)P(-OH)GEQGVP(-OH)GDLGAP(-OH)GP, and **GND**GARGSDGQPGPP(-OH)GP(-OH)P(-OH)GTAGFP(-OH)GSP(-OH)GAK(-OH)GEVGP. MS/MS spectra confirming the hydroxylation pattern pf GND and GPP are present in the supplemental (S-5), the MS/MS spectra of AGP have already been published.[12] While AGP and GPP originate from collagen alpha chain 1(I), GND originates from collagen alpha chain 1(III). AGP and GPP have an almost identical amino acid sequence, whereby AGP has an additional alanine at the N-terminus. The hydroxylation pattern of AGP and GPP are different, whereby GPP contains a hydroxylysine, and AGP does

not. See supplement S-2 for, among other settings, the optimized collision energies and the selected b- and y-ions.

LOGISTIC REGRESSION MODEL

The predictor selection process was applied on the discovery set 2 and the full discovery set. Prior to fitting the full-LRM, the molecular markers (AGP, GPP, GND, and CEA) were tested for a linear relationship with any of the patient characteristics (age, gender, BMI, urine creatinine, and serum creatinine levels). Significant linear relationships were not found, except for urine creatinine that had a significant relationship with AGP, GPP, and GND. Therefore, urine creatinine is added to the full LRM model. The individual molecular markers and urine creatinine were also tested for individual significance by fitting a univariable LRM per marker. The results are shown in Table 3. Individually, all molecular markers showed to be significant and were included in the full-LRM. There was no multicollinearity present between the molecular markers. Therefore all molecular markers were included into the full-LRM in the full discovery set and in discovery set 2. In the full discovery set neither a significant linear relationship was present between any of the molecular markers individually, nor with any of the molecular markers and size of the largest tumor and number of tumors.

Table 3. Predictor selection, significant predictors are marked in bold

Predictor	Univariable (Disc 2)		Multivariable (Disc 2)		Univariable (Full disc)		Multivariable (Full disc)	
	OR[i]	p-value	OR[i]	p-value	OR[i]	p-value	OR[i]	p-value
AGP	1.7 [1.4-2.2]	**$4.3*10^{-5}$**	2.9 [1.6-6.2]	**2.0*10-3**	1.6 [1.4-2.0]	**$3.1*10^{-7}$**	2.1 [1.4 – 3.4]	**1.310^{*-3}**
GPP	4.0 [1.7-11.3]	0.0039	- [iii]		7.1 [3.1-18]	**$1.3*10^{-5}$**	0.8 [0.3-2.7]	0.73
GND	21.1 [8.0-67]	**$1.5*10^{-8}$**	16 [4.6-72]	**$6.2*10^{-5}$**	20 [9.4-48]	**$3.4*10^{-13}$**	17 [5.3-63]	**$7.9*10^{-6}$**
CEA	12.7 [5.1-38]	**$6.8*10^{-7}$**	43 [9.5-301]	**$1.6*10^{-5}$**	21 [8.8-56]	**$1.3*10^{-10}$**	45 [13-205]	**$5.1*10^{-8}$**
Urine creatinine	1.1 [0.99-1.1]	**0.11**	0.70 [0.6-0.8]		1.1 [1.0 – 1.11]	**0.06**	0.7 [0.6 – 0.9]	**$1.0 *10^{-4}$**

[i] The odds ratio (OR) is presented as the OR and the 95% confidence level

[ii] The critical p-value used is $0.05/19 = 2.7*10^{-3}$

[iii] This molecular marker was individually not significant, and therefore not selected for the full-LRM

The optimal-LRM was formed with backwards elimination of the non-significant predictors based on the multivariate analysis for the full discovery set and the discovery set 2, which are described in table 3. For both data sets, this resulted in a model containing AGP, GND, CEA, and urine creatinine. The predictor selection was irrespective of the use of the full discovery set or discovery set 2. The remaining analyses were, therefore, performed solely with the full discovery set to prevent a loss of statistical power. The final formula formed after backwards elimination to predict the probability of an individual of having CRLM is shown in formula 1. The OR with 95% CI for AGP is 2.1 [1.4-3.4], GND is 15 [2.6-46], CEA is 44 [13-201], and urine

creatinine is 0.7 [0.6-0.9].

Formula 1: Predicted probability of having CRLM

$$\frac{1}{1+e^{-1*(-21.2121+0.7395*AGP+2.6976*GND+3.7889*CEA-0.3028*urine\ creatinine)}}$$

The Cook's distance was calculated to ensure that this formula is not heavily influenced by outliers/leverage points. Seventeen data points were above the threshold and were manually inspected. None appeared to be a wrong measurement, and therefore none was removed. On re-measuring, AGP values from the old-LRM and the new optimal-LRM data sets were highly correlated (correlation = 0.89, p-value < $2.2*10^{-16}$). The estimated linear relationship between the old AGP values (AGP_old) and the current AGP values is: AGP = 0.9 + 1.57*AGP_old. The AUCs of the old-LRM and the optimal-LRM were compared using DeLong's test. The old-LRM had an AUC of 0.8824, which is significantly different from the optimal-LRM AUC of 0.9499 (p-value = $.3*10^{-4}$). A scatter plot containing the values of the optimal-LRM and old-LRM is available in the supplement S-6.

Based on the ROC curve for values calculated by the optimal-LRM, a cut-off value of 0.515 was chosen resulting in a balanced sensitivity and specificity. This cut-off value results in an 88% sensitivity and 88% specificity in the full discovery set, and in the validation set in 75% sensitivity and 100% specificity.

Table 4. Overview of the GND (from targeted analysis of the full discovery set) and CEA values and the obtained sensitivity and specificity.

					Optimal-LRM		Old-LRM	
		AGP[i]	GND[i]	CEA[i]	Sensitivity	Specificity	Sensitivity	Specificity
Discovery	Control	2.5 [1.4-3.4]	6.9 [6.7 - 7.2]	0.3 [0.1 - 0.5]	88%	88%	80%	79%
	CRLM	5.0 [2.4-5.9]	7.7 [7.4 - 8.0]	1.0 [0.6 - 1.4]	(80-93)	(80-93)	(71-87)	(70-86)
Validation	Control	4.3 [1.9-4.3]	6.9 [6.7 - 7.1]	0.2 [0.0 - 0.5]	100%	75%	90%	67%
	CRLM	6.7 [4.4-9.2]	7.5 [7.3 - 7.6]	1.5 [0.6 - 2.1]	(68-100)	(46-91)	(57-100)	(39-86)

[i] Mean [1st quartile – 3rd quartile]

The 95% confidence intervals of the sensitivity and specificity based on the Agresti-Coull interval is presented between brackets.

The McNemar test did not show a significant difference (p = 0.3588).between the Optimal-LRM validation specificity (92%) and the Old-LRM validation specificity (67%), which was the largest difference between the Optimal LRM and the Old-LRM.

To estimate the reproducibility of the sample processing with respect to the GND values, we measured three samples, selected in the lower, middle, and higher range of all measured values, five times. The following samples were measured, with in brackets the ^{10}log of the area and %CV: VMS-248 (low, 6.1 ± 1.4%), VMS-253 (middle, 6.9 ± 1.8%), and VMS-163 (high, 7.6 ± 1.4%).

COMPARISON OF IDENTIFIED URINE NOPs TO PREVIOUS WORK

The NOPs AGP, GPP, and GND were not tryptic peptides. Van Huizen et al. [26] published an overview of the differences between collagen proteins in healthy liver tissue and in CRLM tissue. We compared the tryptic peptides identified from these collagen proteins [26] with the NOPs of interest in this study. Table 5, lists the tryptic peptides from healthy liver tissue and CRLM tissue that had an amino acid sequence overlapping with the NOP GND. Five out of seven peptides having an overlapping amino acid sequence with GND are significantly upregulated in CRLM in comparison to liver tissue. One of these five peptides showed a similar hydroxylation pattern as that of GND, and this tryptic peptide was significantly upregulated in CRLM tissue (fold change 3.67, p-value = $1.2^{*}10^{-8}$) in comparison to liver tissue. For AGP and GPP, two tryptic peptides with an overlapping amino acid sequence were identified in healthy liver tissue and CRLM tissue, but these tryptic peptides had a different hydroxylation pattern. These two have an identical amino acid sequence, GEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPSGAR, whereby hydroxylation occurs at position 20, 21, 27, 33, and 39, and at the other peptide at position 9, 15, 17 20, 21, 26, and 27.

Table 5. Tryptic peptides with an overlapping amino acid sequence with GND in healthy liver tissue and in CRLM tissue.

Location	Fold Change	T-test	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Urine			G	N	D	G	A	R	G	S	D	G	Q	P	G	P	P	G	P	P	G	T	A	G	F	P	G	S	P	G	A	K	G	E	V	G	P														
Tissue	3.67	1.2E-08							G	S	D	G	Q	P	G	P	P	G	P	P	G	T	A	G	F	P	G	S	P	G	A	K																		
Tissue	3.58	5.1E-07							G	S	D	G	Q	P	G	P	P	G	P	P	G	T	A	G	F	P	G	S	P	G	A	K	G	E	V	G	P	A	G	S	P	G	S	N	G	A	P	G	Q	R
Tissue	3.19	2.1E-05							G	S	D	G	Q	P	G	P	P	G	P	P	G	T	A	G	F	P	G	S	P	G	A	K	G	E	V	G	P	A	G	S	P	G	S	N	G	A	P	G	Q	R
Tissue	3.16	2.4E-04							G	S	D	G	Q	P	G	P	P	G	P	P	G	T	A	G	F	P	G	S	P	G	A	K																		
Tissue	3.59	5.5E-04							G	S	D	G	Q	P	G	P	P	G	P	P	G	T	A	G	F	P	G	S	P	G	A	K																		
Tissue	0.95	8.0E-01							G	S	D	G	Q	P	G	P	P	G	P	P	G	T	A	G	F	P	G	S	P	G	A	K																		
Tissue	1.23	3.3E-01																															G	E	V	G	P	A	G	S	P	G	S	N	G	A	P	G	Q	R

The left-hand column shows if the peptide originates from urine, or healthy liver tissue and CRLM tissue. The fold chance and t-test (p-value) show if a tryptic peptide is significantly up- or downregulated in CRLM tumor tissue in comparison to liver tissue.[26] Glycine (G) is marked yellow; hydroxylated proline (P) is marked green; hydroxylated lysine (K) is marked light blue.

DISCUSSION

The collagen NOPs identified in urine of patients with CRLM were strongly upregulated compared to urine from controls. Collagen is part of the extracellular matrix, changes in the extracellular matrix promote tumor proliferation and metastasis.[24, 25, 150] The newly identified collagen NOPs were included in a new logistic regression model that proved to have a significantly higher AUC than the model based on AGP and CEA. The sensitivity increased from 88% in the discovery set to 100% in the validation set; the specificity decreased from 88% to 75%, respectively. The validation set was an independently collected sample set. We conclude that the new model is experimentally validated. We do recommend further independent validation of the new model in larger and global cohorts that also cover differences in ethnicity.

The sensitivity achieved in the validation set with the optimal model is at least 15-20% higher than that of the currently used techniques, ranging from 57% to 70%.[6, 19, 20] The specificity is lower than that of the currently used techniques, ranging from 90% to 96%.[5, 13, 14] Overall, the performance of the new model is better than that of the currently used techniques. This property may be clinically beneficial, because earlier detection of CRLM can be foreseen and reduces health care costs.

The sensitivity and specificity we report here for the combined method (old-LRM) are lower than those reported in our previous study.[12] Nevertheless, the Pearson's correlation test shows that both measurements for AGP are highly correlated (correlation = 0.89, p-value $< 2.2*10^{-16}$). The CEA values have not been remeasured and the values used in the studies of Lalmahomed et al. and Broker et al. were used.[11, 12]

Even though the current measurements on GND and GPP were done without SIL peptides, five consecutive preparations of three different samples showed a high reproducibility (CV < 2%). The use of a SIL peptide will likely make the measurements even more reproducible between different sample sets [89, 90], because now we needed to correct GPP and GND values of the validation set to match the discovery set. This correction was based on mean values of the control samples per set. SIL peptides were not yet commercially available, because incorporating 5-hydroxylysine as a building block for peptide synthesis is rather complicated. This is most likely technically possible, although a chemical synthesis study is required before it can be attempted.

The molecular markers: BMI, age, gender, and serum creatinine levels were not significantly related to the standard patient characteristics tested, even though the age and gender related

to the control samples and CRLM samples in the discovery set were significantly different. Only urine creatinine showed to be related to urine NOPs and was therefore included in the new LRM. The difference in the gender distribution in the discovery set is explained by a higher incidence rate in males[23], The molecular markers were not significantly associated with the patient gender, therefore we consider it acceptable that the validation set only contains females, although it is not ideal. In addition, no significant correlation was found between any of the molecular markers and the size of largest tumor, nor with the number of tumors. This makes it difficult to prove that up- or downregulated NOPs in urine originate from the CRLM present in the liver. It is still likely that the NOPs we selected (AGP, GPP, and GND) originate from CRLM. In addition, the three collagen alpha chains that have the largest number of significantly upregulated NOPs are also more strongly upregulated in CRLM tissue than in healthy liver tissue.[26]

Furthermore, proline and lysine can be hydroxylated in collagen to give collagen a specific fingerprint. For the NOPs AGP and GPP, this hydroxylation fingerprint could not be found in the data of Van Huizen et al. [26]. Still, for GND a tryptic peptide with the same hydroxylation pattern was more strongly upregulated in CRLM tissue than in healthy liver tissue. The finding of the tryptic peptide with a similar hydroxylation pattern as GND makes it more plausible that the measured GND originates from the CRLM tumor.

In conclusion, the new model is a significant improvement of the previously published combined method containing AGP and CEA. In addition, the new model appears to be competitive with the currently used techniques, but needs to be further proven in large global validation sets. This study shows, once more, the high potential of urine markers, whether or not in combination with other markers, to detect colorectal liver metastases.

ACKNOWLEDGEMENTS

This activity has received funding from the EIT Health under grant agreement No 19622. This European body receives support from the Horizon 2020 research and innovation program.

AUTHOR CONTRIBUTIONS

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

CHAPTER 4

Up-regulation of collagen proteins in colorectal liver metastasis compared with normal liver tissue

N.A. van Huizen[1,3], R.R.J. Coebergh van den Braak[1], M. Doukas[2], L.J.M. Dekker[3], J.N.M. IJzermans[1], and T.M. Luider[3*]

[1]Department of Surgery, [2]Department of Pathology, [3]Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands. Mailing address for all authors: P.O. Box1738, 3015 GE Rotterdam, The Netherlands.

The Journal of Biological Chemistry. 2019 Jan 4;294(1):281-289

ABSTRACT

Changes to extracellular matrix (ECM) structures are linked to tumor cell proliferation and metastasis. We previously reported that naturally occurring peptides of collagen type I are elevated in urine of patients with colorectal liver metastasis (CRLM). In the present study, we took an MS-based proteomic approach to identify specific collagen types that are up-regulated in CRLM tissues compared with healthy, adjacent liver tissues from the same patients. We found that 19 of 22 collagen alpha chains are significantly up-regulated ($P < 0.05$) in CRLM tissues compared with the healthy tissues. At least four collagen-alpha chains were absent or had low expression in healthy colon and adjacent tissues, but were highly abundant in both colorectal cancer (CRC) and CRLM tissues. This expression pattern was also observed for six non-collagen colon-specific proteins, two of which (CDH17 and PPP1R1B/DARP-32) had not previously been linked to CRLM. Furthermore, we observed CRLM-associated up-regulation of 16 proteins (out of 20 associated proteins identified) known to be required for collagen synthesis, indicating increased collagen production in CRLM. Immunohistochemistry validated that collagen type XII is significantly up-regulated in CRLM. The results of this study indicate that most collagen isoforms are up-regulated in CRLM compared with healthy tissues, most likely as a result of an increased collagen production in the metastatic cells. Our findings provide further insight into morphological changes in the ECM in CRLM and help explain the finding of tumor metastasis–associated proteins and peptides in urine, suggesting their utility as metastasis biomarkers.

Reprinted with permission from The Journal of Biological Chemistry. 2019 Jan 4;294(1):281-289. Copyright 2019 American Society for Biochemistry and Molecular Biology.

INTRODUCTION

The collagen protein family is well-studied, having 8000+ publications annually over the last five years (based on PubMed index, July 2018). Despite the large number of annual publications, there is still a need to explore collagen and its functions. It is believed that deepening on the understanding of collagen will eventually lead to significant contributions in important fields such as cancer, fibrosis, and regenerative medicine [268].

It is well known that the loss of the normal structure of the extracellular matrix (ECM), including the major component collagen, has been associated with enhanced tumor proliferation [269, 270]. Tumor proliferation changes the function of collagen [271, 272] and it has been shown that these changes are dependent on changes of individual collagen levels, as shown by Nyström et al. for colorectal cancer (CRC) [273]. These changes have been suggested to play a role in chemoresistance and increased cell proliferation becoming a vicious circle. Naba et al. showed differences of ECM proteins in colorectal liver metastasis (CRLM) compared to healthy liver in an underpowered sample set (n=3) [152]. However, a sufficiently powered study regarding the concentration of specific collagen types in CRLM is lacking.

As a result of CRLM, it has been shown that, an upregulation of collagen can be observed in urine [208, 274] and plasma [178]. Increased excretion of collagen in urine and plasma may be explained by a higher rate of tissue remodeling and turnover, as matrix metalloproteinases (MMPs) [159-161] are differently expressed in CRLM and CRC compared to healthy liver and colon tissue, respectively [275-277]. However, the upregulation in CRLM of proteins related to collagen production has not been shown.

We aim to study the changes in collagen induced by CRLM with mass spectrometry in a well powered data set. We determined which specific collagen types are altered, if proteins in the collagen turnover pathway are altered, and if we can define a specific protein footprint of the primary tumor in the metastasis.

RESULTS

PATIENT CHARACTERISTICS

Patients with CRLM had a median age 69.2 years (interquartile range [IQR] 60.5-74.7) and were mostly male (66.7%). Patients had a median of 1 (maximum 7) tumor, being moderately differentiated in 28 out of 30 patients. The CRLM tumor had a median size of 2.4 cm (IQR 1.4-3.5). The primary tumor was located in colon (47%), rectum (43%), and sigmoid (10%). Patients with

CRC had a median overall age of 71.1 years (IQR 54.8-80.9 years). The patients with liver fibrosis (FIB, ranging from moderate to severe fibrosis) had a median overall age of 59 years (IQR 58-59 years) and were all male. These patients developed fibrosis due to a combination of different causes: viral hepatitis (3 out of 5), HCC (5 out of 5), steatosis (1 out of 5), steatohepatitis (2 out of 5). Additional information is found in the supporting information S-4.

DATA ASSESSMENT

Normalization

Normalized values obtained from healthy adjacent liver tissue (control) and CRLM tissue did not differ significantly (P=0.58)

Permutation

In the true data set of control vs. CRLM tissue, 5812 of the 20,635 peptides (28.1%) were tested significantly different. In the permutated data, an average of 631 (3%) peptides was significantly different. The true data set contained more significant peptides than the average of the permutation plus twice the SD (5812> > ~ 631; P< 0.05) of the permutated data.

PROTEOMICS

Top 10 proteins down- and upregulated

In total 2817 proteins were identified in CRLM and control tissue (see supporting information S-8). The 10 proteins with the highest absolute number of down- and upregulated peptides are shown in table 1. Many downregulated proteins were present in liver specific pathways (table 1). Furthermore, seven of the top 10 upregulated proteins in CRLM tissue originated from ECM.

Table 1. Top 10 proteins with the highest absolute number of peptides down- or upregulated in CRLM.

Protein	Gene name	Fold change[i]		# peptides[ii]
		<0	>0	
Downregulated				
Carbamoyl-phosphate synthase [ammonia], mitochondrial	CPS1	135	0	141
Fatty acid synthase	FASN	87	0	91
Spectrin alpha chain, non-erythrocytic 1	SPTAN1	62	0	87
C-1-tetrahydrofolate synthase, cytoplasmic	MTHFD1	55	0	57
Glycogen debranching enzyme	AGL	46	0	46
Cytosolic 10-formyltetrahydrofolate dehydrogenase	ALDH1L1	44	0	48
Aldehyde oxidase	AOX1	40	0	43
Pyruvate carboxylase, mitochondrial	PC	40	1	46
3-ketoacyl-CoA thiolase, mitochondrial	ACAA2	39	0	41
Alcohol dehydrogenase 4	ADH4	39	0	41

Table 1 continued.

Protein	Gene name	Fold change[i]		# peptides[ii]
		<0	>0	
Upregulated				
Collagen alpha-1(I) chain	COL1A1	11	238	286
Collagen alpha-2(I) chain	COL1A2	10	158	189
Collagen alpha-1(III) chain	COL3A1	16	135	210
Fibrillin-1	FBN1	1	99	107
Filamin-A	FLNA	3	90	105
Fibronectin	FN1	1	77	85
Myosin-9	MYH9	2	66	125
Plectin	PLEC	5	50	128
Collagen alpha-1(XII) chain	COL12A1	2	50	58
Collagen alpha-2(V) chain	COL5A1	1	49	50

[i] The number of peptides with a fold change > or < 0 and P<0.05. The fold change is based on ^{10}log values. [ii] Number of identified peptides per protein. Gene names in bold are involved in liver processes (upper half, downregulated), gene names underlined are related to the ECM (lower half, upregulated).

Collagen in CRLM

We then analyzed the number of differentially expressed collagen peptides and collagen alpha chains. In CLRM tissue were 1137 collagen peptides identified, from which 819 peptides (72%) were up- and 55 peptides (4,9%) were downregulated. Overall, we observed an increased collagen level in CRLM vs. control tissue (P<0.0001, linear-fold change 1.67). Furthermore, 19 of the 22 collagen alpha chains were significantly upregulated (table 2). COL12A1 was the most differentially expressed collagen alpha chain (2-fold change, 86% of identified peptides [n=50] significantly upregulated). An unsupervised clustering was performed based on all collagen peptides that separated both groups, except for four samples (CRLM-5, -9, -10, and -25) The four samples (CRLM-8, -12, -19, and -21) containing up to 50% control tissue were evenly clustered within the CRLM group (Figure 1). CRLM-10 had much lower mass spectrometry (MS) intensities for collagen peptides than the rest of the CRLM samples. However, in the normalization process this sample remained within the normal range of samples.

Table 2. Types of collagen identified and the number of peptides significantly up- and downregulated.

Collagen types		# Peptides[i]	Fold change		P-value[iii]	Fold change[iv]
			>0[ii]	<0[ii]		
P02452	COL01A1	286	239	11	$1.84x10^{-09}$	1.7
P08123	COL01A2	189	158	10	$9.4x10^{-10}$	1.7
P02458	COL02A1	5	1	0	$2.4x10^{-03}$	1.2
P02461	COL03A1	210	135	16	$7.6x10^{-06}$	1.5
P02462	COL04A1	33	24	3	$7.0x10^{-07}$	1.5
P08572	COL04A2	33	30	0	$2.92x10^{-12}$	2.0

Table 2 continued.

Collagen types		# Peptides[i]	Fold change >0[ii]	Fold change <0[ii]	P-value[iii]	Fold change[iv]
P53420	COL04A4[v]	1	1	0	5.19×10^{-06}	1.9
P20908	COL05A1	31	30	0	3.15×10^{-14}	2.0
P05997	COL05A2	50	49	1	7.73×10^{-16}	2.3
P25940	COL05A3[v]	1	1	0	6.2×10^{-04}	1.6
P12109	COL06A1	35	10	4	2.8×10^{-02}	1.2
P12110	COL06A2	30	8	1	1.7×10^{-03}	1.3
P12111	COL06A3	96	39	2	6.5×10^{-04}	1.3
Q02388	COL07A1	2	2	0	9.7×10^{-11}	2.2
P27658	COL08A1[v]	1	0	1	1.1×10^{-02}	0.7
Q03692	COL10A1	7	6	0	2.3×10^{-08}	1.7
P12107	COL11A1	5	4	0	3.5×10^{-14}	2.0
Q99715	COL12A1	58	50	2	6.9×10^{-14}	2.0
Q05707	COL14A1	46	29	3	8.8×10^{-07}	1.5
P39059	COL15A1	2	2	0	7.1×10^{-08}	1.8
Q07092	COL16A1	2	1	0	6.7×10^{-01}	1.0
P39060	COL18A1	14	1	1	6.2×10^{-01}	1.0

[i] The number of peptides identified per collagen alpha chain. [ii] The number of peptides with a fold change > or < 0 and P<0.05. The fold change is based on ^{10}log values. [iii] Indicates the P-value at the protein level. [iv] Indicates the true fold change at the protein level (not the ^{10}log value).
[v] These proteins are identified with one peptide and their identification is considered as less reliable compared to proteins identified with 2 or more peptides.

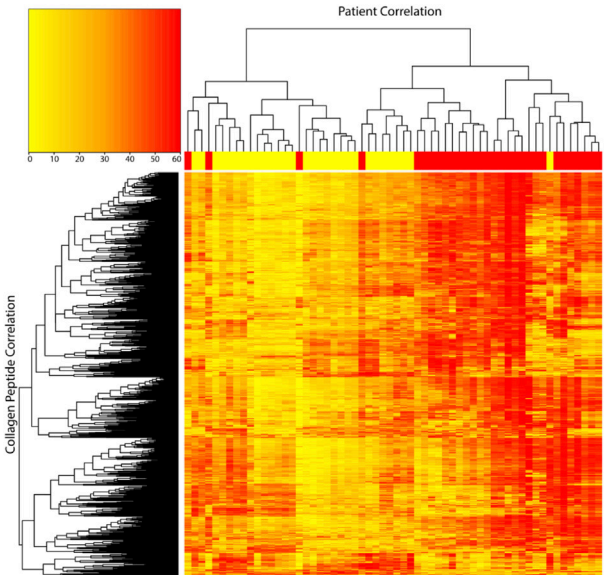


Figure 1. Unsupervised clustering at the patient level (horizontal) and collagen peptide level (vertical). The yellow and red regions at the top indicate control liver and CRLM tissue, respectively. The color scheme in the cluster analysis itself is from yellow-low to red-high.

COLLAGEN TURNOVER RELATED PROTEINS

The observed upregulation of collagen in CRLM may be due to changes in the normal turnover of collagen in the liver. To test this hypothesis, 65 different proteins included in collagen turnover pathways were retrieved from the "Reactome Pathway Knowledgebase" [278, 279]. Twenty-four of these 65 proteins were identified, 17 (~68%) of which were significantly upregulated in CRLM tissue. Sixteen of the 17 upregulated proteins are included in the collagen synthesis and/or fibril forming pathways (table 3). Supporting information S-5 contains addition information.

Table 3. An overview of the identified proteins involved in collagen synthesis and degradation.

Biological role	Protein (Uniprot)	Protein name	# Peptides[i]	Fold change[ii]		# Proteins[iii]	
				>0	<0	P-value	Fold change
B	Q8NBJ5	COLGALT1[iv]	1	1	0	6.6E-05	1.78
B	P14780	MMP9	4	2	0	1.8E-03	1.44
B	Q32P28	P3H1[iv]	1	0	0	1.7E-02	1.42
B	P13674	P4HA1	7	3	1	1.9E-06	1.44
B	Q15113	PCOLCE	4	4	0	4.6E-13	2.22
B	Q15149	PLEC	128	50	5	4.7E-05	1.32
B	Q02809	PLOD1	2	2	0	2.6E-11	2.17
B	O00469	PLOD2[iv]	1	1	0	6.6E-03	1.38
B	O60568	PLOD3	5	3	0	2.8E-06	1.61
B,F	P07237	P4HB	42	1	36	1.6E-18	0.48
B,F	P50454	SERPINH1	13	11	0	7.4E-14	1.94
D	P07339	CTSD	20	0	15	1.2E-05	0.65
D	P08246	ELANE	6	3	0	1.7E-04	1.39
D	P12955	PEPD	4	4	0	6.2E-08	0.59
D	P23284	PPIB	14	0	11	9.4E-08	0.63
F	P25774	CTSS	2	1	0	8.9E-02	1.23
F	Q03001	DST[iv]	1	0	1	1.3E-06	0.50
F	P23229	ITGA6	8	6	0	3.6E-07	1.61
F	P16144	ITGB4	17	13	3	1.4E-09	1.60
F	Q16787	LAMA3	3	3	0	7.4E-09	1.81
F	Q13751	LAMB3	5	5	0	6.0E-11	1.93
F	Q13753	LAMC2	2	1	0	3.2E-05	1.56
F	Q8IUZ5	PHYKPL[iv]	1	0	1	4.9E-12	0.40
F,D	P07858	CTSB	13	2	5	8.5E-02	0.88

The listed proteins are involved in three different pathways, annotated as: F, assembly of collagen fibrils and other multimetric structures; B, collagen biosynthesis and modifying enzymes; D, collagen degradation.

[i] Number of peptides identified per protein. [ii] Number of peptides that are significantly different and have a fold change > or <0. The fold-change is based on ^{10}log values. [iii] P-values calculated over the summed ranked peptides followed by the fold change. The fold change shown is a true value (not a ^{10}log value).

[iv] These proteins are identified with one peptide and their identification is considered as less reliable compared to proteins identified with 2 or more peptides.

CRC PROTEIN SIGNATURE IN CRLM

Next, we studied the presence of a correlation between collagen levels in CRLM and CRC tissue. For COL10A1, COL12A1, COL14A1, and COL15A1, an absence or low number of unique peptides was observed in healthy colon and healthy liver tissue, but a higher number of unique peptides was present in CRC and CRLM tissues (supporting information S-6), being most apparent for COL12A1.

Furthermore, presence of proteins specific for colon tissue in CRLM and control tissues was studied. Elevated expression of 165 genes was found in colon tissue and in CRLM tissue 25 corresponding proteins were identified (table 4). Supporting information S-7 contains additional information. Six proteins matched the set criteria and were considered unique for colon tissue vs. control liver tissue and were found significantly higher in CRLM tissue than control tissue.

IMMUNOHISTOCHEMISTRY

MS data was cross-validated by IHC staining of COL12A1 in CRLM and control tissue. CRLM stroma tissue stained positive for COL12A1 (Figure 2a,b), except for three tissues (CRLM-19, -20, and -28, Figure 2c). Control tissue did not show positive staining (Figure 2d). The three negatively stained CRLM tissue sections did cluster with the rest of the CRLM samples by proteomics (Figure 1). Based on summed peptide ranking the COL12A1 values of the three negative IHC tissues were above 3*SD of the control, and 0, 9, and 18 peptides were respectively identified per sample, indicating that by proteomics these samples can be detected correctly. The IHC staining of COL12A1 is present in the ECM. However, the staining was not equally distributed in the ECM of the CRLM tissue. Furthermore, studying tissue sections to determine IHC scoring revealed distortions in the CRLM morphology compared with the control tissue (Figure 2). In addition, CRC and healthy colon tissue (control C) were stained. CRC tissues showed, similar to CRLM, staining of the ECM for COL12A1. Staining was also observed in the nuclei of epithelial cells in healthy colon tissue by COL12A1 antibody NBP1-88062 (Figure 2g). The other COL12A1 antibody (HPA009143) did not show staining of the epithelial cells, but stained in addition to the collagen present in tumor stroma smooth muscle tissue (Figure 2h). Fibrotic liver tissue (FIB) was stained by IHC for COL12A1 and no positive staining was observed (Figure 2f).

Table 4. Colon-specific proteins are upregulated in CRLM.

Protein	Gene[i]	Fold change[ii]		# peptides[iii]
		<0	>0	
Villin 1	VIL1	1	18	26
Polymeric immunoglobulin receptor	PIGR	2	12	19
Fatty acid binding protein 1	FABP1	17	0	17
Cadherin 17	CDH17	1	11	16
Keratin 20, type I	KRT20	0	11	15
Carbonic anhydrase II	CA2	11	0	12
Galectin-4	LGALS4	0	6	10
Carbonic anhydrase I	CA1	7	1	8
UDP-glucuronosyltransferase 1-8	UGT1A8	7	0	8
Carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM5	0	5	7
IgGFc-binding protein	FCGBP	1	3	7
Meprin A subunit alpha	MEP1A	0	2	7
Glycoprotein A33	GPA33	0	5	6
UDP-glucuronosyltransferase 2B15	UGT2B15	5	0	6
Mucin 13	MUC13	0	5	5
Protein phosphatase 1 regulatory subunit 1B	PPP1R1B/ DARPP32	0	5	5
Epidermal growth factor receptor kinase substrate 8-like protein 3	EPS8L3	1	2	4
Protein FAM3D	FAM3D	0	2	4
Carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	0	3	3
Left-right determination factor 1	LEFTY1	1	1	3
Sulfotransferase family cytosolic 1B member 1	SULT1B1	2	0	3
Carcinoembryonic antigen-related cell adhesion molecule 1	CEACAM1	0	2	2
Carcinoembryonic antigen-related cell adhesion molecule 7	CEACAM7	0	2	2
Protein-arginine deiminase type-2	PADI2	0	1	2
UDP-glucuronosyltransferase 1-10	UGT1A10	2	0	2

These 25 proteins were identified with proteomics and, according to The Human Protein Atlas, are elevated in colon tissue compared with all other tissue types.

[i] Genes written in bold are considered unique for colon tissue. [ii] Number of peptides that are significantly different ($p<0.05$) and have a fold change > or <0. The fold change is based on ^{10}log values. [iii] Total number of peptides identified for this protein.

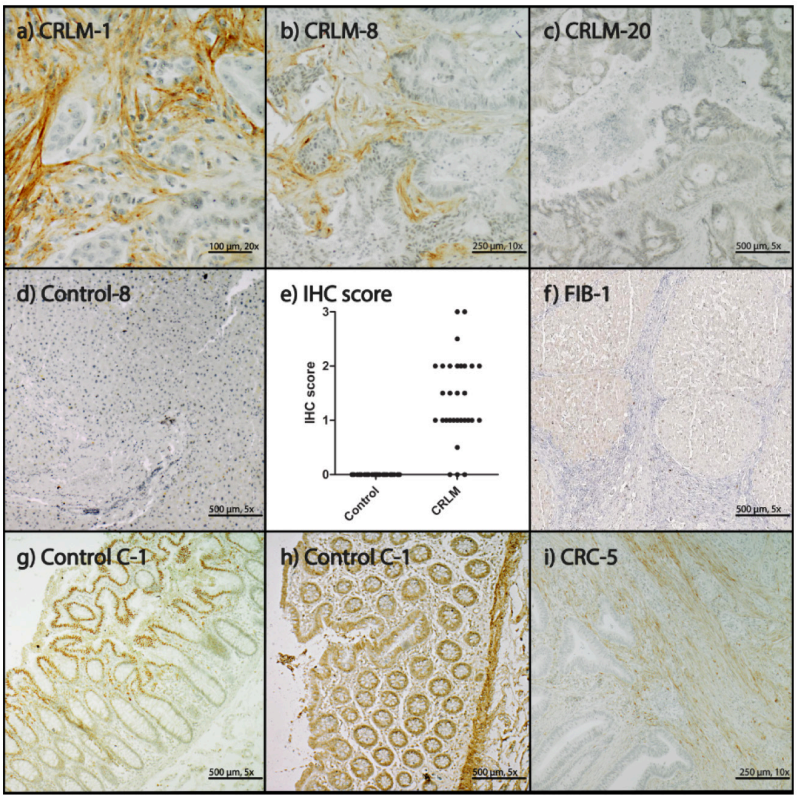


Figure 2. IHC staining of COL12A1. a) CRLM-1, IHC score 2; b) CRLM-8, IHC score 1.5; c) CRLM-20, IHC score 0; d) Control-8, IHC score 0; e) IHC scores for COL12A1 in control and CRLM tissue: 0, no staining; 0.5, small focal staining; 1, focal staining; 1.5, few focal staining areas; 2, several focal staining areas; 2.5, staining more than 40% and less than 50%; 3, >50% staining; f) FIB-1, IHC score 0; g) Control C-1, antibody NBP1-88062; h) Control C-1, antibody HPA009143, healthy tissue; i) CRC-5, IHC score: 1.

DISCUSSION

In this study we focused on the presence of collagen in CRLM and we showed by using proteomics that we could distinguish all 30 CRLM samples correctly from normal liver tissue in a relative detailed protein level. Twenty-two of the 44 known collagen alpha chains were observed in CRLM and/or control tissue (healthy adjacent liver tissue). The remaining 22 chains were not identified, possibly due to low concentration or absence in the liver.

Previously, Naba et al., showed the presence of specific collagen alpha chains in colon, CRC, liver, and CRLM tissue in a small sample set (n=3) including a pooled control (n=2) [152]. Naba

et al. observed more collagen types, although they performed additional enrichment and extensive fractionation. The authors indicated that validation needs to be performed in a larger sample set, which was confirmed by our power calculation with an estimation of (minimum) 25 samples per group. In the present study we analyzed from 30 individuals matched control and CRLM tissue providing a well-powered data set. An upregulation of collagen peptides ($P<0.0001$) in CRLM tissue compared to control tissue was observed as well as upregulation of individual collagen alpha-chains. Three collagen alpha chains (COL8A1, COL16A1, COL18A1) were not significantly affected. COL18A1 is mainly involved in the development of the eye [280], central nervous system, and liver structures [281, 282]. Musso et al. demonstrated that COL18A1 production is not increased in CRLM in comparison with liver tissue, and therefore may be considered as a negative control [283]. The low number (two or less) of identified peptides for COL4A4, COL5A3, COL7A1, COL8A1, COL15A1, and COL16A1 indicate relative low levels in CRLM. The increase in collagen is accompanied by an increase in proteins related to collagen synthesis, of which several have been described separately to stimulate tumor proliferation: P4HA1[284, 285], PCOLCE [285], PLOD1[285], PLOD2 [286], PLOD3[285], and SERPINH1 [287]. The number of proteins identified from the collagen degradation pathway were not sufficient to draw a conclusion regarding down- or upregulation. Only one collagenase was identified, MMP-9 (fold change 1.44). MMP-9 cleaves collagen types IV, V, XI, and XIV [159-161]. Other MMPs might also be upregulated [275, 276], although they were not identified.

COL12A1 MS data was cross-validated with IHC and both techniques showed significant upregulation of COL12A1. During scoring of the tissues, no abnormalities were observed in healthy adjacent liver tissue. Nevertheless, the presence of molecular abnormalities cannot be ruled out with certainty. We tried to exclude direct tumor effects by taking control tissue which was at least at a distance of 1 cm away from the tumor. Many articles referring to pre-metastatic niches (PMN) were published [288-291]. A PMN is formed by interaction of the potential metastasis site with proteins excreted from the primary tumor, creating a niche that is favorable for the growth of a metastasis. We were not able to visualize COL12A1 in three CRLM tissue sections with IHC, although increased COL12A1 levels were detected with MS. In CRC, COL12A1 was mainly present in stroma, analogous to CRLM.

However, in colon tissue it was surprisingly observed inside the nuclei of epithelial cells, with repetitive staining. We could not validate the staining of the nuclei of the epithelial cells with another antibody (HPA009143). This indicates that staining of the nuclei is most likely false positive. This antibody (HPA009143) which stains COL12A1 also showed cross-reactivity. In this case smooth muscle tissue, stained false positively. Both antibodies showed exactly the same staining for collagen structures and showed both additional cross-reactivity (nuclei endothelial cell and smooth muscle tissue).

The shotgun proteomics data allowed us to search for specific proteins produced by CRLM, which originate from the primary colon tumor tissue. First, we looked into collagen types that were expressed higher in CRC with respect to colon tissue, and were also more abundant in CRLM with respect to liver tissue. Four collagen types matched these criteria: COL10A1, COL12A1, COL14A1, and COL15A1, and the abundance was strongest for COL12A1. We hypothesize that these four collagen types visible in CRLM are a sign of origin of the primary tumor and highly specific for CRLM. We cannot exclude the possibility that fibrotic tissue is triggered by the metastatic cells to produce e.g. COL12A1, however we observed in five fibrotic tissue sections that COL12A1is not present in just fibrotic tissue itself. Collagen type X [292] and type XII [248] are described to be upregulated in CRC, however collagen type XIV and XV are not described. In CRLM collagen type X, XII, XIV, and XV are not described in relation with CRLM in literature.

Colon-specific proteins were selected from the Protein Atlas, after applying the selection criteria six proteins remained. These remaining proteins (CDH17, KRT20, CEACAM5, GPA33, MUC13, and PPP1R1B/DARPP-32) are colon specific and were significantly present in the CRLM tissue. Previous reports state that KRT20 [293], MUC13 [152, 294], CEACAM5 [295], and GPA33 [296, 297] are differently expressed in CRLM tissue; while, CDH17 and PPP1R1B/DARP-32 have not been described previously. It is likely that CDH17 is involved in cell organization and stimulates tumor proliferation; [298] while, PPP1R1B/DARPP-32 is highly expressed in CRC tissue and is a predictor for metastasis [299]. Furthermore, six of the 10 top downregulated proteins are related to liver-specific processes. The relative loss of proteins involved in liver processes in the tumor relates to a known decrease of hepatocytes in tumor tissue [300]. The top 10 upregulated proteins included collagen and two other ECM proteins, fibrillin-1 (FBN1) and fibronectin (FN1), suggesting the ECM composition strongly deviates from normal.

We conclude that collagen is upregulated in CRLM compared to control tissue and that specific collagen types (COL10A1, COL12A1, COL14A1, and COL15A1) from the primary tumor can be detected in the CRLM tissue as well. The collagen changes found in CRC and CRLM may reflect changes of the ECM related to tumor proliferation and seeding into the liver. These data may help to define new biomarkers to be used in CRLM detection after treatment of the primary tumor or its metastases.

EXPERIMENTAL PROCEDURES

EXPERIMENTAL DESIGN AND STATISTICAL RATIONALE

This study was approved by the Erasmus MC review board (MEC-2007-088) and we have

worked according to the Declaration of Helsinki. We identified collagen types which are differently present in paired (n=30) CRLM tissue compared to normal adjacent liver tissue (control). Presence of colon or CRC specific proteins in CRLM was studied by comparing CRLM, control, healthy adjacent colon (n=5), and CRC tissue (n=5). Fibrotic tissue (n=5) was analyzed to determine if COL12A1 upregulation is a general liver process or caused by metastatic cells. CRLM, control liver, CRC, and colon tissue were analyzed by IHC (COL12A1 antibody: NBP1-88062) and mass spectrometry. Fibrotic liver (FIB) tissue was analyzed by IHC (COL12A1 antibody: NBP1-88062). Colon tissue was additionally stained with a second COL12A1 antibody (HPA009143) to verify the unexpected staining of epithelial nuclei.

Sample size of the CRLM and control tissue were based on a power analysis (alpha=0.05, beta=0.20). Mean and standard deviation (SD) used for the power analysis were calculated on the overall data of log-transformed significant upregulated collagen peptides in CRLM and control tissue of five patients (control mean=5.00, CRLM mean=5.94, pooled SD=1.18). The power analysis depicted a sample size larger than 25 samples per group to determine the observed differences in the subset. No replicate measurements were performed, we assumed the biological variation to be much larger than the technical variation as has also been described in literature for the technique used [263].

Quality of the LC-MS/MS runs was monitored by measuring samples first on a test system to identify incomplete digestions and determine the presence of other components influencing the chromatography. Furthermore, after every 10 samples an QC sample was measured containing a set of peptides spread over the whole gradient to monitor possible retention time shifts or loss of intensity in the mass spectrometer. The peptides showed a retention time variation which remained within 0.2 min.

Data was assessed by analysis of the normalization by a t-test and background was assessed by permutation testing, see paragraph Data assessment (Experimental section). Proteins included in collagen turnover pathways were analyzed. Statistics are described in the paragraphs Data Analysis and in Statistical Analysis (Experimental section).

Mass spectrometry data was orthogonal cross-validated by IHC on the most distinctive collagen type.

SAMPE SELECTION CRITERIA

All samples were provided by the department of Pathology, Erasmus MC, and examined (by a GI-pathologist) for the presence of healthy and tumor tissue in the same section. Four CRLM sections were included that contained up to 50% control tissue (CRLM-8, -12, -19, and -21), the

other sections were free of control tissue. Control liver sections were demonstrated to be free of tumor tissue by standard histological examination and were taken at a minimum distance of 1 cm away from the tumor. All CRC sections included contained non-tumorous colon tissue (by standard histological examination). All samples were included in the data analysis. Tissue sections and patient information were handled according to the "Human Tissue and Medical Research: Code of Conduct for Responsible Use (Federa, 2011)". Therefore, no approval of the Medical Research and Ethics Committee was required.

CHEMICALS

Ultra-high pressure liquid chromatography grade nano-LC solvents were obtained from Biosolve (Valkenswaard, the Netherlands). Two COL12A1 antibodies (NBP1-88062 and HPA009143) were obtained from Bio-Connect BV (Huissen, the Netherlands). EnVision Detection Systems Peroxidase/DAB, Rabbit/Mouse was obtained from Dako (Heverlee, Belgium). TRIS-EDTA for Heat-Induced Antigen Retrieval (HIAR) was obtained from Klinipath (Duiven, the Netherlands). Other chemicals were obtained from Sigma-Aldrich (Zwijndrecht, the Netherlands).

IMMUNOHISTOCHEMISTRY STAINING (IHC)

Formalin-Fixed Paraffin-Embedded (FFPE) tissue sections were analyzed by MS and IHC. Consecutive tissue sections (6 µm thickness) were cut. For IHC, tissue sections were deparaffinized and rehydrated, followed by inhibition of endogenous peroxidases. Prior to primary antibody labelling, HIAR was used. Samples were labelled with a primary antibody against COL12A1, followed by secondary labelling, visualized by incubation with DAB, and counterstained with hematoxylin. Tissue sections were scored by a GI-pathologist (MD). Tissues was scored according to the following criteria: 0, no staining; 0.5, small focal staining; 1, focal staining; 1.5, few focal staining areas; 2, several focal staining areas; 2.5, staining more than 40% and less than 50%; 3, >50% staining; The full IHC protocol is available in the supporting information S-1.

SAMPLE TREATMENT FOR MASS SPECTROMETRY (MS)

Tissue sections for MS analysis were placed in an Eppendorf cup, per sample one tissue section was analyzed, irrespective from the size of the resection material. Injection volumes were normalized based on the UV-absorbance measured on a test system. The tissue sections were deparaffinized and rehydrated, followed by removal of formalin cross-links. Formalin cross-links were removed by incubation with a solution of Tris and Rapigest [301]. Followed by reduction, alkylation, and overnight trypsin digestion. Trypsin cleaves exclusively C-terminal to lysine or arginine unless proline is located at the C-terminal side [302]. The full protocol for sample treatment for MS analysis, and nano-LC and MS settings are available in the supporting

information S-1. Mass spectrometry data was made publicly accessible via the PRIDE archive, accession number: PXD008383.

DATA ANALYSIS

MGF peaklist files were extracted from raw files by ProteoWizard (v3.0.9166). MGF were searched using the Mascot search engine (v2.3.2, Matrix Science Inc., London, UK) and the UniProt/Swiss-prot database (20194 entries), as described by Singh et al. [303]. In short, the database search contained a mass tolerance of 10 ppm for the peptide mass, and 0.5 Da for the fragment mass. A maximum of 4 miss cleavages was allowed. Hydroxylation (+16 Da) of proline, lysine, and methionine were included as variable modifications and carbamidomethylation (+57 Da) of cysteine as fixed modification. Mascot search results were further analyzed by Scaffold (v4.6.2, Portland, OR, USA). In Scaffold protein confidence levels were set to a 1% false discovery rate (FDR), at least 1 peptide per protein, and a 1% FDR at the peptide level. FDRs were estimated by inclusion of a decoy database search generated by Mascot. Subsequently, the Scaffold data were imported into Progenesis QI (v4, Nonlinear Dynamics, Newcastle-upon-Tyne, United Kingdom) to align the LC runs. Features not matching the selection criteria: charge between 2 and 5, and at least two isotopes were excluded from further analysis. Identifications from Scaffold were imported via Scaffold into Progenesis QI, followed by exporting the normalized abundance to Excel 2010 (Microsoft, Redmon, WA, USA). Duplicate feature intensities were summed. Data was further processed with Excel, GraphPad Prism (v5.01, La Jolla, CA, USA), and R (v3.3.1, Vienna, Austria). Prior to ^{10}log-transformation, a relative small value of "10" was added, this to include missing values for further data analysis. The ^{10}log-transformed data was used for statistics, thereby the data was assumed to be normally distributed after log-transformation. Protein significance was determined by ranked peptide values, summation of all peptides per protein, and followed by an unequal variance t-test. In all following comparisons, an unequal variance t-test was used unless stated otherwise. P-values <0.05 were considered significant unless stated otherwise.

DATA ASSESSMENT

CRLM and control MS data were assessed, this included the analysis of normalization for injection volumes, permutation testing, and intra- and intercollagen alpha chain correlations. To exclude a bias a priori in the data, normalization was tested by summation of all individual peptide values per patient, followed by testing for significance between CRLM and control tissue.

To exclude the possibility that results were found by chance in large datasets, permutation testing (n=1000 iterations) was performed using R. The R-script has been added to the supporting information S-2. Briefly, the data was randomly divided into two groups at the

peptide level; significant differences between the two groups were determined using the Wilcoxon signed-rank test. Significant P-values were summed per permutation and the ^{10}log was taken. It was assumed that the distribution of the ^{10}log summed significant P-values was normally distributed. If the true dataset value was greater than the average value of the permutation test plus twice the SD ($p<0.05$) , then a significant difference was assumed.

STATISTICAL ANALYSIS

An unsupervised cluster analysis was performed with R on ranked peptides. The R-code is available in the supporting information S-3.

The observed upregulation of collagen can be due to a disturbance in its normal turnover. To test this hypothesis, proteins included into collagen turnover pathways were retrieved from the "Reactome Pathway Knowledgebase" [278, 279]. Three pathways present in the "Reactome Pathway Knowledgebase" are related to collagen production and degradation: "Collagen biosynthesis and modifying enzymes", "Assembly of collagen fibrils and other multimetric structures", and "Collagen degradation". Although not included in the "Reactome Pathway Knowledgebase", Xaa-Pro dipeptidase (PEPD) is involved in collagen degradation. PEPD cleaves Xaa-Pro and Xaa-Hyp, but not Pro-Pro, and is essential for the final degradation of collagen.

PEPD was added to the retrieved list of proteins from the "Reactome Pathway Knowledgebase". The CRC protein signature in CRLM was determined at two levels: collagen specific and the general proteome. At the collagen level, the total unique peptide count was used as a measure of abundance. Proteins were considered as "hitchhiked" from the colon to CRLM if they matched the following criteria: 1) were marked in the Human Protein Atlas (www.proteinatlas.org) as tissue enriched, group enriched, or tissue enhanced in colon tissue; 2) had an intensity based absolute quantification (iBAQ) value at the protein level for colon tissue but not for liver tissue; 3) the RNA expression level was 100-fold lower in the liver vs. colon tissue, and had five peptides or more identified per protein. iBAQ and RNA expression levels were taken from https://www.proteomicsdb.org/.

ACKNOWLEDGEMENTS

The authors thank the Erasmus MC Cancer Computational Biology Center for access to their IT-infrastructure and software, which was used for the computations and data analysis in this study (https://ccbc.erasmusmc.nl/).

Conflict of interest. The authors declare that they have no conflicts of interest with the contents of this article.

AUTHOR CONTRIBUTIONS

The study was designed by LD, JIJ, NvH, RCB, and TL. Experiments were performed by NvH, and acquired data was analyzed by NvH and MD. All authors were involved in writing of the manuscript and approved the submitted manuscript.

CHAPTER 5

Down-regulation of collagen hydroxylation in colorectal liver metastasis

N.A. van Huizen[1,2], P.C. Burgers[2], J. van Rosmalen[3], M. Doukas[4], J.N.M. IJzermans[1], and T.M. Luider[2*]

[1]Department of Surgery, [2]Department of Neurology, [3]Department of Biostatistics, [4]Department of Pathology, Erasmus University Medical Center, Rotterdam, the Netherlands.

Under review: FASEB journal (2019)

ABSTRACT

Collagen is significantly upregulated in colorectal liver metastasis (CRLM) compared to liver tissue. Expression levels of specific collagen types in CRLM resemble those in colorectal cancer (CRC) and colon tissue. We investigated whether the collagen hydroxylation pattern from the primary tumor also migrates with the metastatic tumor. The degree of collagen alpha-1(I) hydroxylation in colon, CRC, liver, and CRLM tissue of the same individuals (n=14) was studied with mass spectrometry. The degree of hydroxylation was investigated in 36 collagen alpha-1(I) peptides, covering 54% of the triple helical region. The degree of hydroxylation in liver tissue was similar to that in colon tissue. The overall degree of hydroxylation was significantly lower (9% ± 14%) in CRC tissue and also significantly lower (12% ± 22%) in CRLM tissue compared to colon. Furthermore, eleven peptides with a specific number of hydroxylations are significantly different between CRLM and liver tissue; these peptides could be candidates for the detection of CRLM. For one of these eleven peptides, a matching naturally occurring peptide in urine has been identified as being significantly different between patients suffering from CRLM and healthy controls. The hydroxylation pattern in CRLM resembles partly the pattern in liver, primary colorectal cancer and colon.

INTRODUCTION

Collagen is a family of proteins that form triple helical structures. A triple helix is formed by three protein subunits called alpha-chains. The specific part of an alpha-chain involved in triple helix formation consists of a repeating unit $[\text{-Gly-Xaa-Yaa-}]_n$ where Xaa and Yaa can be different amino acids, frequently proline and lysine.[48] Proline and lysine can be enzymatically modified prior to triple helix formation. Proline is hydroxylated into hydroxyproline (Hyp); two different Hyp isomers exist in collagen, namely 4-hydroxyproline (4Hyp) and – more rarely – 3-hydroxyproline (3Hyp). Normally, 4Hyp at the Yaa-position is formed by 4-prolyl hydroxylase, which requires Gly-Xaa-Pro as substrate.[304] Neither the enzyme, nor the substrate required for the formation of 4xHyp (4Hyp at the Xaa-position) is known.[75] 3Hyp is formed by 3-prolyl hydroxylase, which requires Gly-Pro-4Hyp as substrate.[305] 3Hyp and 4Hyp are involved in triple helix stability and fiber formation, but their specific functions are still not fully understood.[306-308] The function of 4xHyp remains unknown.[75] Lysine can be modified into 5-hydroxylysine (5Hyl) by lysyl hydroxylases.[109] In addition, 5Hyl can be further glycosylated [118] and is involved in collagen cross-linking [309]. After triple helix formation, collagen triple helices are excreted from the cell via the Golgi apparatus into the extracellular matrix. In the extracellular matrix, triple helices form a variety of supramolecular structures.[310]

Triple helix folding is reversible; unfolding occurs when the melting temperature (T_m) is exceeded.[311] The T_m is influenced by the physical properties of the amino acids at the Xaa and Yaa positions[312] and post-translational modifications (PTMs). In short model peptides, hydroxylation of proline into 4Hyp increased the T_m up to 6°C [306], while conversion into 3Hyp decreased the T_m [306]. The effect of 5Hyl on the T_m remains unknown. Leikina et al. demonstrate that the T_m of collagen type I in lung is just below body temperature, in addition, no folded triple helices are present below 30°C.[313]

The 4Hyp formation rate is influenced by the amino acid at the Xaa position [79], the temperature at which collagen-producing cells are cultured [64], and the amount of oxygen present [314]. Collagen produced under hypoxic conditions has approximately 1.7% less Hyp and almost 10% less Hyl.[314] Under hypoxic conditions, proline hydroxylation is the rate limiting step of collagen formation, indicating that a minimum number of PTMs is required to form stable collagen. Because collagen hydroxylation is influenced by many parameters, we hypothesize that the hydroxylation pattern varies strongly between different tissues and within a particular tissue. In agreement with this, Montgomery et al. described that the degree of hydroxylation in collagen alpha-1(I) of ten breast cancer patients varied between 0% and 90%, depending on the proline position.[73]

Collagen levels in tumor tissue are dysregulated in comparison to healthy tissue.[315] In colorectal liver metastasis (CRLM), collagen levels are, in general, significantly upregulated in comparison to adjacent normal liver tissue.[26, 152] Collagen type 10, 12, 14, and 15 expression levels have most likely migrated from the healthy colon tissue, via the primary tumor (colorectal cancer; CRC), to the metastasis. The migration of colon-specific protein expression into CRLM is not limited to collagen; other proteins, e.g. CDH17, KRT20, CEACAM5, GPA33, MUC13, and PPP1R1B/DARPP32 show a similar pattern, although many colon-specific proteins are down-regulated too.[26]

In this study we have investigated if the hydroxylation pattern, taking hydroxyproline and hydroxylysine into account, of collagen alpha-1(I) migrates from colon tissue, via the primary tumor, towards the metastasis similarly to a number of colon-specific proteins and specific collagen types. Collagen hydroxylation was studied with bottom-up mass spectrometry in colon, CRC, CRLM, and liver tissue obtained from the same individual.

MATERIALS AND METHODS

SAMPLE SELECTION

Bottom-up proteomics data of matching normal liver tissue and colorectal liver metastasis (CRLM) tissue of 30 patients were retrieved from the PRIDE Archive (PXD008383).[26] We could collect matching healthy adjacent colon tissue and CRC tissue from 14 out of these 30 patients. The tissue material had been collected during curative surgery. Morphology differences indicated that the CRC tissues contained non-tumor material (most likely adjacent normal tissue), and that one colon tissue contained non-normal tissue (most likely tumorous tissue). None of the CRLM tissues contained non-tumor material, and none of the liver tissues contained non-normal tissue. This study was approved by the Erasmus MC ethics review board (MEC-2007-088) and we have worked according to the Declaration of Helsinki. An experienced pathologist (MD) has reviewed the sections.

SAMPLE PREPARATION AND DATA PROCESSING

All tissues were prepared, measured, and the data was processed as described by van Huizen et al.[26] In short, formalin cross-links were removed by incubation with TRIS, followed by cysteine reduction and alkylation. The proteins in the tissue were cleaved into peptides by overnight incubation with trypsin. Peptides were identified with nanoLC-MS/MS measurements on an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fischer Scientific, San Jose, CA, USA). Raw data was converted into Mascot Generic Files and searched with Mascot (v2.3.2, Matrix Science Inc., London, UK) and the UniProt/Swiss-prot Human database

(20194 entries). Hydroxylations (+16 Da) of proline, lysine, and methionine were included as variable modifications and carbamidomethylation (+57 Da) of cysteine as fixed modification. Mascot search results were further analyzed in Scaffold (v4.6.2, Portland, OR, USA). In Scaffold, protein confidence levels were set to a 1% false discovery rate (FDR), at least 2 peptides per protein, and a 1% FDR at the peptide level. FDRs were estimated by inclusion of a decoy database search generated by Mascot. A Peptide Report exported from Scaffold was used for data analysis. Mass spectrometry data was made publicly accessible via the PRIDE archive, accession number: PXD015015.

PEPTIDE PANEL SELECTION

Collagen alpha-1(I) peptides were selected on the following criteria: 1) does not contain methionine or cysteine; 2) contains at least one proline or lysine moiety; and 3) a unique number of hydroxylations, having been identified in 20 or more liver or CRLM tissues, and in 3 or more colon and CRC tissues. The latter sub criterion was applied because the statistical formulas used require a sample size of 3 or larger.

PRINCIPAL COMPONENTS ANALYSIS

Principal component analysis (PCA) was performed in R (v3.6.0.)[266]; the R-script can be found in the supplemental (S-1). A table was created containing for every sample a value for the absence (value 0) or presence (value 1) for each peptide with an unique number of hydroxylations. Zero to 10 hydroxylations were taken into consideration; where 10 is the maximum number of hydroxylations observed for the peptides in the peptide panel. The PCA was calculated with the 'prcomp' function. If two or more tissue types were compared, then a 2D-PCA-plot was made with the package 'factoextra'. 3D-PCA-plots were made by exporting the coordinates of the first three PCA dimensions, and plotting with 'scatter3d'. For assessment of the data, 2D-PCA-plots were made by exporting the PCA coordinates, and plotting with 'ggplot2'.

HYDROXYLATION ANALYSIS

Statistical analysis was performed in R (v3.6.0.)[266], the R-script can be found in the supplemental (S-2). In the digested tissue, collagen peptides are present with different numbers of hydroxylations per peptide. A difficulty with collagen analysis is the identification of the exact location of a hydroxylation. To avoid the identification of the exact hydroxylation position, we used the number of hydroxylations per peptide obtained by the parent mass. Per tissue type (n=4), per patient (n=14 and n=30), and per peptide from the peptide panel (n=36), we checked the presence of 0 to 10 hydroxylations taking into account both hydroxyproline and hydroxylysine, which generated a three-dimensional data matrix. Per peptide a list was generated containing the numbers of different forms observed, compared between

two tissue types. Testing for differences between tissue types was performed with a paired Wilcoxon signed-rank test. Per tissue a list was generated containing the average number of hydroxylations per peptide, which was compared between two tissue types with a paired Wilcoxon signed-rank test. If a peptide was not identified in tissue from a particular patient, no average could be calculated, then this patient was removed from analysis of the average number of hydroxylations. Besides the overall differences between the number of different forms observed, and the average number of hydroxylations per peptide, we also studied in more detail if there are peptides with a specific number of hydroxylations that is different between different tissues. The difference in a specific hydroxylation form between two tissue types was assessed with the McNemar test. Two-sided p-values below 0.05 were considered as significant. Figure 1 shows an example of the set-up of the data.

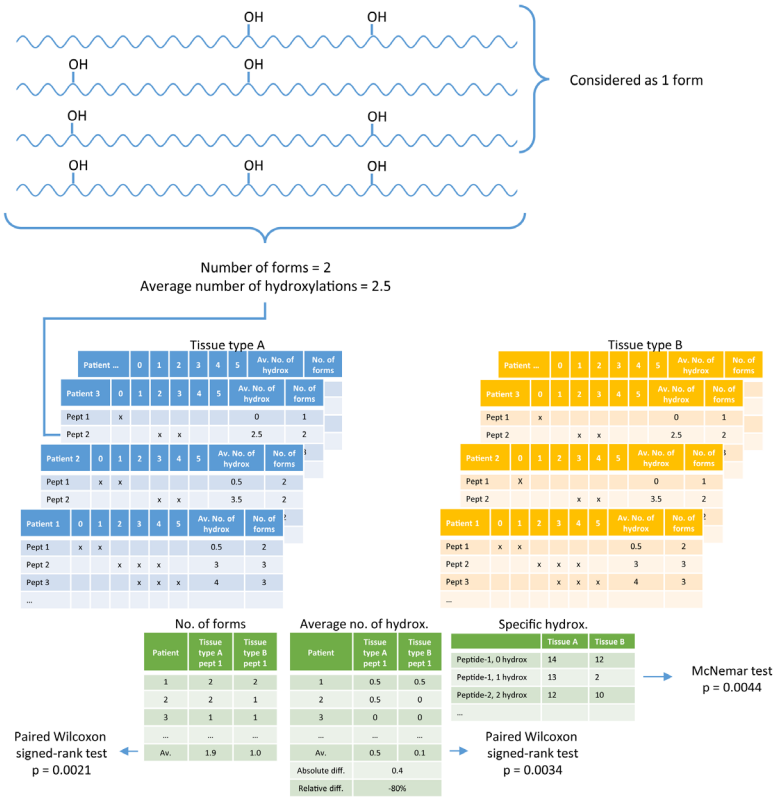


Figure 1. A diagram explaining the data analysis. Per tissue, per patient the number of forms present in collagen peptides, and the average number of hydroxylations per peptide is determined (blue and yellow tables). Per peptide, per patient these two values are compared between two tissue types, left and middle green table. In addition, every form (specific hydroxylation) is separately compared between two tissue types based on the number of patients in which the specific form is identified, right green table.

DATA ASSESSMENT

The data quality was assessed prior to the analysis of the collagen hydroxylation. Per tissue type a 2D-PCA-plot was made, in which the color of the data points indicates the age, the shape indicates the gender, and, in the case of tumor tissue, the shape of the data points indicates the location of the primary tumor. Clustering of the data points based on color, shape, or size would indicate the presence of a bias in the data set. Also, a 2D-PCA-plot was made between Colon and CRC tissue, because they contained non-normal and non-tumor tissue, respectively. If grouping was based on the percentage non-tumor and non-normal tissue, then a bias in the data would be present.

To study the within tissue variation in liver (n=30) and in CRLM (n=30) tissue, permutation testing (n=1000 iterations) was performed. Samples were split randomly in two groups (n=15 each) and compared based on the number of forms, the average number of hydroxylations, and specific hydroxylation forms. Permutation testing (n=1000 iterations) was also performed for the between different tissue comparisons. A frequency table was made based on the number of significant values per iteration. The permutation results shown are the 95% cumulative values.

If, between tissue types, a significant difference in the number of forms was present, then per tissue the presence of linear relation between the number of forms and the average number of hydroxylations was tested with the Pearson correlation test. A linear relation was considered significant if the p-value was below 0.05/4 (Bonferroni correction).

ANALYSIS OF THE HYDROXYLATION VARIATION

The hydroxylation pattern was compared between: colon and liver, colon and CRC, CRC and CRLM, colon and CRLM, and liver and CRLM. The difference in the degree of hydroxylation was analyzed in different directions: 1) 2D- and 3D-PCA-plots; 2) the number of forms observed; 3) detailed analyzes of the average number of hydroxylations; and 4) analyzes of a specific hydroxylation form. Detailed information regarding the average number of hydroxylations was obtained by: plotting the average number of hydroxylations per peptide; the absolute difference in the average number of hydroxylations per peptide of CRC, CRLM, and liver tissue with respect to colon tissue (tissue – colon tissue); and the relative difference with respect to colon tissue ([tissue/colon -1]*100%). Differences between all groups were tested with 2-sided t-tests. For the average number of hydroxylations per peptide, p-values below 0.05/6 (Bonferroni correction, 6 comparisons), and for the relative differences, p-values below 0.05/3 (Bonferroni correction, 3 comparisons) were considered significant. The average of the relative difference was compared to zero with a one-sample t-test; p-values below 0.05/3 (Bonferroni correction, 3 comparisons) were considered as significant.

RESULTS

PATIENT CHARACTERISTICS

An overview of all patient characteristics is present in the supplemental (S-3). CRC patients (n=14) had a median age of 64 years (interquartile range [IQR] 57-73) and were mostly male (64%). The CRC tumor had a median size of 4.0 cm (IQR 3.3-4.8), being moderately differentiated. CRC was located in colon (29%), rectum (57%), or sigmoid (14%). Liver and CRLM tissue were not collected simultaneously with colon and CRC tissue. The CRLM patients had a median age of 65 years (IQR 59-73) and a median of 2 (maximum 7) tumors with a median size of 2.3 cm (IQR 1.2-3.5), being moderately differentiated.

PEPTIDE SELECTION

The number of identified collagen alpha-1(I) peptides with unique numbers of hydroxylation differed per tissue type (ANOVA $p < 0.001$), only colon and CRC were not significantly different (Bonferroni post hoc testing $p > 0.05$) and all other comparisons were significantly different (Bonferroni post hoc testing $p < 0.0001$). In liver 70 (n=30, SD=17), CRLM 111 (n=30, SD=23), colon 162 (n=14, SD=10), and CRC 168 (n=14, SD=28) peptides with unique number of hydroxylations were identified. A total of 113 different peptides with a unique number of hydroxylations were identified, which did neither contain methionine nor cysteine. Thirty-six peptides fulfilled all selection criteria and were selected for the peptide panel. These peptides covered 54% of the triple helical region of collagen alpha-1(I), supplemental S-4. They are numbered corresponding to their occurrence in the primary structure of collagen, and will be referred in the text as, for example, peptide 1.

DATA ASSESSMENT

In the constructed 2D-PCA-plots, the degree of collagen hydroxylation in CRLM and CRC tissue showed no clustering based on patient age, gender, or location of the (primary) tumor. Likewise, the degree of hydroxylation in liver and colon tissue showed, based on 2D-PCA-plots, no clustering based on patient age or gender. CRC tissue contained non-tumor material, and one colon tissue contained non-normal tissue. Nevertheless, colon and CRC samples still clustered based on the tissue type. The 2D-PCA-plots are available in the supplementary (S-4).

Permutation testing results are presented in table 1. The within tissue variation in liver and CRLM tissue is similar to the number of differences observed by chance between the different tissue comparisons. The permutation testing results indicate that, differences between two tissue types are not random effects if more than 4 differences in the number of forms, 3 differences between the average number of hydroxylations per peptide, or 4 specific hydroxylation forms are found.

Table 1. Permutation (n=1000 iterations) testing results, showing the maximum numbers of significant hits (p <0.05) that were found in 95% of the iterations.

Tissue comparisons	Number of forms	Average number of hydroxylations	Specific hydroxylation form
Liver[i]	3	2	4
CRLM[i]	4	2	3
Liver vs colon	3	3	3
Colon vs CRC	3	2	2
CRLM vs CRC	3	3	3
Colon vs CRLM	4	3	3
Liver vs CRLM	3	2	3

[i] Within tissue variation, sample set was split in half and compared multiple times

TISSUE COMPARISON

In the 2D-PCA-plot shown in figure 2, liver, CRLM, and colon are separated based on their hydroxylation pattern. CRC is separated from liver and CRLM tissue, but overlaps with colon tissue. Two exceptions are observed. Patient CRLM-10 is located (coordinates: -3.98, 0.22) in between the liver samples, most likely because only 23 different peptides with a unique number of hydroxylations were identified (liver average = 62). Patient CRC-19 is located (coordinates: -1.01, 0.38) in between the liver samples, because only 69 different peptides with a unique number of hydroxylations were identified (CRC average = 92). A 3D-PCA-plot based on the first three dimensions results in a complete separation between liver, CRLM, and Colon+CRC, while colon and CRC are still partly overlapping. The 3D-PCA-plot is available in the supplemental (S-4). The 2D-PCA-plots and 3D-PCA-plots for individual tissue comparisons are also available in the supplemental (S-4). For the comparison 'liver vs colon', there is a completely separation in a 2D-PCA-plot and 3D-PCA-plot. For 'colon vs CRC' there is a partial overlap in the 2D-PCA-plot and 3D-PCA-plot. For 'CRC vs CRLM', and 'liver vs CRLM' (14 vs 14) there is a slight overlap in the 2D-PCA-plot and they are fully separated in the 3D-PCA-plot. 'Liver vs CRLM' (30 vs 30) are partly overlapping in the 2D-PCA-plot, and fully separated in the 3D-PCA-plot.

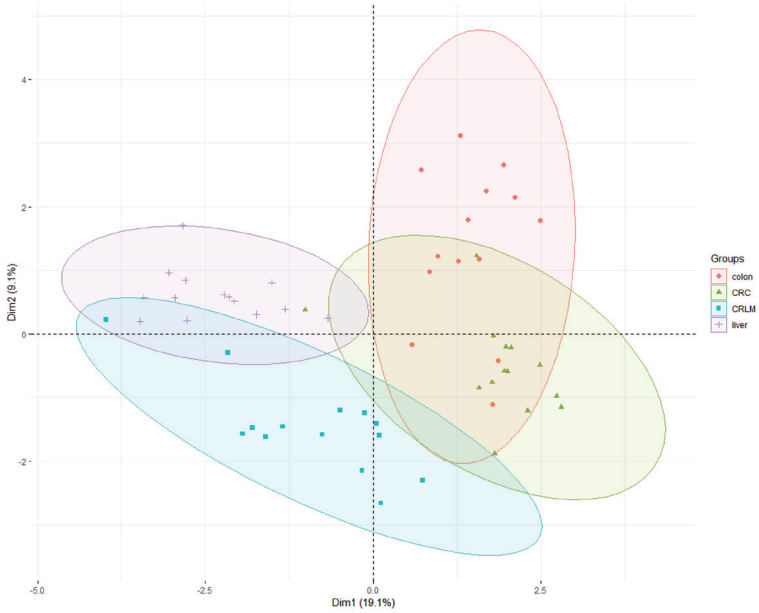


Figure 2. 2D-PCA-plot showing the separation of all 4 tissue types based on their degree of hydroxylation. The y-axis is the first dimension of the PCA, which explains 19.1% of the variation. The x-axis is the second dimension of the PCA, which explains 9.1% of the variation.

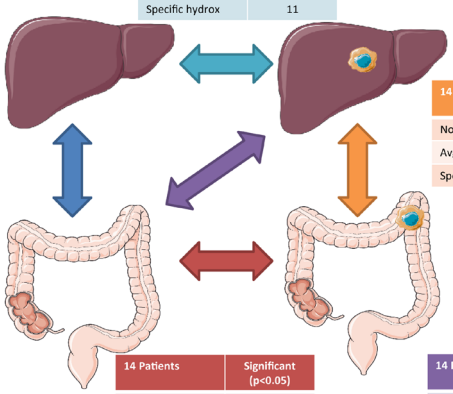
An overview of the results of the tissue comparisons is presented in figure 3. The number of significant peptides and the direction (down- or upregulated) indicate the similarity between the different tissue types. To compare the effect of group size, the comparison 'liver vs CRLM' was performed with the patients of whom colon and CRC tissue was collected (14 vs 14) and with the full data set (30 vs 30). In the larger data set, more and stronger significant differences were observed, most likely due to the larger data set.

A Pearson correlation test was performed to exclude a relation between the number of different forms observed and the average number of hydroxylations per peptide. Significant correlations were not found (p-critical = 0.05/4=0.0125, Bonferroni correction); liver: r=0.40, p-value=0.018 CRLM: r=0.31, p-value=0.066; colon: r=0.21, p-value=0.21; CRC: r=0.16, p-value=0.35.

	No. of forms		Avg. No. of hydrox		Specific hydrox.	
	Down	up	Down	up	Down	Up
Liver vs colon	2	26	8	5	3	27
Colon vs CRLM	18	3	12	1	22	5
Colon vs CRC	4	3	11	0	7	3
CRC vs CRLM	3	8	4	7	13	5
Liver vs CRLM (30)	2	22	19	3	7	30
Liver vs CRLM (14)	1	5	14	0	3	8

30 Patients	Significant (p<0.05)
No. of forms	24
Avg. No. of hydrox	22
Specific hydrox	37

14 Patients	Significant (p<0.05)
No. of forms	6
Avg. No. of hydrox	14
Specific hydrox	11



14 Patients	Significant (p<0.05)
No. of forms	28
Avg. No. of hydrox	13
Specific hydrox	30

14 Patients	Significant (p<0.05)
No. of forms	11
Avg. No. of hydrox	11
Specific hydrox	17

14 Patients	Significant (p<0.05)
No. of forms	7
Avg. No. of hydrox	11
Specific hydrox	18

14 Patients	Significant (p<0.05)
No. of forms	21
Avg. No. of hydrox	13
Specific hydrox	27

Figure 3. Tissue comparison analysis. The black table shows the number of peptides down- or upregulation regarding the differences in the number of forms, the average number of hydroxylations, and specific hydroxylations. The arrows indicate the comparisons made and tables surrounding the tissue cartoons having the same color belong together.

In all tissue comparisons the total number of significant differences with respect to the average number of hydroxylations is similar. However, the comparisons of 'liver vs colon', and 'CRC vs CRLM' have a similar number of down- and upregulated averages (black table fig 3), while the comparisons 'colon vs CRC', 'colon vs CRLM', and 'liver vs CRLM' have an overall increase in downregulation of the average.

Figure 4 shows a more detailed analysis of the average number of hydroxylations. In figure 4A), the average number of hydroxylations per peptide per tissue is shown. The overall average numbers of hydroxylations in CRC and CRLM are significantly lower than those in colon and liver tissue. Liver and colon tissue do not differ in number of hydroxylations, and also CRC and CRLM do not differ in this respect. Figure 4B) shows the differences in the average number of hydroxylations in the comparison of colon tissue and the other tissues investigated. Figure 4B) visualizes the overall shift in the average number of hydroxylations per peptide. The overall shift in the number of hydroxylations per peptide in liver tissue, with

respect to colon tissue, does not significantly differ from zero. Overall, the mean of all average number of hydroxylations per peptide is significantly downregulated in CRC ($p=6.3*10^{-3}$) and in CRLM ($p=6.3*10^{-5}$) tissue compared to liver tissue. CRC and CRLM do not have a different shift in comparison to each other (p=0.27). In figure 4C), the relative differences in the average with respect to colon tissue are shown. There is a significant decrease in the average degree of hydroxylation, -9% in CRC (SD = 14%, p-value=0.00056), and -12% in CRLM (SD = 22%, p = 0.0026) with respect to colon tissue, and liver tissue does not significantly differ from colon tissue (-3%, SD = 27%, p = 0.55).

Figure 4C) visualizes how the degree of hydroxylation changes per peptide from colon to CRC, to CRLM, to liver tissue. For peptide 1, the average number of hydroxylations remains constant with the transition from colon to CRC tissue. However, in CRLM tissue the average number of hydroxylations decreases strongly. The degree of hydroxylation of peptide 1 in CRLM is similar to that in liver tissue. Regarding peptide 21, the average number of hydroxylations decreases strongly from colon to CRC, and from CRC to CRLM, and is fairly similar between CRLM and liver tissue. Regarding peptides 15 and 16, the average number of hydroxylations decreases from colon to CRC, and increases from CRC to CRLM to liver tissue. Peptides 1, 15, 16, and 21 appear to move towards liver tissue, when transitioning from colon, via CRC, to CRLM.

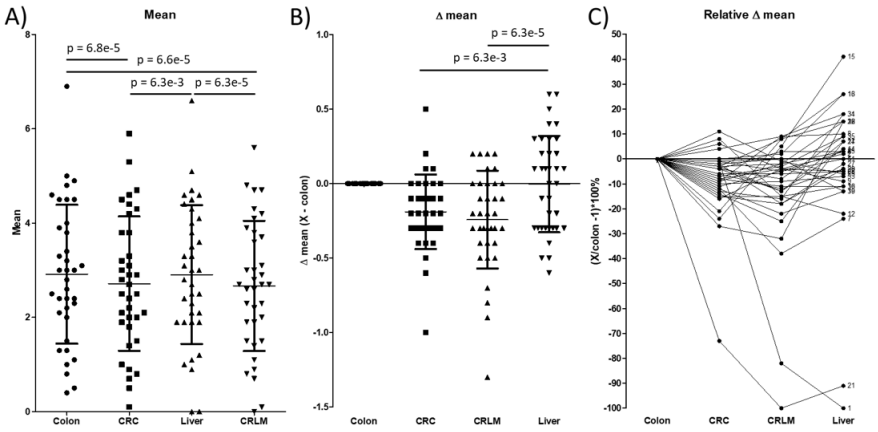


Figure 4. A) The average number of hydroxylations per peptide. Tested for differences with a repeated paired t-test. P-values below 0.05/6=0.0083 are considered significant. B) Absolute difference in the average number of hydroxylations per peptide with respect to colon tissue. P-values below 0.05/3=0.0167 are considered significant. C) Relative difference in the number of hydroxylations with respect to colon tissue. The 'X' in the legend of the Y-axis represents CRC, CRLM, or liver tissue, depending on the comparison made. Only significant values are presented as horizontal bars in A) and B).

The largest number (n=30) of different specific hydroxylation sites were detected between liver and colon tissue; a comparable number (n=27) of different specific hydroxylation sites is present between colon and CRLM tissue. A closer study of specific hydroxylations that differ between liver and CRLM tissue, shows several unique differences, see table 2. Peptide 7 with 0 hydroxylation is more prevalent in CRC and CRLM than in healthy colon and liver tissue. Peptide 13 with 1, peptide 15 with 1, peptide 16 with 1, peptide 18 with 0, peptide 19 with 3, and peptide 34 with 2 hydroxylations are abundantly present in colon, CRC and CRLM, but hardly in liver tissue. Peptide sequences are presented in the supplemental (S-4). Peptide 17 with 2, and peptide 27 with 5 hydroxylation are less abundant in CRLM compared to colon, CRC, and liver. Peptide 14 with 6, and peptide 24 with 5 hydroxylations, are present at similar levels in colon and liver, less in CRC, and the least in CRLM. The general pattern emerging from this data is consistent with a lower degree of hydroxylation in CRLM than in liver or colon tissue.

Table 2. Peptides with a specific hydroxylation form that are significantly different between liver and CRLM tissue, and the number of patients, per tissue, in whom the peptide was identified.

Peptide	Number of prolines and lysines	Number of hydroxylations	14 Patients (matching colon, CRC, liver, and CRLM)				30 Patients (matching liver, and CRLM)	
			Colon	CRC	CRLM	Liver	CRLM	Liver
7	7	0	0	2	8	0	15	0
13	4	1	14	14	12	3	26	13
14	8	6	13	7	3	14	5	27
15	5	1	14	14	12	3	26	6
16	9	2	13	13	8	1	20	3
17	4	2	14	10	4	11	10	18
18	4	0	7	9	11	1	26	3
19	6	3	8	11	11	0	22	3
24	9	5	5	3	0	8	2	13
27	9	5	13	12	3	9	5	17
34	9	2	12	11	10	0	20	2

In urine, natural occurring peptides (NOPs) were identified that overlap with six peptides from table 2.[manuscript in submission] Only for peptide 34, from table 2, a NOP (AGPPGAPGAPGAPGPVGPAGKSGDRGETGP with 2 hydroxylations) was significantly different (p-value of $3.1*10^{-4}$) between urine from patients suffering from CRLM and healthy control urine. On the basis of the b- and y-ions from MS/MS spectra, the peptide AGPP(-OH)GAP(-OH) GAPGAPGPVGPAGKSGDRGETGP is identified to be hydroxylated at the 5th and 8th amino acid in colon tissue, CRC tissue, liver tissue, CRLM tissue, control urine, and CRLM urine.

DISCUSSION

The hydroxylation pattern of collagen alpha-1(I) is tissue-specific and each studied tissue has a unique hydroxylation pattern and a degree of hydroxylation as could be detected by PCA and statistical analysis. The variation in the hydroxylation pattern between liver and colon is larger than the variation within each tissue. Even though the average number of hydroxylations is comparable between colon and liver tissue, locally along the collagen alpha-1(I) chain the degree of hydroxylation is significantly different. The variation within a tissue and between healthy tissues implicates that every collagen triple helix has a unique hydroxylation pattern. Further studies are required to investigate whether prolines at specific sites are always hydroxylated. In addition, it would be of interest to ascertain whether a lack of specific hydroxyprolines could inhibit triple helix formation, or prevent incorporation of such a triple helix into supra-molecular structures. The hydroxylation pattern has a strong influence on the T_m. We suggest that every triple helix consequently has a slightly different T_m, and therefore interacts differently with the surroundings, and has different micro-unfoldings. In previous research it has been assumed that all prolines are hydroxylated; for example, to determine the T_m *in silico*.[98] We show that this is not the case and that this assumption leads to an overestimation of the T_m.

The collagen hydroxylation pattern is studied based on 36 selected peptide, of the selected peptides, several contain a miss cleavage present at arginine at lysine. A possible explanation is the presence of hydroxylation at the lysine moiety, which has been described to reduce the activity of trypsin.[316] Although, peptide 1, which only contains one lysine moiety, is observed with and without a hydroxylation at lysine. This example shows that trypsin is able to cleave at a hydroxylated lysine. It might be more likely that the presence of aspartic acid and glutamic acid in close proximity to an arginine and lysine moiety reduces the efficiency of trypsin activity.[317, 318]

We investigated if the hydroxylation pattern in a tumor migrates with the metastasis. The results of this study indicate that the hydroxylation degree in tumor tissue decreases significantly in comparison to degree of hydroxylation in healthy tissue. The degree of hydroxylation in CRC tissue is 9% lower and that in CRLM is 12% lower than in colon and liver tissue. Still, in the comparisons 'colon vs liver' and 'CRLM vs CRC' similar numbers of peptides were down- and upregulated. The hydroxylation pattern of several of the studied peptides appears to be tissue-dependent. For instance, peptide 1 was present in colon tissue with 0 and 1 hydroxylation, but in liver tissue only with 0 hydroxylation. CRLM tissue seems to be in an intermediate state, containing both peptide 1 with 0 hydroxylation and – to a lesser extent – peptide 1 with 1 hydroxylation. Peptide 1 contains just one lysine moiety, and zero proline

moieties, further modification of 5-hydroxylysine can prevent us from detecting peptide 1 with 1 hydroxylation. To our knowledge additional modifications of this specific lysine are not described in literature; nevertheless, if an additional modification is present, then this still indicates a significant difference between the tissues that have been compared. In line with the peptide 1 pattern, the hydroxylation degree of peptide 21 in colon tissue is similar to that in CRC tissue, and strongly reduced in both CRLM and liver tissue. The hydroxylation degree of peptides 15 and 16 decreases in CRC tissue, increases in CRLM tissue, and is even higher in liver tissue. Despite the reduced degree of hydroxylation in CRLM tissue, the general hydroxylation pattern does partly resemble liver tissue but also CRC, and colon.

A recently described method to detect CRLM is by measuring collagen natural occurring peptides (NOPs) in urine.[11, 12] Higher sensitivity and specificity can potentially be obtained by adding distinctive collagen NOPs to the panel of molecular markers as described by us. In the present study, we found an overall reduction in the degree of collagen hydroxylation in tumor tissue, and identified specific peptides which are distinctive between CRLM and liver tissue. Comparison with previous data [manuscript in submission] showed that a NOP, which overlaps with peptide 34, was significantly different in urine of patients suffering from CRLM and in urine of healthy controls. This NOP is a promising candidate to improve the detection of CRLM in urine. If the degree of collagen hydroxylation is too low, then the T_m of the formed triple helix can be too low and as a result this collagen will not be built into the supramolecular structures. Consequently we may find that there is a possibility to find NOPs in urine that have an even lower degree of hydroxylation than observed in this study, which could possibly be even more distinctive.

Hypoxia is a possible explanation of the overall decreased degree of hydroxylation in tumor tissue. McKeown clarified differences in the partial pressure of oxygen (pO_2) in healthy and tumor tissue.[319] The pO_2 in rectal cancer tissue (30 mm Hg) is approximately 2 times lower than in healthy rectal tissue (52 mmHg); the pO_2 in liver tumor (6.0 mmHg) is approximately 5 times lower than in healthy liver tissue (30 mmHg). Utting et al. showed that hypoxia reduces the degree of hydroxylation and has the largest effect on Hyl formation, and a relative small effect on Hyp formation.[68] In contrast to the above reasoning, the hydroxylation degree of peptides 1 and 21 is decreased in liver tissue in comparison to colon tissue. Peptide 15 and 16, also in contrast to the overall trend, have an increasing hydroxylation degree from CRC to CRLM to liver tissue. Hypoxia can explain the general trend observed, although specific cases such as peptides 1, 15, 16, and 21 do not agree with these observations of McKeown en Utting et al. and may be explained by unknown mechanisms by which hydroxylation is regulated specifically.

Proline and lysine can be enzymatically hydroxylated. Van Huizen et al. described the general upregulation of enzymes related to collagen production and hydroxylation in CRLM.[26] Except the specific enzyme P4HB, the beta unit of 4-prolyl hydroxylase, is strongly reduced in CRLM (11, SD=5 unique peptides) compared to liver tissue (20, SD=5 unique peptides) (fold change = 0.48, p-value = $1.6*10^{-18}$). P4HB is crucial for the solubility and folding of 4-prolyl hydroxylase. In the present study in colon tissue, on average, 12.8 peptides (SD=4.1) belonging to P4HB were identified, and in CRC on average 16.1 peptides (SD=3.9). The downregulation of P4HB in CRLM in comparison to liver tissue might explain the decreased degree of hydroxylation; in CRC tissue, however, P4HB is higher than in colon tissue while the hydroxylation degree is also decreased.

The extracellular matrix is involved in tumor development.[150]. It has been proposed that the extracellular matrix can act as a tumor suppressor as long as it remains in its 'natural shape'; if the 'natural morphology' is lost, then it could become a tumor promoter [150]. In this study we have shown that besides the collagen expression, also the collagen hydroxylation is significantly different between tumor and normal tissue. Further research is required to understand the effect of the reduced degree of collagen hydroxylation on tumor proliferation and to identify the mechanisms that are involved in the regulation of changed hydroxylation levels in cancer.

CHAPTER 6

Identification of 4-hydroxyproline at the Xaa position in collagen by mass spectrometry

Nick A. van Huizen[1,2], Peter C. Burgers[2], Fabrice Saintmont[3,4], Patrick Brocorens[4], Pascal Gerbaux[3], Christoph Stingl[2], Lennard J.M. Dekker[2], Jan N.M. IJzermans[1], and Theo M. Luider[2*]

[1]Department of Surgery, [2]Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands. Mailing address for all authors: P.O. Box1738, 3015 GE Rotterdam, The Netherlands.

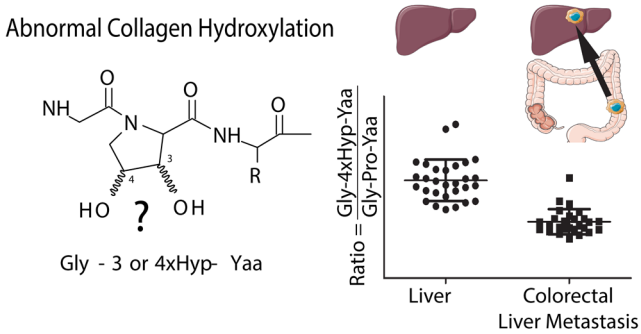
[3]Organic Synthesis & Mass Spectrometry Laboratory, Interdisciplinary Center for Mass Spectrometry (CISMa), Center of Innovation and Research in Materials and Polymers (CIRMAP), University of Mons - UMONS, 23 Place du Parc, 7000 Mons, Belgium. [4]Laboratory for Chemistry of Novel Materials, Center of Innovation and Research in Materials and Polymers, Research Institute for Science and Engineering of Materials, University of Mons, UMONS, 23 Place du Parc, 7000 Mons, Belgium

Journal of Proteome Research. 2019 May 3;18(5):2045-2051

ABSTRACT

Collagen has a triple helix form, structured by a [-Gly-Xaa-Yaa-] repetition, where Xaa and Yaa are amino acids. This repeating unit can be post-translationally modified by enzymes, where proline is often hydroxylated into hydroxyproline (Hyp). Two Hyp isomers occur in collagen: 4-hydroxyproline (4Hyp, Gly-Xaa-Pro, substrate for 4-prolyl hydroxylase) and 3-hydroxyproline (3Hyp, Gly-Pro-4Hyp, substrate for 3-prolyl hydroxylase). If 4Hyp is lacking at the Yaa position, then Pro at the Xaa position should remain unmodified. Nevertheless, in literature 41 positions have been described where Hyp occurs at the Xaa position (?xHyp) lacking an adjacent 4Hyp. We report four additional positions in liver and colorectal liver metastasis tissue (CRLM). We studied the sequence commonalities between the 45 known positions of ?xHyp. Alanine and glutamine were frequently present adjacent to ?xHyp. We showed that proline, position 584 in COL1A2, had a lower rate of modification in CRLM than in healthy liver.

The isomeric identity of ?xHyp, i.e. 3- and/or 4Hyp, remains unknown. We present a proof of principle identification of ?xHyp. This identification is based on liquid chromatography retention time differences and mass spectrometry using ETD-HCD fragmentation, complemented by ab initio calculations. Both techniques identify ?xHyp at position 584 in COL1A2 as 4-hydroxyproline (4xHyp).



Reprinted with permission from Journal of Proteome Research. 2019 May 3;18(5):2045-2051. Copyright 2019 American Chemical Society.

INTRODUCTION

Collagen is a family of proteins well-known to be naturally present as a triple helix. The triple helix is structured by a $[\text{-Gly-Xaa-Yaa-}]_n$ repeating unit, where Xaa and Yaa can be different amino acids. This repeating unit can be post-translationally modified, the most common modification being the enzymatic hydroxylation of proline into hydroxyproline (Hyp). Two Hyp isomers have been observed, namely 4-hydroxyproline (4Hyp) and 3-hydroxyproline (3Hyp). 4Hyp is formed by 4-prolyl hydroxylase, which requires Gly-Xaa-Pro as substrate [304]. 3Hyp is formed by 3-prolyl hydroxylase, which requires Gly-Pro-4Hyp as substrate.[305] 3Hyp and 4Hyp are involved in triple helix stability and fiber formation; the exact functional differences of 3Hyp and 4Hyp are not yet fully understood.[306-308]

In the literature, 45 positions have been reported where Hyp is present at the Xaa position [69-71, 73, 320]. However, these substrate positions do not match with known enzyme activity. Therefore, the identity of the enzyme generating Hyp at the Xaa position remains unknown. Hyp occurring at the Xaa position will be referred to as ?xHyp, where the "?" indicates the third or fourth position of the hydroxyl group positioned on proline, and the "x" to the Xaa position in the repeating unit observed in collagen (Gly-Xaa-Yaa). 40 out of the 41 positions have been identified by collagen PTM mapping with mass spectrometry, (in brackets the adjacent amino acid at the Yaa position is annotated): in COL4A1 of Murine Engelbreth-Holm-Swarm position: 481 (Gln) [71]; in COL4A1 of Human lens capsule positions: 71 (Gln), 514 (Gln), and 662 (Gln) [71]; in COL2A1 of Bovine positions: 214 (Ala), 220 (Gln), 358 (Val), 502 (Ala), 646 (Ala), 696 (Ala), 796 (Ala), and 1099 (Ala) [70]; in COL5A1 of Bovine placenta: positions 509 (Val), and 587 (Ala) [320]. In COL1A1 of human breast cancer 32 positions were identified, at the Yaa position were present: 2 times Gln, 3 times Arg, 10 times Ala, 1 time Gly, 1 time Pro, 2 times Val, 1 time Thr, 1 time Asp, 2 times Ile, and 2 times Ser.[73] Lysine was present 7 times at the Yaa position, however hydroxylysine was not included in the database search and therefore this finding is likely to be an artefact and is excluded from further analysis. The 41thposition has been identified in a specific peptide of COL1A2 in Bovine gelatin (position 812 (Ala)) which is part of a panel of collagen peptides to distinguish different animal sources of gelatin [69]. In addition to these 41 known positions, we report here four additional positions. We studied the sequence commonalities between these 45positions to identify a possible substrate for the unknown enzyme which modifies Pro into ?xHyp.

Kassel et al. distinguished 3Hyp and 4Hyp based on w-ions in mussel adhesive proteins by mass spectrometry (MS) [74]. Their method applied to our case did not result in isomer identification due to interference of the expected transitions of interest. We therefore developed a new MS-based method to distinguish 3Hyp and 4Hyp in collagen peptides. As a

proof of principle, we here show the identification of ?xHyp in one of the newly identified positions (GLHGEFGLP(4Hyp)GP(?xHyp)AGPR) based on both liquid chromatography retention time difference and MS experiments (electron transfer dissociation (ETD), followed by high-energy collision induced (HCD) fragmentations). By integration of experiments and ab initio calculations, we derive fragmentation pathways which rationalize our observations. We further validated both the liquid chromatography and the mass spectrometry method by analysis of synthetic peptides. In addition, we investigated if ?xHyp is differently present in metastatic tumors of colon rectal carcinoma.

MATERIALS AND METHODS

DATA COLLECTION

By re-analysis of bottom-up proteomics data uploaded at the PRIDE Archive (PXD008383), we were able to identify four new locations, based on MS^2 fragmentation, of ?xHyp. The mass spectrometric data retrieved from the PRIDE Archive (PXD008383) was obtained from formalin-fixed, paraffin-embedded healthy liver tissue. Sample processing is described by van Huizen et al[26]. Two healthy liver tissue samples and corresponding colorectal liver metastasis (CRLM) tissue (control-6, control-18, CRLM-6, and CRLM-18) were used for further identification of ?xHyp.

The publication by van Huizen et al.[26] showed a strong upregulation of COL1A2 in CRLM compared to liver tissue ($p= 9.4*10^{-10}$, fold change = 1.7). Based on their data we analyze in this manuscript if the ratio between COL1A2 containing the ?xHyp modification at location 584 and COL1A2 lacking the ?xHyp modification is different between CRLM (n=30 samples) and healthy adjacent liver tissue (n=30 samples). Therefore, we analyzed the ^{10}log-transformed data (the data is available in the supplemental information, S-1) for the ratio between the peptide containing ?xHyp GLHGEFGLP(4Hyp)GP(?xHyp)AGPR and peptide lacking ?xHyp GLHGEFGLP(4Hyp)GP(Pro)AGPR. In one liver sample (control-17), and in three CRLM samples (CLRM-4, CRLM-8, CRLM-14) the peptide containing ?xHyp could not be properly assigned and these data were therefore removed from data analysis. The overall COL1A2 levels of these four samples do not deviate from the other samples. Ratios are reported as an averaged ratio ± standard deviation. The difference between the ratios was tested for significance by use of a t-test (with assumption of unequal variance). A p-value below 0.05 was considered as significant.

MATERIALS

Synthetic peptides were ordered at Pepscan (Lelystad, the Netherlands) whereby the proline

at the Xaa position was modified: GLHGEFGLP(4Hyp)GP(Pro)AGPR, GLHGEFGLP(4Hyp)GP(3Hyp)AGPR, GLHGEFGLP(4Hyp)GP(4Hyp)AGPR. These synthetic peptides will be abbreviated below as GLH-Pro, GLH-3Hyp, GLH-4Hyp, respectively.

MASS SPECTROMETRIC ANALYSIS ON SYNTHETIC PEPTIDES BY DIRECT INFUSION

By direct infusion of the three synthetic peptides into an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fischer Scientific, San Jose, CA, USA), we set-up an ETD-HCD method to distinguish the three synthetic peptides; ETD is in MS^2 mode and subsequent HCD in MS^3 mode. Synthetic peptides were infused at a concentration of 0.02 mM in a 50:50 water:methanol, and 0.1% formic acid solution at a speed of 4 µL/min. Direct infusion was performed using an ESI-source with a spray voltage of 3.7 kV in the positive ion mode and kept at a temperature of 275 °C. Both doubly and triply charged ions were observed, of which the triply charged ions were selected for further ETD fragmentation. Triple charged ions were collected for 118 ms or until the AGC target of 5.0e4 was reached. Fluoranthene was used as an ETD reagent and was collected for 200 ms or until the AGC target of 2.0e5 was reached. Optimal reaction time between fluoranthene and triply charged peptide ions of interest was 200 ms. After ETD fragmentation of the triply charged ions at m/z 793.25 for GLH-Pro and m/z 798.58 for GLH-3Hyp and GLH-4Hyp, we observed charge-reduced radical cations, $[M+3H]^{2+\bullet}$. These charge-reduced peptide ions were further fragmented by HCD collision. HCD fragmentation was performed on the charge-reduced fragment ions with m/z 739.8 for GLH-Pro and m/z 498.58 for GLH-3Hyp and GLH-4Hyp. HCD collision energy was set to 35 %. ETD-HCD fragments were collected for 500 ms or until the AGC target of 5.0e5 was reached, followed by detection in the Orbitrap mass analyzer set at 120,000 resolution. The spectra shown in figure 5 have been averaged over 4 min.

AB INITIO CALCULATIONS

Ab initio calculations were performed to validate the ETD/HCD fragmentation. In the ab initio calculations, we used a model version of the peptides (as shown in figure 4) whereby the charged C- and N-termini were blocked with $NH\text{-}CH_3$ and $(C{=}O)\text{-}CH_3$, respectively. All model molecules were built using GaussView 6 and optimized with Gaussian 16 at the MP2 level of theory with the 6-31g(d,p) basis set [321]. For the open-shell stable species investigated here, the unprotected <S2> values were within 0.7501-0.7504 indicating negligible spin contamination. All presented molecular structures are fully optimized and minima are confirmed through the vibrational frequency analysis. Vibrational frequency calculations also provide zero-point energy correction (ZPE) which is applied to the energy displayed.

IDENTIFICATION BY NANOLC-ESI-ETD-HCD OF ?XHYP

Synthetic peptides were diluted to a concentration of 2 nM and mixed in a 1:1 ratio with

the liver tissue digest. The following mixtures were analyzed: the digest with 0.1% TFA, with GLH-Pro, with GLH-Pro + GLH-3Hyp, and with GLH-Pro + GLH-4Hyp. This was performed for control-6, control-18, CRLM-6, and CRLM-18 [26]. The same nanoLC method was used as described by van Huizen et al. [26], the only difference being that a 30 min gradient instead of a 90 min gradient was used. The nanoLC was coupled to a nanospray source of an Orbitrap Fusion. The above mentioned ETD-HCD mass spectrometry method was used for detection. To increase the number of MS^3 scans on a LC time scale, ions were analyzed in the ion trap of the Orbitrap Fusion. LC-MS/MS/MS runs were made publicly accessible via the PRIDE archive, accession number: PXD011784.

SEQUENCE LOGO

Sequence commonalities were studied by generating a sequence logo for the 41locations reported in literature [69-71, 73, 320] and the four additional locations reported in this study. Montgomery et al. reported the lysine to be present at the Yaa position next to ?xHyp.[73] We rejected the presence of ?xHyp next to lysine, hydroxylysine was not included in the database search and is therefore likely to be an artefact.[73] In total, 45 positions were taken into account for the sequence logo. The sequence of nine amino acids in front and after ?xHyp were selected for the generation of a sequence logo. ?xHyp was set to position "0" in the sequence logo. Hydroxylation of proline at other locations were not considered. The Sequence Logo was generated via an online tool (http://weblogo.berkeley.edu/logo.cgi).

6

RESULTS

REANALYSIS OF PRIDE DATABASE DATA

New Hyp at Xaa locations

Re-analysis of bottom-up data of liver tissue (PXD008383) resulted in the identification of four peptides containing ?xHyp, namely COL1A1: GP(?xHyp, pos. 417)SGPQGPGGPP(4Hyp)GPK; COL1A1: GPSGPQGP(?xHyp, pos. 423)GGPP(4Hyp)GPK; COL1A2: GLHGEFGLP(4Hyp)GP(?xHyp, pos. 584)AGPR; and COL1A2: GEAGAAGP(?xHyp, pos. 701)AGPAGPR. Peptides lacking the aberrant hydroxylation were also detected.

Sequence logo

The identity of the enzyme modifying Pro into ?xHyp is unknown. To get more insight into this unknown enzyme, we search for sequence commonalities by generating a sequence logo. The generated sequence logo (figure 1) was based on 41 ?xHyp locations reported in literature [69-71, 73, 320] and the additional four locations reported in this manuscript. Glycine is dominantly present at every third position due to the repeating unit. At the Yaa position

adjacent to ?xHyp (position 1), alanine (47%), or glutamine (16%) is frequently present. Glutamine is mainly observed in collagen type IV. Four out of four ?xHyp positions reported in collagen type IV contained glutamine at the Yaa (position 1). The repeating units adjacent to ?xHyp frequently contain alanine and proline.

Figure 1. Sequence logo based on 16 ?xHyp locations described in literature [69-71, 320] and the additional 4 reported in this manuscript. ?xHyp was set at position "0".

?xHyp in colorectal liver metastasis

We used proteomics data of colorectal liver metastasis (n=30) and healthy adjacent liver tissue (n=30) as described[26] to calculate the ratio between the peptide containing ?xHyp and the peptide lacking ?xHyp. It was found in that the ratio in liver tissue was 0.54 ± 0.11 while the ratio in CRLM was 0.31 ± 0.07 (figure 2). Thus the ratio in healthy adjacent liver tissue is significant different from CRLM tissue (p-value = $1*10^{-9}$). The decreased ratio in CRLM is caused by a strong significant increase of the peptide lacking ?xHyp from 6.8 ± 0.3 in liver to 7.4 ± 0.4 in CRLM (p-value = $2*10^{-8}$), while the peptide containing ?xHyp increased significantly but only slightly from 6.5 ± 0.3 in liver to 6.8 ± 0.4 in CRLM (p-value = $9*10^{-4}$).

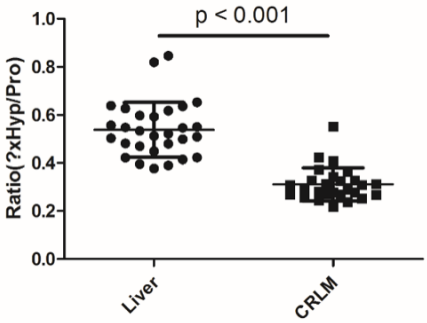


Figure 2. The ratio between the modified and unmodified peptides (GLHGEFGLP(4Hyp)GP(Pro)AGPR and GLHGEFGLP(4Hyp)GP(?xHyp)AGPR, respectively)

In figure 2 two liver samples have a higher ratio (Control-16, Control-19) than the others and in the CRLM one sample (CRLM-24) has a deviating ratio. These three samples do not have deviating COL1A2 levels.

MASS SPECTROMETRY ANALYSIS ON SYNTHETIC PEPTIDES BY DIRECT INFUSION

Partial mass spectra, acquired by ETD-HCD (MS^3) fragmentation for GLH-Pro, GLH-3Hyp, and GLH-4Hyp are shown in figure 3. Comparison of these ETD-HCD fragmentation mass spectra of GLH-3Hyp and GLH-4Hyp reveals that the m/z 400 signal is more intense for GLH-3Hyp.The m/z 454 signal is not observed for GLH-3Hyp, but is present for GLH-4Hyp. While the peak at m/z 400 corresponds to the y4 fragment ion ($AGPR^+$), the peak at m/z 454 does not match any a-, b-, c-, x-, y-, or z-ions.

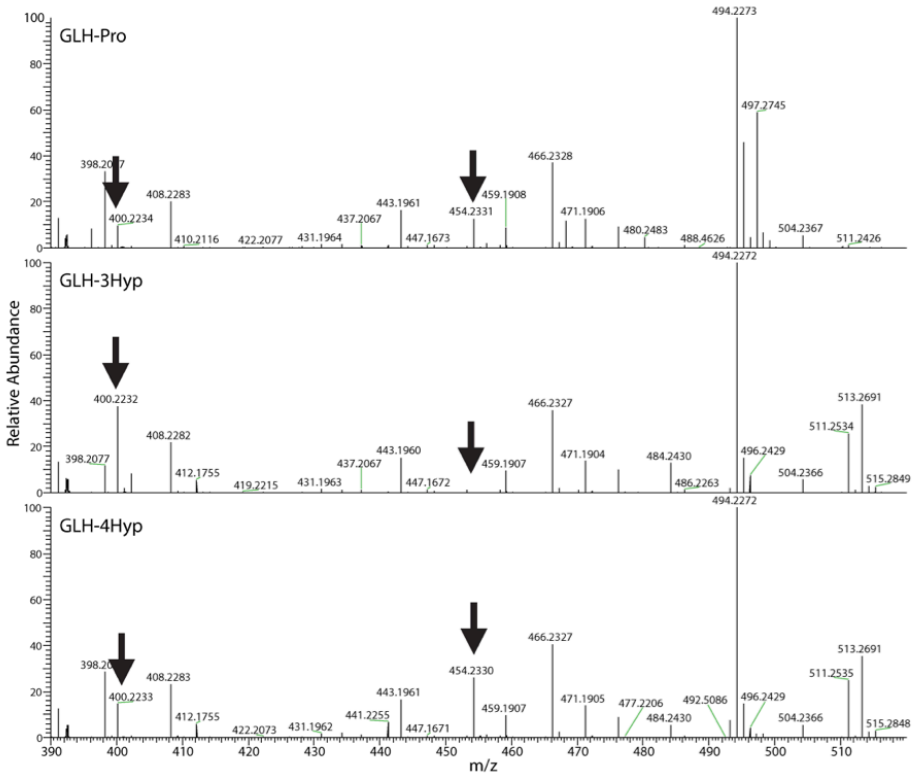


Figure 3. Zoom-in of the ETD-HCD mass spectra of the synthetic peptides acquired by direct infusion. GLH-3Hyp and GLH-4Hyp are distinguishable by singly charged fragments at m/z 400 and 454.

We propose a fragmentation pathway (figure 4) for GLH-3Hyp and GLH-4Hyp which explains the difference observed between GLH-3Hyp and GLH-4Hyp fragmentations in the synthetic peptides. Electron transfer dissociation (ETD) of the triply charged ions results in the neutralization of the proline moiety, whereby a hypervalent nitrogen species is formed which undergoes immediate ring opening [322] leading to a carbon-centered radical. The latter species does not dissociate further under ETD conditions. The doubly-charged ring-opened structure generated by ETD is subsequently activated by high energy dissociation (HCD). The activation of the ring-opened structure results in β-cleavage at the carbon-centered radical, as shown in figure 4. This process produces telltale fragment ions detected at m/z 470 for GLH-3Hyp and at m/z 454 for GLH-4Hyp. Fragment ions at m/z 454 are indeed observed in the MS^3 spectra of the GLH-4Hyp peptide ions, but the m/z 470 ions are not detected in the GLH-3Hyp peptide ions. Also, we do not observe the complementary fragments at m/z 1024 and 1040 for GLH-3Hyp and GLH-4Hyp, respectively. The high mass part of the mass spectrum mainly shows intense C- and Z-ions.

To get further mechanistic insight on the collision-induced dissociation of the ETD generated radical ions, we calculated at the ab initio level of theory the reaction enthalpies for these dissociation processes. We used model molecules presenting the carbon-centered radical proline ring, respectively 3Hyp an 4Hyp in figure 4 with R_1 and R_2 being methyl groups. The relative energies for the HCD fragmentations are shown below the structures, see figure 4. The initial fragmentation of the collisional activated model 3Hyp radicals requires 88 kJ/mol starting from the ring-opened structure, whereas the same fragmentation for the model 4Hyp radicals is 138 kJ/mol endothermic. Thus, m/z 470 ions are expected to be detected in the ETD/HCD mass spectrum of the GLH-3Hyp ions since the corresponding ions, generated by a more energy demanding process, are clearly detected at m/z 454 in the ETD/HCD mass spectrum of the GLH-4Hyp ions.

However, our ab initio calculations indicate that the m/z 470 ions can consecutively undergo a rearrangement initiated by a 1,5-hydrogen shift to eliminate 3-oxoprop-2-enal to produce m/z 400 ions as observed in the MS^3 spectrum in figure 3. This secondary process only requires 146 kJ/mol and is thus expected to occur. Such a process is not expected to occur for the GLH-4Hyp ions. When comparing the three mass spectra of figure 3, the higher intensity of the m/z 400 signal in the specific case of the GLH-3Hyp peptide may be ascribed to the presence of the consecutive m/z 470 to m/z 400 decomposition.

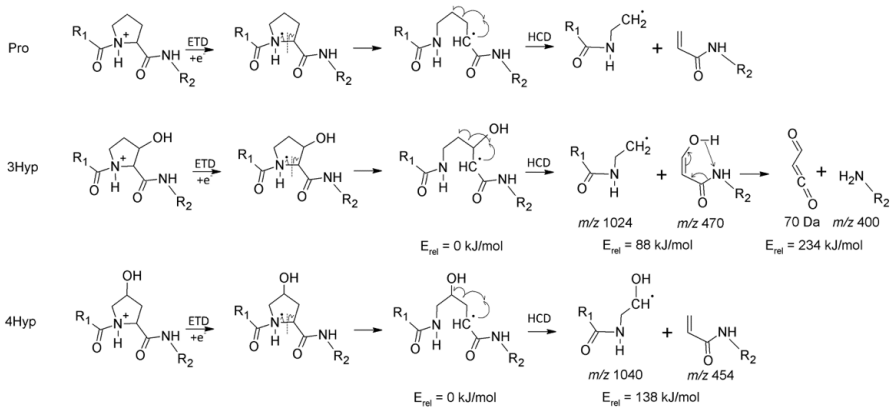


Figure 4. Proposed ETD-HCD fragmentation pathway. R_1 = [GLHGEFGLP(4Hyp)G + H]$^+$, R_2 = [AGPR + H]$^+$ with corresponding m/z values shown below the structures which refer to figure 3. For the model compound used for ab initio calculations $R_1 = R_2 = CH_3$. Energy levels calculated by our ab initio method are shown below the corresponding structures. Heat of formation of the HCD fragments are calculated relative to the third structure, where the proline backbone has been broken by ETD to yield a carbon-centered radical.

?XHYP IDENTIFICATION BY NANOLC-ESI-ETD-HCD

Identification of ?xHyp in the mixtures of liver tissue digest and synthetic peptides was performed by nanoLC-ESI-ETD-HCD analysis. In figure 5, the chromatograms obtained for the different mixtures for "control-18" are shown. In figure 5A, the top chromatogram corresponds to the trypsin digest of liver tissue mixed with TFA, showing the retention time of peptide with ?xHyp. A zoom-in of the MS^3 mass spectra is shown in figure 5C, this spectrum shows the characteristic peak m/z 454 and low intensity for m/z 400, therefore matching with the mass spectrum of GLH-4Hyp. In addition, the retention time of GLH-4Hyp (bottom chromatogram) overlaps with the peptide present in the digest, while GLH-3Hyp elutes 0.5 min later. That GLH-3Hyp and GLH-4Hyp have such different retention times comes as a surprise to us and this finding alone allows differentiation of GLH-3Hyp and GLH-4Hyp. Hydroxylation of ?xHyp is not complete, hence we also observe the peptide containing unmodified proline. Retention times of GLH-Pro and the peptide originating from the digest containing Pro are identical. The other samples analyzed (Control-6, CRLM-18, and CRLM-6) gave similar results as shown in figure 5.

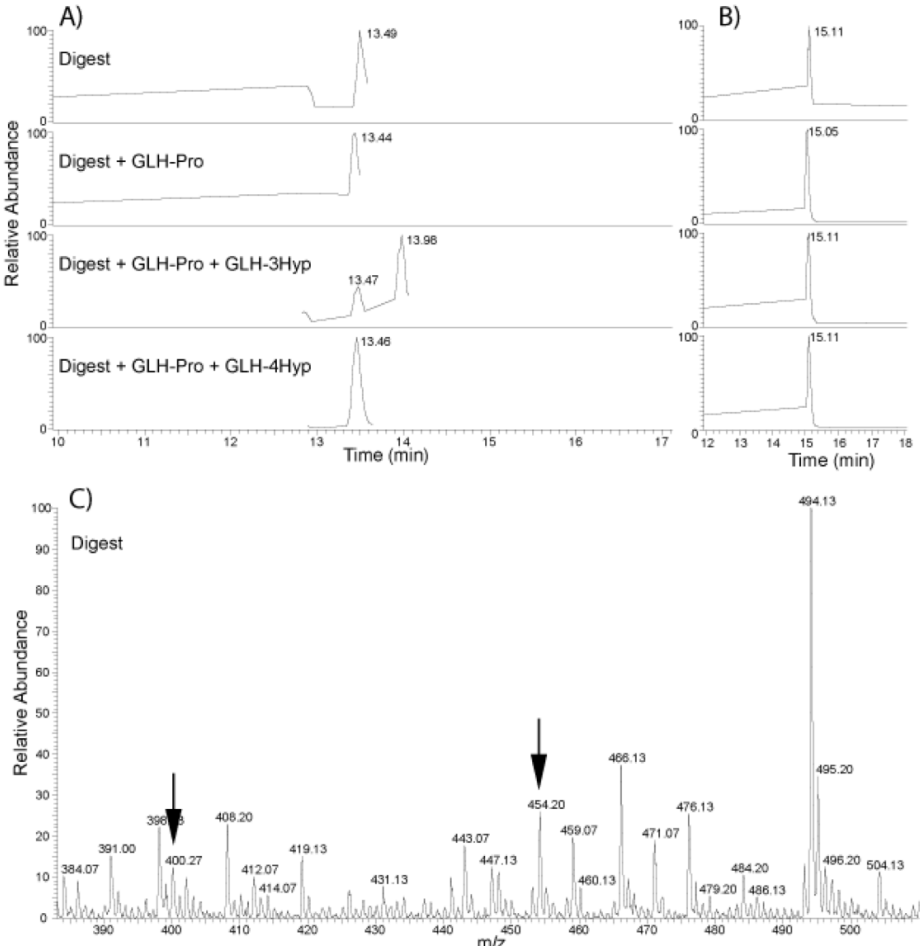


Figure 5. Identification of ?xHyp in sample control-18 based on LC retention time differences and MS^3 fragmentation. Part A) and B) show retention times based on extracted ion chromatography for m/z 498.57 and m/z 493.24, respectively, whereby chromatograms in B) originate from the same samples as A). The following peptides are seen in A) GLH-?xHyp (t_R = 13.49 min), GLH-3Hyp (t_R = 13.96 min), and GLH-4Hyp (t_R = 13.47 min), and in B) GLH-Pro (t_R = 15.11 min). C) Zoom-in of the MS^3 mass spectra belonging to peptide present in the digest containing ?xHyp, top chromatogram in A).

DISCUSSION

We conclude that the unknown ?xHyp at position 584 in COL1A2 is 4Hyp. We do not have any evidence for 3Hyp to be present at the Xaa position; however we cannot exclude the possibility that a small amount of 3Hyp is present below the detection limit. To prevent confusion with

respect to the nomenclature of collagen PTMs, we propose to address 4Hyp at the Xaa position as 4xHyp. The identification of in vivo 4xHyp is based on both retention time differences and mass spectrometry analysis. We have validated the proposed fragmentation scheme by calculating, with an ab initio method, the energy required for the observed fragmentations. We consider the required energy as relatively low, considering that the dissociation of a carbon-carbon bond requires about 350 kJ/mol. Therefore, HCD readily activates the further dissociation of the ring-opened structure.

Since we showed the identification of 4xHyp only in one position as a proof of principle, we cannot conclude that all 45 known positions of ?xHyp are 4xHyp. Hydroxyproline is not exclusively present in collagen and also occurs in other proteins. Spahr et al. describes the presence of hydroxyproline in a GGGGP linker of an FC-fusion protein.[323] However, the identity 3- or 4-hydroxyproline remains unknown. Zhou et al. report 562 sites in 272 proteins in HeLa cells were hydroxyproline is present.[324] Zhou et al. assume that the found hydroxyprolines are 4-hydroxyprolines, but no evidence is provided with respect to the identity of the hydroxyprolines. The described method can also be used to identify 3- or 4-hydroxyproline in non-collagen proteins. In collagen, sequence commonalities are observed between different locations, tissue types, and organisms which make it likely that all ?xHyp are in fact 4xHyp. Collagen type IV is a possible exception, where glutamine instead of alanine is dominantly present at position 1. The known enzyme substrate[304] for prolyl 4-hydroxylase does not match these sequence commonalities and this indicates the presence of an unknown substrate for prolyl 4-hydroxylase or an unknown substrate for a not yet identified enzyme with this function. This could be investigated by offering a guest peptide containing Gly-Pro-Ala or Gly-Pro-Gln to prolyl 4-hydroxylase or liver lysates. Such an experiment was previously performed to study the effect of an amino acid at the Xaa position on the hydroxylation rate of proline into 4Hyp.[79] It was shown that an amino acid at the Xaa position has a major influence on the hydroxylation rate of proline at the Yaa position. Furthermore, modification of proline into 4xHyp is not observed in all COL1A2 molecules, in liver tissue the peptide containing unmodified proline is present at a significant higher level than the peptide containing 4xHyp. In CRLM tissue, the unmodified peptide is at significant higher levels present, which is not surprising because COL1A2 is strongly upregulated in CRLM compared to liver tissue ($p= 9.4*10^{-10}$, fold change = 1.7) [26]. However, the peptide containing 4xHyp did increase significantly but not as strong as the peptide lacking ?xHyp. The calculated ratio between peptides containing 4xHyp and lacking 4xHyp is significantly different between CRLM and liver tissue, and this clearly distinguishes the adjacent liver from CRLM tissue.

The role of 4xHyp is unknown, although it has been shown that 4Hyp stabilizes the triple

helix [306]. In general, 4Hyp provides the required thermal stability to form a triple helix at body temperature by forming hydrogen bonds with its surroundings. However, an amino acid at the Yaa position has less interaction with its surroundings than an amino acid at the Xaa position where 4xHyp is located [325]. Therefore, a different stabilization effect can be expected. Further investigation is required to identify the role of 4xHyp in stabilizing the triple helix.

The R configuration of the hydroxyl group in 4Hyp, included in the (2S,4R)-4-hydroxyproline moiety, is crucial for the stabilization of the triple helix [326]. However, it is unknown how 4xHyp is formed, so we cannot assume that the configuration is similar to 4Hyp. Further research is required to determine the configuration of the hydroxyl group.

The understanding of collagen PTMs is crucial for the understanding of the role of collagen in, among others, tissue formation, cell attachment, tissue support, tumor proliferation, and handling physical stress. The identification of ?xHyp as 4xHyp at position 584 in COL1A2 brings us another step closer to the understanding of collagen PTMs and possibilities to apply this knowledge for understanding tumor metastasis and generate possibilities for diagnosis of cancer.

ACKNOWLEDGMENTS

The UMONS MS laboratory acknowledges the Fonds National de la Recherche Scientifique (FRS-FNRS) for continuing support. FS thanks the UMONS for a PhD grant. Computational resources for calculations in the Laboratory for Chemistry of Novel Materials have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11.

CHAPTER 7

General discussion
Summary
Samenvatting

GENERAL DISCUSSION

In cancerous tissue, collagen can be strongly modified at the molecular level as well as at the supramolecular level. Changes in the extracellular matrix, of which collagen is a major component, may stimulate tumor proliferation and even induce tumorigenesis.[24, 269, 327-330] The opposite is also true: collagen may inhibit tumor cells, and even prevent cells from becoming malignent.[24, 269] Collagen changes have mostly been studied at the protein or supramolecular structure level but specific studies including posttranslational modifications in different types of cancer are lacking. Nonetheless, changes at the posttranslational modification level will strongly affect the supramolecular structure organization; leading to dysregulated collagen, as described in **Chapter 2**. For example, 3-hydroxyproline, 4-hydroxyproline, and glycosylation influence triple helix formation, inter-triple helix interaction, and cross-linking. On top of that, a change of the triple helix composition can have severe consequences. Studies have shown that collagen type 1 homotrimers are immune for MMPs and enhance tumor mobility[331], whereas the loss of specific collagen type IV alpha chains correlates with tumor differentiation.[332] Within a tumor large differences in collagen that correlate with different tumor characteristics maybe be present.[333]

Collagen is affected in CRLM [152, 334], and this is reflected in serum [178, 335] and urine [11, 12]. The combination of collagen in urine and carcinoembryonic antigen (CEA) in serum showed to be very promising to detect CRLM.[11, 12] CEA is already used in the detection of CRLM, however the sensitivity (68%) is too low.[14]The aim of this thesis is to improve analysis of collagen NOPs in urine and combined with serum CEA in order to detect CRLM. The advantage of urine is that it can be obtained non-invasively. Larger volumes of urine can be easily collected also outside the hospital. In addition, there is no need for imaging in the hospital, With the use of urine, analysis can be performed at a higher frequency than with the current techniques with the advantage of potentially detecting metastasis at an earlier stage.

We describe the possibility of detecting CRLM with a high sensitivity and specificity through a combination of a collagen NOP, GNDGARGSDGQPGPP(-OH)GP(-OH)P(-OH)GTAGFP(-OH) GSP(-OH)GAK(-OH)GEVGP, in urine and CEA in serum. Besides GND, also other collagen NOPs are significantly different between urine of healthy controls and patients with CRLM. Collagen NOPs might originate from different processes including: 1) collagen turnover inside the tumor; 2) collagen turnover in healthy tissue in close proximity of the tumor; or 3) a more unlikely coincidental non-related tumor process. The following checks should be performed to confirm that collagen NOPs selected for validation are related to CRLM: 1) The collagen NOP and corresponding hydroxylation pattern should be differentially present in CRLM and control tissue; and 2) the collagen NOP should belong to a collagen type that is

significantly different between healthy and tumor tissue. The presence of these NOPs in CRLM tissue does not prove that the NOP is caused by the tumor, but shows at least an association.

In this thesis we have shown that most collagen types appeared to be affected by CRLM: the collagen abundance is significantly upregulated in CRLM compared to liver tissue; specific collagen type levels are more comparable to colon tissue than to liver tissue; and the degree of collagen hydroxylation in cancer is significantly lower than that in normal tissue. In addition, we identified 4-hydroxyproline at the exceptional Xaa-position (4xHyp) instead of at the Yaa position. Collagen affected by CRLM contained also less 4xHyp compared to healthy liver tissue.

A proper understanding of the collagen composition and hydroxylation mechanisms in CRLM is important and could be extended to other types of cancer. For diagnostic purposes, it would be interesting to study, for example, the change in collagen composition and the degree of hydroxylation at the transition zone between CRLM and healthy adjacent liver tissue. Especially studying differences between the various known growth patterns observed in CRLM is of interest.[333] Technically, this could be performed by selecting tissue corresponding to the different growth patterns with laser microdissection microscopy, followed by mass spectrometry analysis. In addition, electron microscopy can aid in visualizing the effects of CRLM on the supramolecular organization of collagen.

PERSPECTIVE

We have demonstrated that urine is a very promising body fluid to detect CRLM based on collagen NOPs. Currently, the method described in **Chapter 3** has a similar or an even better sensitivity and specificity as the currently-used techniques. We foresee that detection of CRLM based on collagen NOPs in urine will replace present techniques as first line of detection. Most ideally, an improvement will be made by which CRLM detection can be solely based on urine instead of a combination with serum.

Maxim et al. give an overview of 142 different screening tests for various diseases; the overall sensitivity is 94.9% (median; interquartile range (IQR) [85.1-99.1]), and the specificity is 80.1% (median; IQR [67.4-89.5]).[336] The sensitivity was higher than the specificity in almost three-quarters of the screening tests. To prevent false negatives, high sensitivity is favored over a high specificity. The test we have developed is intended for patients at risk and is therefore not necessarily a population screening test; therefore, a somewhat lower specificity can be acceptable for these patients at risk.

To reach levels of more than 95% sensitivity and specificity, additional collagen NOPs in urine

need to be screened for their value as molecular marker. We reported over 100 significant different collagen NOPs, and, so far, only three peptides (AGP, GPP, and GND) have been validated. This implicates that still many promising candidates are left to be tested. Candidate NOPs should be selected on the basis of the two criteria described above, and a third criteria should be added: the NOP should have a low degree of hydroxylation, because collagen hydroxylation is significantly downregulated in tumor tissue. Easy screening methods to tests all these molecules should be developed to reach most optimal sensitivities and specificities.

Collagen NOPs in urine of patients with CRLM, which have a lower hydroxylation pattern than in healthy tissue, will revert to normal levels after tumor resection. Up to 70% of the liver can be removed; the liver can regenerate and will do so over the course of a year.[337] Surgery will cause scarring of tissue, which takes more than a year to heal.[338] Collagen is strongly involved in the processes of regeneration and tumor formation, which results in higher collagen turnover and remodeling. It is essential to get longitudinal data of patients at risk; with the information deduced from these data we might improve the sensitivity and specificity of the test.

The European Union supports our COLO-MET project, which allows us to continue investigating collagen NOPs in urine to detect CRLM, and to bring the test a step closer to clinical implementation (https://www.eithealth.eu/en_US/colo-met). We hope that we can demonstrate that an urine-based detection method is able to detect CRLM at an earlier stage than the currently used techniques. Earlier detection would benefit the patient, because smaller tumor size and smaller number of tumors may affect decission making in the management of a patient developing CRLM.[339-341]

A better insight in tumor biology of collagen may help to optimize treatment strategies, thus improving the outcome for people with CRLM.

SUMMARY

The research described in this thesis focusses on the detection of collagen natural occurring peptides (NOPs) in urine in order to improve detection of colorectal liver metastasis (CRLM). In addition, to understand the observed differences in urine levels of NOPS, collagen was studied in organs affected by CRLM.

Chapter 2 contains a review of collagen and collagen analysis with mass spectrometry. Collagen is a family of 45 different proteins, called alpha chains, which are divided over 28 collagen types. Each collagen type consists of up to six different unique alpha chains. Alpha-chains contain a region (triple helical region) in the primary amino acid sequence that consists of the repeating unit $[\text{-Gly-Xaa-Yaa-}]_n$, wherein Xaa and Yaa can be different amino acids. The triple helical region makes it possible for three collagen alpha chains to form a triple helix. Another characteristic of collagen is the extensive posttranslational modification (PTM) of proline and lysine. PTMs are crucial for triple helix formation and the formation of large supramolecular structures. Mass spectrometry is a versatile technique that can be used to analyze collagen, more specifically to analyze the amino acid sequence, proline and lysine PTMs, as well as the general composition in various body fluids (e.g. serum, urine) and tissue types (e.g. liver, bone). Mass spectrometry can also be used to quantitatively measure collagen concentrations in various diseases.

In **Chapter 3** we described the identification and validation of collagen NOPs in urine as a way to improve the detection of CRLM. Collagen NOPs were identified in urine with a standard bottom-up proteomics approach. The most distinctive NOPs were selected and validated with a targeted mass spectrometry method. A newly validated collagen NOP in combination with carcinoembryonic antigen levels in blood formed a new molecular marker panel with better predictive power than the currently used techniques.

In **Chapter 4** we described the compositional difference of collagen between liver and CRLM tissue. In general, collagen is significantly upregulated in CRLM compared to liver tissue. Also proteins related to collagen synthesis are mostly upregulated. The composition of collagen in CRLM tissue differs from that in liver tissue. The abundance of collagen types 10, 12, 14, and 15 is more similar to colon tissue than to liver tissue, and could have migrated via the primary tumor (colorectal cancer, CRC) to CRLM.

In the data presented in **Chapter 4**, we observed – in contrast to a higher abundance of collagen – a decrease in the hydroxylation degree of collagen in CRLM compared to liver tissue. In **Chapter 5** we analyzed, in more detail, the degree of collagen alpha-1(I)

hydroxylation in colon, CRC, liver, and CRLM tissues. We confirmed the significant decrease in the hydroxylation degree of collagen alpha-1(I) in CRLM and CRC, compared to liver and colon tissue, respectively. Furthermore, it appeared that within patients' healthy tissue, collagen alpha-1(I) is present with different hydroxylation forms. The degree of hydroxylation of collagen alpha-1(I) in colon and liver tissue did not differ. However, locally significant differences were present. In addition, we showed that peptides with a specific degree of hydroxylation that are significantly different in tissue, can be differentially present in urine.

In the data presented in **Chapter 4**, an unexpected proline PTM was observed at the Xaa position. In literature, multiple locations have been described in which this unexpected proline PTM occurs. However, the identity of this proline PTM remained unknown. **Chapter 6** describes the proof-of-principle identification of 4-hydroxyproline at position 584 (4xHyp) in collagen alpha-2(I). We have developed a mass spectrometry method which can differentiate between 3-hydroxyproline and 4-hydroxyproline. This method consists of serial fragmentation of a peptide containing 4xHyp with electron transfer dissociation (ETD) and higher-energy collisional dissociation (HCD). First the proline backbone is broken with ETD, followed by complete fragmentation of the proline ring with HCD. 4xHyp was observed both in liver tissue and in CRLM tissue. Similar levels of 4xHyp are present in both tissue types, while the overall collagen content in CRLM was strongly increased. Therefore, the ratio between collagen with and without 4xHyp was lower in CRLM tissue.

SAMENVATTING

Het onderzoek beschreven in dit proefschrift betreft de detectie van natuurlijk voorkomende collageenpeptiden (collageen-NOPs) in urine, waarmee de detectie van colorectale lever metastase (CRLM) kan worden verbeterd. Om de verschillen in collageen-NOPs in urine te kunnen begrijpen, hebben we ook collageen bestudeerd in organen waar CRLM zich manifesteert.

In **Hoofdstuk 2** wordt een overzicht gegeven van de eigenschappen van collageen, en de analyse van collageen met behulp van massaspectrometrie. Collageen is een familie van 45 verschillende eiwitten, alpha chains genaamd die onderverdeeld kunnen worden in 28 typen. Elk collageen type kan bestaan uit 1 t/m 6 verschillende unieke alpha chains. In de aminozuursequentie van elke alpha-chain zit een gebied dat bestaat uit $[\text{-Gly-Xaa-Yaa-}]_n$, waarbij Xaa en Yaa verschillende aminozuren kunnen zijn. Dit gebied maakt het mogelijk dat drie alpha chains samen een triple helix kunnen vormen. De uitgebreide posttranslationele modificaties (PTMs) van lysine en proline vormen een ander kenmerk van collageen. PTMs zijn cruciaal voor de vorming van collageen triple helices, en de vorming van grote supramoleculaire structuren. Massaspectrometrie is een veelzijdige techniek die uitermate geschikt is om collageen op moleculaire schaal te analyseren. Bijvoorbeeld: de aminozuursequentie, proline en lysine PTMS, de samenstelling in verschillende lichaamsvloeistoffen (bijv. urine en serum), en de samenstelling in verschillende soorten weefsels (bijv. lever, darm, tumor). Massaspectrometrie kan ook gebruikt worden voor de kwantitatieve analyse van collageen, al dan niet in verband met verschillende ziektes.

In **Hoofdstuk 3** beschrijven we de identificatie en validatie van collageen-NOPs in urine om de detectie van CRLM te verbeteren. De collageen-NOPs zijn geïdentificeerd door middel van een standaard bottom-up proteomicsmethode. De meest onderscheidende NOPs zijn geselecteerd voor verdere validatie met een kwantitatieve massaspectrometriemethode. Een gevalideerd collageen-NOP gecombineerd met carcino-embryonaal antigen (CEA) in serum vormt het nieuwe moleculaire marker panel – met een betere voorspellende waarde dan verkregen met de technieken die momenteel in het ziekenhuis worden gebruikt.

In **Hoofdstuk 4** beschrijven we het verschil in de samenstelling van collageen tussen CRLM en leverweefsel. In het algemeen is collageen significant verhoogd in CRLM- vergeleken met leverweefsel. Ook zijn de eiwitten die gerelateerd zijn aan de vorming van collageen verhoogd. Collageen in CRLM-weefsel heeft een andere samenstelling dan collageen in leverweefsel. De expressieniveaus van collageen type 10, 12, 14, en 15 lijken sterker op darm- dan op leverweefsel. Het is mogelijk dat de expressie van deze collageen types via de primaire

tumor (CRC) is gemigreerd naar CRLM.

In de data verzameld in **Hoofdstuk 4** hebben we ontdekt dat, in tegenstelling tot de verhoogde expressie van collageen, de mate van collageenhydroxylatie verlaagd is in CRLM-weefsel ten opzichte van leverweefsel. In **Hoofdstuk 5** hebben we de mate van collageen alpha-1(I) hydroxylatie nader onderzocht in darm-, lever-, CRLM-, en CRC-weefsel. Hieruit volgde dat collageen inderdaad in mindere mate gehydroxyleerd is in CRLM- en CRC-weefsel vergeleken met respectievelijk lever- en darmweefsel. Uit onderzoek bleek dat zelfs binnen weefsel van eenzelfde patiënt meerdere hydroxylatievormen van collageen voorkomen. Over het geheel genomen verschilt de mate van hydroxylatie niet tussen darm- en leverweefsel, maar op bepaalde plaatsen zijn wel significante verschillen aanwezig. Vervolgens laten we ook zien dat het mogelijk is de verschillen die in urine gezien worden te koppelen aan het verschil in de mate van collageenhydroxylatie tussen CRLM- en leverweefsel.

In de data verzameld in **Hoofdstuk 4** hebben we een onbekende proline PTM gevonden op de Xaa-positie. In de literatuur is de aanwezigheid hiervan op verschillende posities meerdere keren beschreven, maar de identiteit was tot nu toe onbekend. In **Hoofdstuk 6** laten we zien dat het mogelijk was deze te identificeren als 4-hydroxyproline (4xHyp) in collageen alpha-2(I) positie 584. We hebben een massaspectrometriemethode ontwikkeld die onderscheid kan maken tussen 3- en 4-hydroxyproline. Deze methode bestaat uit het fragmenteren met behulp van ETD en HCD fragmentatie technieken. Met ETD wordt de backbone van proline gebroken, en verdere fragmentatie vindt plaats door middel van HCD. 4xHyp is aanwezig in zowel lever- als in CRLM-weefsel. Beide weefsels bevatten ongeveer evenveel 4xHyp, maar collageen is sterk verhoogd in CRLM-weefsel. Hierdoor was de ratio tussen collageen dat wel 4xHyp bevat en collageen dat niet 4xHyp bevat significant lager in CRLM-weefsel.

REFERENCES

REFERENCES

1. J. Ferlay, M. Colombet, I. Soerjomataram, T. Dyba, G. Randi, M. Bettio, A. Gavin, O. Visser, F. Bray, Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018, Eur J Cancer 103 (2018) 356-387.
2. A. Figueredo, R.B. Rumble, J. Maroun, C.C. Earle, B. Cummings, R. McLeod, L. Zuraw, C. Zwaal, C. Gastrointestinal Cancer Disease Site Group of Cancer Care Ontario's Program in Evidence-based, Follow-up of patients with curatively resected colorectal cancer: a practice guideline, BMC Cancer 3 (2003) 26.
3. A. Al-Asfoor, Z. Fedorowicz, M. Lodge, Resection versus no intervention or other surgical interventions for colorectal cancer liver metastases, Cochrane Database Syst Rev (2) (2008) CD006039.
4. E. Gregoire, E. Hoti, D.L. Gorden, S. de la Serna, G. Pascal, D. Azoulay, Utility or futility of prognostic scoring systems for colorectal liver metastases in an era of advanced multimodal therapy, Eur J Surg Oncol 36(6) (2010) 568-74.
5. I. Grossmann, G.H. de Bock, C.J. van de Velde, J. Kievit, T. Wiggers, Results of a national survey among Dutch surgeons treating patients with colorectal carcinoma. Current opinion about follow-up, treatment of metastasis, and reasons to revise follow-up practice, Colorectal Dis 9(9) (2007) 787-92.
6. A.I. Valderrama-Trevino, B. Barrera-Mera, J.C. Ceballos-Villalva, E.E. Montalvo-Jave, Hepatic Metastasis from Colorectal Cancer, Euroasian J Hepatogastroenterol 7(2) (2017) 166-175.
7. G.Y. Locker, S. Hamilton, J. Harris, J.M. Jessup, N. Kemeny, J.S. Macdonald, M.R. Somerfield, D.F. Hayes, R.C. Bast, Jr., Asco, ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer, J Clin Oncol 24(33) (2006) 5313-27.
8. S. Pita-Fernandez, M. Alhayek-Ai, C. Gonzalez-Martin, B. Lopez-Calvino, T. Seoane-Pillado, S. Pertega-Diaz, Intensive follow-up strategies improve outcomes in nonmetastatic colorectal cancer patients after curative surgery: a systematic review and meta-analysis, Ann Oncol 26(4) (2015) 644-56.
9. J. Lemke, G. Cammerer, J. Ganser, J. Scheele, P. Xu, S. Sander, D. Henne-Bruns, M. Kornmann, Survival and Prognostic Factors of Colorectal Liver Metastases After Surgical and Nonsurgical Treatment, Clin Colorectal Cancer 15(4) (2016) e183-e192.
10. K. Wang, W. Liu, X.L. Yan, J. Li, B.C. Xing, Long-term postoperative survival prediction in patients with colorectal liver metastasis, Oncotarget 8(45) (2017) 79927-79934.
11. M.E.E. Broker, Z.S. Lalmahomed, H.P. Roest, N.A. van Huizen, L.J.M. Dekker, W. Calame, C. Verhoef, J.N.M. IJzermans, T.M. Luider, Collagen Peptides in Urine: A New Promising Biomarker for the Detection of Colorectal Liver Metastases, Plos One 8(8) (2013) e70918.
12. Z.S. Lalmahomed, M.E.E. Broker, N.A. van Huizen, R. van den Braak, L.J.M. Dekker, D. Rizopoulos, C. Verhoef, E.W. Steyerberg, T.M. Luider, J.N.M. IJzermans, Hydroxylated collagen peptide in urine as biomarker for detecting colorectal liver metastases, Am J Cancer Res 6(2) (2016) 321-330.
13. J. Kievit, Follow-up of patients with colorectal cancer: numbers needed to test and treat, Eur J Cancer 38(7) (2002) 986-99.
14. E. Tan, N. Gouvas, R.J. Nicholls, P. Ziprin, E. Xynos, P.P. Tekkis, Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer, Surg Oncol 18(1) (2009) 15-24.
15. H. Lodish, A. Berk, S.L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, Molecular Cell Biology, Chapter

R

22.3, W.H. Freeman, New York, United states of America, 2000.

16. W. Friess, Collagen–biomaterial for drug delivery, Eur J Pharm Biopharm 45(2) (1998) 113-36.
17. W. Hoehenwarter, N.M. Kumar, M. Wacker, U. Zimny-Arndt, J. Klose, P.R. Jungblut, Eye lens proteomics: from global approach to detailed information about phankinin and gamma E and F crystallin genes, Proteomics 5 (2005) 245-257.
18. W. Hoehenwarter, J. Klose, P.R. Jungblut, Eye lens proteomics, Amino Acids 30(369-389) (2006).
19. Y.J. Lee, R.H. Rice, Y.M. Lee, Proteome analysis of human hair shaft: from protein identification to posttranslational modification, Mol Cell Proteomics 5(5) (2006) 789-800.
20. R.H. Rice, Proteomic analysis of hair shaft and nail plate, J Cosmet Sci 62(2) (2011) 229-36.
21. J. Myllyharju, K.I. Kivirikko, Collagens, modifying enzymes and their mutations in humans, flies and worms, Trends Genet 20(1) (2004) 33-43.
22. M. van der Rest, R. Garrone, Collagen family of proteins, Faseb J 5(13) (1991) 2814-23.
23. C. Frantz, K.M. Stewart, V.M. Weaver, The extracellular matrix at a glance, J Cell Sci 123(Pt 24) (2010) 4195-200.
24. M. Fang, J. Yuan, C. Peng, Y. Li, Collagen as a double-edged sword in tumor progression, Tumour Biol 35(4) (2014) 2871-82.
25. G.F. Xiong, R. Xu, Function of cancer cell-derived extracellular matrix in tumor progression, J Cancer Metastasis Treat 2 (2016) 357-64.
26. N.A. van Huizen, R. Coebergh van den Braak, M. Doukas, L.J. Dekker, J.N.M. IJzermans, T.M. Luider, Up-regulation of collagen proteins in colorectal liver metastasis compared with normal liver tissue, J Biol Chem 294(1) (2018) 281-289.
27. M.A. Karsdal, S.H. Nielsen, D.J. Leeming, L.L. Langholm, M.J. Nielsen, T. Manon-Jensen, A. Siebuhr, N.S. Gudmann, S. Ronnow, J.M. Sand, S.J. Daniels, J.H. Mortensen, D. Schuppan, The good and the bad collagens of fibrosis - Their role in signaling and organ function, Adv Drug Deliv Rev 121 (2017) 43-56.
28. Y.M. Michelacci, Collagens and proteoglycans of the corneal extracellular matrix, Braz J Med Biol Res 36(8) (2003) 1037-46.
29. R.A. Harvey, D.R. Ferrier, Lippincott's illustrated reviews, Biochemistry, 5 ed., Wolters Kluwer2011.
30. J. Vuust, K.A. Piez, A kinetic study of collagen biosynthesis, J Biol Chem 247(3) (1972) 856-62.
31. S.J. Eyles, Proline not the only culprit?, Nat Struct Biol 8(5) (2001) 380-1.
32. H.P. Bachinger, The influence of peptidyl-prolyl cis-trans isomerase on the in vitro folding of type III collagen, J Biol Chem 262(35) (1987) 17144-8.
33. L. Bonfanti, A.A. Mironov, Jr., J.A. Martinez-Menarguez, O. Martella, A. Fusella, M. Baldassarre, R. Buccione, H.J. Geuze, A.A. Mironov, A. Luini, Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation, Cell 95(7) (1998) 993-1003.
34. E.G. Canty, K.E. Kadler, Procollagen trafficking, processing and fibrillogenesis, J Cell Sci 118(Pt 7) (2005) 1341-53.
35. Y. Ogawa, M.S. Razzaque, K. Kameyama, G. Hasegawa, S. Shimmura, M. Kawai, S. Okamoto, Y. Ikeda, K. Tsubota, Y. Kawakami, M. Kuwana, Role of heat shock protein 47, a collagen-binding chaperone, in lacrimal gland pathology in patients with cGVHD, Invest Ophthalmol Vis Sci 48(3) (2007) 1079-86.
36. K.E. Kadler, Y. Hojima, D.J. Prockop, Assembly of collagen fibrils de novo by cleavage of the type I pC-collagen with procollagen C-proteinase. Assay of critical concentration demonstrates that collagen

self-assembly is a classical example of an entropy-driven process, J Biol Chem 262(32) (1987) 15696-701.

37. L.I. Smith-Mungo, H.M. Kagan, Lysyl oxidase: properties, regulation and multiple functions in biology, Matrix Biol 16(7) (1998) 387-98.

38. I.S. Brown, Scanning electron microscopy of human dermal fibrous tissue, Journal of anatomy 113(2) (1972) 159-168.

39. T. Tateya, I. Tateya, D.M. Bless, Immuno-scanning electron microscopy of collagen types I and III in human vocal fold lamina propria, Ann Otol Rhinol Laryngol 116(2) (2007) 156-9.

40. K.E. Kadler, Fell Muir Lecture: Collagen fibril formation in vitro and in vivo, Int J Exp Pathol 98(1) (2017) 4-16.

41. W. Cheng, R. Yan-Hua, N. Fang-Gang, Z. Guo-an, The content and ratio of type I and III collagen in skin differ with age and injury, African journal of biotechnology 10(13) (2011) 2524-2529.

42. D.F. Holmes, K.E. Kadler, The 10+4 microfibril structure of thin cartilage fibrils, PNAS 103(46) (2003) 17249-17254.

43. J.P.R. Orgel, T.C. Irving, A. Miller, T.J. Wess, Microfibrillar structure of type I collagen *in situ*, PNAS 103(24) (2006) 9001-9005.

44. J.J. Wu, M.A. Weis, L.S. Kim, D.R. Eyre, Type III collagen, a fibril network modifier in articular cartilage, J Biol Chem 285(24) (2010) 18537-44.

45. A. Gautieri, S. Uzel, S. Vesentini, A. Redaelli, M.J. Buehler, Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils, Biophys J 97(3) (2009) 857-65.

46. H. Kuivaniemi, G. Tromp, D.J. Prockop, Mutations in collagen genes: causes of rare and some common diseases in humans, Faseb J 5(7) (1991) 2052-60.

47. D.J. Prockop, Mutations that alter the primary structure of type I collagen. The perils of a system for generating large structures by the principle of nucleated growth, J Biol Chem 265(26) (1990) 15349-52.

48. J.A. Ramshaw, N.K. Shah, B. Brodsky, Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides, J Struct Biol 122(1-2) (1998) 86-91.

49. A. Bhattacharjee, M. Bansal, Collagen structure: The madras triple helix and the current scenario, Life 57(3) (2005) 161-172.

50. M. Aikio, I. Alahuhta, S. Nurmenniemi, J. Suojanen, R. Palovuori, S. Teppo, T. Sorsa, C. López-Otín, T. Pihlajaniemi, T. Salo, R. Heljasvaara, P. Nyberg, Arresten, a collagen-derived angiogenesis inhibitor, suppresses invasion of squamous cell carcinoma, Plos One 7(12) (2012) e51044.

51. T.M. Mundel, R. Kalluri, Type IV collagen-derived angiogenesis inhibitors, Microvasc Res 74(2-3) (2007) 85-9.

52. G.D. Kamphaus, P.C. Colorado, D.J. Panka, H. Hopfer, R. Ramchandran, A. Torre, Y. Maeshima, J.W. Mier, V.P. Sukhatme, R. Kalluri, Canstatin, a novel matrix-derived inhibitor of angiogenesis and tumor growth, J Biol Chem 275(2) (2000) 1209-15.

53. C. Magnon, A. Galaup, B. Mullan, V. Rouffiac, C. Bouquet, J.M. Bidart, F. Griscelli, P. Opolon, M. Perricaudet, Canstatin acts on endothelial and tumor cells via mitochondrial damage initiated through interaction with alphavbeta3 and alphavbeta5 integrins, Cancer Res 65(10) (2005) 4353-61.

54. Y. Hamano, M. Zeisberg, H. Sugimoto, J.C. Lvely, Y. Maeshima, C. Yang, R.O. Hynes, Z. Werb, A. Sudhakar, R. Kalluri, Physiological levels of tumstatin, a fragment of colagen IV A3 chain, are generated by MMP-9 proteolysis and suppress angiogenesis via aVB3 integrin, Cancer cell 3(6) (2003) 589-601.

55. A.G. Marneros, B.R. Olsen, The role of collagen-derived proteolytic fragments in angiogenesis, Matrix Biol 20(5-6) (2001) 337-45.

56. U. Felbor, L. Dreier, R.A. Bryant, H.L. Ploegh, B.R.

Olsen, W. Mothes, Secreted cathepsin L generates endostatin from collagen XVIII, Embo J 19(6) (2000) 1187-94.

57. I.M. Olfert, O. Baum, Y. Hellsten, S. Egginton, Advances and challenges in skeletal muscle angiogenesis, Am J Physiol Heart Circ Physiol 310(3) (2016) H326-36.

58. L. Standker, M. Schrader, S.M. Kanse, M. Jurgens, W.G. Forssmann, K.T. Preissner, Isolation and characterization of the circulating form of human endostatin, FEBS Lett 420(2-3) (1997) 129-33.

59. H. John, W.G. Forssmann, Determination of the disulfide bond pattern of the endogenous and recombinant angiogenesis inhibitor endostatin by mass spectrometry, Rapid Commun Mass Spectrom 15(14) (2001) 1222-8.

60. H. John, K. Radtke, L. Standker, W.G. Forssmann, Identification and characterization of novel endogenous proteolytic forms of the human angiogenesis inhibitors restin and endostatin, Biochim Biophys Acta 1747(2) (2005) 161-70.

61. E. Leikina, M.V. Mertts, N. Kuznetsova, S. Leikin, Type I collagen is thermally unstable at body temperature, Proc Natl Acad Sci U S A 99(3) (2002) 1314-8.

62. R.G. Paul, A.J. Bailey, Chemical stabilisation of collagen as a biomimetic, ScientificWorldJournal 3 (2003) 138-55.

63. P.L. Privalov, Stability of proteins. Proteins which do not present a single cooperative system, Adv Protein Chem 35 (1982) 1-104.

64. A. Torre-Blanco, E. Adachi, Y. Hojima, J.A. Wootton, R.R. Minor, D.J. Prockop, Temperature-induced post-translational over-modification of type I procollagen. Effects of over-modification of the protein on the rate of cleavage by procollagen N-proteinase and on self-assembly of collagen into fibrils, J Biol Chem 267(4) (1992) 2650-5.

65. M. Yamauchi, M. Shiiba, Posttranslational modifications of proteins, Humana Press2002.

66. J. Uitto, D.J. Prockop, Synthesis and secretion of under-hydroxylated procollagen at various temperatures by cells subject to temporary anoxia, Biochem Biophys Res Commun 60(1) (1974) 414-23.

67. B. Depalle, Z. Qin, S.J. Shefelbine, M.J. Buehler, Influence of cross-link structure, density and mechanical properties in the mesoscale deformation mechanisms of collagen fibrils, J Mech Behav Biomed Mater 52 (2015) 1-13.

68. J.C. Utting, S.P. Robins, A. Brandao-Burch, I.R. Orriss, J. Behar, T.R. Arnett, Hypoxia inhibits the growth, differentiation and bone-forming capacity of rat osteoblasts, Exp Cell Res 312(10) (2006) 1693-702.

69. G. Zhang, T. Liu, Q. Wang, L. Chen, J. Lei, J. Luo, G. Ma, Z. Su, Mass spectrometric detection of marker peptides in tryptic digests of gelatin: A new method to differentiate between bovine and porcine gelatin, Food Hydrocolloids 23(7) (2009) 2001-2007.

70. E. Song, Y. Mechref, LC-MS/MS identification of the O-glycosylation and hydroxylation of amino acid residues of collagen alpha-1 (II) chain from bovine cartilage, J Proteome Res 12(8) (2013) 3599-609.

71. T. Basak, L. Vega-Montoto, L.J. Zimmerman, D.L. Tabb, B.G. Hudson, R.M. Vanacore, Comprehensive Characterization of Glycosylation and Hydroxylation of Basement Membrane Collagen IV by High-Resolution Mass Spectrometry, J Proteome Res 15(1) (2016) 245-58.

72. C. Yang, A.C. Park, N.A. Davis, J.D. Russell, B. Kim, D.D. Brand, M.J. Lawrence, Y. Ge, M.S. Westphall, J.J. Coon, D.S. Greenspan, Comprehensive mass spectrometric mapping of the hydroxylated amino acid residues of the alpha1(V) collagen chain, J Biol Chem 287(48) (2012) 40598-610.

73. H. Montgomery, N. Rustogi, A. Hadjisavvas, K. Tanaka, K. Kyriacou, C.W. Sutton, Proteomic Profiling of Breast Tissue Collagens and Site-specific Characterization of Hydroxyproline Residues of Collagen Alpha-1-(I),

J proteome research 11 (2012) 5890-5902.

74. D.B. Kassel, K. Biemann, Differentiation of hydroxyproline isomers and isobars in peptides by tandem mass spectrometry, Anal Chem 62(15) (1990) 1691-5.

75. N.A. van Huizen, P.C. Burgers, F. Saintmont, P. Brocorens, P. Gerbaux, C. Stingl, L.J.M. Dekker, I.J. JNM, T.M. Luider, Identification of 4-Hydroxyproline at the Xaa Position in Collagen by Mass Spectrometry, J Proteome Res 18(5) (2019) 2045-2051.

76. K.I. Kivirikko, J. Myllyharju, Prolyl 4-hydroxylases and their protein disulfide isomerase subunit, Matrix Biol 16(7) (1998) 357-68.

77. J. Myllyharju, Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis, Matrix Biol 22(1) (2003) 15-24.

78. K.L. Gorres, R.T. Raines, Prolyl 4-hydroxylase, Crit Rev Biochem Mol Biol 45(2) (2010) 106-24.

79. R.S. Rapaka, V. Renugopalakrishman, D.W. Urry, R.S. Bhatnagar, Hydroxylation of proline in polytripeptide models of collagen: stereochemistry of polytripeptide-prolyl hydroxylase interaction, Biochemistry 17(14) (1978) 2892-2898.

80. J.F. Lees, N.J. Bulleid, The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen, J Biol Chem 269(39) (1994) 24354-60.

81. B. Sacca, C. Renner, L. Moroder, The chain register in heterotrimeric collagen peptides affects triple helix stability and folding kinetics, J Mol Biol 324(2) (2002) 309-18.

82. L.E. Bretscher, C.L. Jenkins, K.M. Taylor, M.L. DeRider, R.T. Raines, Conformational stability of collagen relies on a stereoelectronic effect, J Am Chem Soc 123(4) (2001) 777-8.

83. C. Giunta, N.H. Elcioglu, B. Albrecht, G. Eich, C. Chambaz, A.R. Janecke, H. Yeowell, M. Weis, D.R. Eyre, M. Kraenzlin, B. Steinmann, Spondylocheiro dysplastic form of the Ehlers-Danlos syndrome–an autosomal-recessive entity caused by mutations in the zinc transporter gene SLC39A13, Am J Hum Genet 82(6) (2008) 1290-305.

84. P.D. Metcalfe, J. Wang, H. Jiao, Y. Huang, K. Hori, R.B. Moore, E.E. Tredget, Bladder outlet obstruction: progression from inflammation to fibrosis, BJU Int 106(11) (2010) 1686-94.

85. L. Yang, T. Chan, J. Demare, T. Iwashina, A. Ghahary, P.G. Scott, E.E. Tredget, Healing of burn wounds in transgenic mice overexpressing transforming growth factor-beta 1 in the epidermis, Am J Pathol 159(6) (2001) 2147-57.

86. S.W. Chan, J. Greaves, N.A. Da Silva, S.W. Wang, Assaying proline hydroxylation in recombinant collagen variants by liquid chromatography-mass spectrometry, BMC Biotechnol 12 (2012) 51.

87. J.F. Wang, H. Jiao, T.L. Stewart, H.A. Shankowsky, P.G. Scott, E.E. Tredget, Fibrocytes from burn patients regulate the activities of fibroblasts, Wound Repair Regen 15(1) (2007) 113-21.

88. E.E. Tredget, N. Falk, P.G. Scott, A.M. Hogg, J.F. Burke, Determination of 4-hydroxyproline in collagen by gas chromatography/mass spectrometry, Anal Biochem 190(2) (1990) 259-65.

89. K. Lanckmans, S. Sarre, I. Smolders, Y. Michotte, Use of a structural analogue versus a stable isotope labeled internal standard for the quantification of angiotensin IV in rat brain dialysates using nano-liquid chromatography/tandem mass spectrometry, Rapid Commun Mass Spectrom 21(7) (2007) 1187-95.

90. E. Stokvis, H. Rosing, J.H. Beijnen, Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not?, Rapid Commun Mass Spectrom 19(3) (2005) 401-7.

91. D.D. Dziewiatkowski, V.C. Hascall, R.L. Riolo, Epimerization of trans-4-hydroxy-L-proline to cis-4-hydroxy-D-proline during acid hydrolysis of

collagen, Anal Biochem 49(2) (1972) 550-8.

92. W.J. Lindblad, R.F. Diegelmann, Quantitation of hydroxyproline isomers in acid hydrolysates by high-performance liquid chromatography, Anal Biochem 138(2) (1984) 390-5.

93. P. Tiainen, A. Pasanen, R. Sormunen, J. Myllyharju, Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV, J Biol Chem 283(28) (2008) 19432-9.

94. W.A. Cabral, W. Chang, A.M. Barnes, M.A. Weis, M.A. Scott, S. Leikin, E. Makareeva, N. Kuznetsova, K.N. Rosenbaum, C. Tifft, D.I. Bulas, C. Kozma, P.A. Smith, D.R. Eyre, J.C. Marini, Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta, Nat Genet 39(3) (2007) 359-365.

95. J.A. Vranka, L.Y. Sakai, H.P. Bachinger, Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes, J Biol Chem 279(22) (2004) 23615-21.

96. D. Baldridge, U. Schwarze, R. Morello, J. Lennington, T.K. Bertin, J.M. Pace, M.G. Pepin, M.A. Weis, D.R. Eyre, J. Walsh, D. Lambert, A. Green, H. Robinson, M. Michelson, G. Houge, C. Lindman, J. Martin, J. Ward, E. Lemyre, J.J. Mitchell, D. Krakow, D.L. Rimoin, D.H. Cohn, P.H. Byers, B. Lee, CRTAP and LEPRE1 mutations in recessive osteogenesis imperfecta, Hum Mutat 29(12) (2008) 1435-1442.

97. J. Khoshnoodi, V. Pedchenko, B.G. Hudson, Mammalian collagen IV, Microsc Res Tech 71(5) (2008) 357-70.

98. A.V. Persikov, J.A. Ramshaw, B. Brodsky, Prediction of collagen stability from amino acid sequence, J Biol Chem 280(19) (2005) 19343-9.

99. D.R. Eyre, M. Weis, D.M. Hudson, J.J. Wu, L. Kim, A novel 3-hydroxyproline (3Hyp)-rich motif marks the triple-helical C terminus of tendon type I collagen, J Biol Chem 286(10) (2011) 7732-6.

100. D.M. Hudson, M. Garibov, D.R. Dixon, T. Popowics, D.R. Eyre, Distinct post-translational features of type I collagen are conserved in mouse and human periodontal ligament, J Periodontal Res 52(6) (2017) 1042-1049.

101. M.A. Weis, D.M. Hudson, L. Kim, M. Scott, J.J. Wu, D.R. Eyre, Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly, J Biol Chem 285(4) (2010) 2580-90.

102. M. Fang, M.M. Banaszak Holl, Variation in type I collagen fibril nanomorphology: the significance and origin, BoneKEy 394 (2013).

103. T.J. Wess, Collagen fibril form and function, Adv Protein Chem 70 (2005) 341-74.

104. L. Bozec, G. van der Heijden, M. Horton, Collagen fibrils: nanoscale ropes, Biophys J 92(1) (2007) 70-5.

105. C.L. Jenkins, L.E. Bretscher, I.A. Guzei, R.T. Raines, Effect of 3-hydroxyproline residues on collagen stability, J Am Chem Soc 125(21) (2003) 6422-7.

106. K. Mizuno, D.H. Peyton, T. Hayashi, J. Engel, H.P. Bachinger, Effect of the -Gly-3(S)-hydroxyprolyl-4(R)-hydroxyprolyl- tripeptide unit on the stability of collagen model peptides, Febs J 275(23) (2008) 5830-40.

107. K. Uzawa, W.J. Grzesik, T. Nishiura, S.A. Kuznetsov, P.G. Robey, D.A. Brenner, M. Yamauchi, Differential expression of human lysyl hydroxylase genes, lysine hydroxylation, and cross-linking of type I collagen during osteoblastic differentiation in vitro, J Bone Miner Res 14(8) (1999) 1272-80.

108. D.R. Eyre, M.A. Paz, P.M. Gallop, Cross-linking in collagen and elastin, Annu Rev Biochem 53 (1984) 717-48.

109. M. Yamauchi, M. Sricholpech, Lysine post-translational modifications of collagen, Essays Biochem 52 (2012) 113-33.

110. C. Wang, M. Valtavaara, R. Myllyla, Lack of collagen type specificity for lysyl hydroxylase isoforms, DNA Cell Biol 19(2) (2000) 71-7.

111. D.K. Mercer, P.F. Nicol, C. Kimbembe, S.P. Robins, Identification, expression, and tissue distribution of the three rat lysyl hydroxylase isoforms, Biochem Biophys Res Commun 307(4) (2003) 803-9.

112. J.E. Lee, Y. Kim, A tissue-specific variant of the human lysyl oxidase-like protein 3 (LOXL3) functions as an amine oxidase with substrate specificity, J Biol Chem 281(49) (2006) 37282-90.

113. Q. Xiao, G. Ge, Lysyl oxidase, extracellular matrix remodeling and cancer metastasis, Cancer Microenviron 5(3) (2012) 261-73.

114. J.E. Wagenseil, R.P. Mecham, Vascular extracellular matrix and arterial mechanics, Physiol Rev 89(3) (2009) 957-89.

115. H.M. Kagan, W. Li, Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell, J Cell Biochem 88(4) (2003) 660-72.

116. H.M. Kagan, M.A. Williams, P.R. Williamson, J.M. Anderson, Influence of sequence and charge on the specificity of lysyl oxidase toward protein and synthetic peptide substrates, J Biol Chem 259(18) (1984) 11203-7.

117. R. Chen, X. Jiang, D. Sun, G. Han, F. Wang, M. Ye, L. Wang, H. Zou, Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry, J Proteome Res 8(2) (2009) 651-61.

118. I. Perdivara, M. Yamauchi, K.B. Tomer, Molecular Characterization of Collagen Hydroxylysine O-Glycosylation by Mass Spectrometry: Current Status, Aust J Chem 66(7) (2013) 760-769.

119. P. Chen, M. Cescon, P. Bonaldo, Collagen VI in cancer and its biological mechanisms, Cell 19(7) (2013) 410-417.

120. D.F. Zielinska, F. Gnad, K. Schropp, J.R. Wisniewski, M. Mann, Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery, Mol Cell 46(4) (2012) 542-8.

121. W. Morelle, J.C. Michalski, Analysis of protein glycosylation by mass spectrometry, Nat Protoc 2(7) (2007) 1585-602.

122. K.L. Ford, W. Zeng, J.L. Heazlewood, A. Bacic, Characterization of protein N-glycosylation by tandem mass spectrometry using complementary fragmentation techniques, Front Plant Sci 6 (2015) 674.

123. B. Schegg, A.J. Hulsmeier, C. Rutschmann, C. Maag, T. Hennet, Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases, Mol Cell Biol 29(4) (2009) 943-52.

124. C.L. Yang, H. Rui, S. Mosler, H. Notbohm, A. Sawaryn, P.K. Muller, Collagen II from articular cartilage and annulus fibrosus. Structural and functional implication of tissue specific posttranslational modifications of collagen molecules, Eur J Biochem 213(3) (1993) 1297-302.

125. [125] E. Gineyts, p. Garnero, P.D. Delmas, Urinary excretion of glucosyl-galactosyl pyridinoline: a specific biochemical marker of synovium degradation, Rheumatology (Oxford) 40 (2001) 315-323.

126. M. Terajima, I. Perdivara, M. Sricholpech, Y. Deguchi, N. Pleshko, K.B. Tomer, M. Yamauchi, Glycosylation and cross-linking in bone type I collagen, J Biol Chem 289(33) (2014) 22636-47.

127. J. Uribarri, S. Woodruff, S. Goodman, W. Cai, X. Chen, R. Pyzik, A. Yong, G.E. Striker, H. Vlassara, Advanced glycation end products in foods and a practical guide to their reduction in the diet, J Am Diet Assoc 110(6) (2010) 911-16 e12.

128. [128] Y. Takagi, A. Kashiwagi, Y. Tanaka, T. Asahina, R. Kikkawa, Y. Shigeta, Significance of fructose-induced

protein oxidation and formation of advanced glycation end product, J Diabetes Complications 9(2) (1995) 87-91.

129. G. Fessel, Y. Li, V. Diederich, M. Guizar-Sicairos, P. Schneider, D.R. Sell, V.M. Monnier, J.G. Snedeker, Advanced glycation end-products reduce collagen molecular sliding to affect collagen fibril damage mechanisms but not stiffness, PLoS One 9(11) (2014) e110948.

130. J.M. Bohlender, S. Franke, G. Stein, G. Wolf, Advanced glycation end products and the kidney, Am J Physiol Renal Physiol 289 (2005) 645-659.

131. T.T. Andreassen, K. Seyer-Hansen, A.J. Bailey, Thermal stability, mechanical properties and reducible cross-links of rat tail tendon in experimental diabetes, Biochim Biophys Acta 677(2) (1981) 313-7.

132. C.C. Danielsen, T.T. Andreassen, Mechanical properties of rat tail tendon in relation to proximal-distal sampling position and age, J Biomech 21(3) (1988) 207-12.

133. Y. Li, G. Fessel, M. Georgiadis, J.G. Snedeker, Advanced glycation end-products diminish tendon collagen fiber sliding, Matrix Biol 32(3-4) (2013) 169-77.

134. S.Y. Tang, D. Vashishth, The relative contributions of non-enzymatic glycation and cortical porosity on the fracture toughness of aging bone, J Biomech 44(2) (2011) 330-6.

135. E.A. Zimmermann, E. Schaible, H. Bale, H.D. Barth, S.Y. Tang, P. Reichert, B. Busse, T. Alliston, J.W. Ager, 3rd, R.O. Ritchie, Age-related changes in the plasticity and toughness of human cortical bone at multiple length scales, Proc Natl Acad Sci U S A 108(35) (2011) 14416-21.

136. N. Verzijl, J. DeGroot, Z.C. Ben, O. Brau-Benjamin, A. Maroudas, R.A. Bank, J. Mizrahi, C.G. Schalkwijk, S.R. Thorpe, J.W. Baynes, J.W. Bijlsma, F.P. Lafeber, J.M. TeKoppele, Crosslinking by advanced glycation end products increases the stiffness of the collagen network in human articular cartilage: a possible mechanism through which age is a risk factor for osteoarthritis, Arthritis Rheum 46(1) (2002) 114-23.

137. M. Saito, K. Marumo, K. Fujii, N. Ishioka, Single-column high-performance liquid chromatographic-fluorescence detection of immature, mature, and senescent cross-links of collagen, Anal Biochem 253(1) (1997) 26-32.

138. D.R. Sell, V.M. Monnier, Structure elucidation of a senescence cross-link from human extracellular matrix. Implication of pentoses in the aging process, J Biol Chem 264(36) (1989) 21597-602.

139. K.B. Holte, N.G. Juel, J.I. Brox, K.F. Hanssen, D.S. Fosmark, D.R. Sell, V.M. Monnier, T.J. Berg, Hand, shoulder and back stiffness in long-term type 1 diabetes; cross-sectional association with skin collagen advanced glycation end-products. The Dialong study, J Diabetes Complications 31(9) (2017) 1408-1414.

140. K. Mikulikova, A. Eckhardt, S. Pataridis, I. Miksik, Study of posttranslational non-enzymatic modifications of collagen using capillary electrophoresis/mass spectrometry and high performance liquid chromatography/mass spectrometry, J Chromatogr A 1155(2) (2007) 125-33.

141. D.R. Eyre, M.A. Weis, J.J. Wu, Advances in collagen cross-link analysis, Methods 45(1) (2008) 65-74.

142. R. Vanacore, A.J. Ham, M. Voehler, C.R. Sanders, T.P. Conrads, T.D. Veenstra, K.B. Sharpless, P.E. Dawson, B.G. Hudson, A sulfilimine bond identified in collagen IV, Science 325(5945) (2009) 1230-4.

143. A.M. McLaren, L.D. Hordon, H.A. Bird, S.P. Robins, Urinary excretion of pyridinium crosslinks of collagen in patients with osteoporosis and the effects of bone fracture, Ann Rheum Dis 51(5) (1992) 648-51.

144. D.R. Eyre, T.J. Koob, K.P. Van Ness, Quantitation of hydroxypyridinium crosslinks in collagen by high-performance liquid chromatography, Anal Biochem 137(2) (1984) 380-8.

145. C. Couppe, P. Hansen, M. Kongsgaard, V. Kovanen, C. Suetta, P. Aagaard, M. Kjaer, S.P. Magnusson, Mechanical properties and collagen cross-linking of the patellar tendon in old and young men, J Appl Physiol (1985) 107(3) (2009) 880-6.

146. S. Ricard-Blum, M. Van der Rest, B. Dublet, Unconventional Collagens Types VI, VII, VIII, IX, X, XII, XIV, XVI, and XIX, Oxford university press2000.

147. S. Ricard-Blum, The collagen family, Cold spring harb perspectives in biology 3(1) (2011).

148. K. Gelse, E. Poschl, T. Aigner, Collagens--structure, function, and biosynthesis, Adv Drug Deliv Rev 55(12) (2003) 1531-46.

149. M.D. Shoulders, R.T. Raines, Collagen structure and stability, Annu Rev Biochem 78 (2009) 929-58.

150. M.J. Bissell, W.C. Hines, Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression, Nat Med 17(3) (2011) 320-9.

151. X.J. Ma, S. Dahiya, E. Richardson, M. Erlander, D.C. Sgroi, Gene expression profiling of the tumor microenvironment during breast cancer progression, Breast Cancer Res 11(1) (2009) R7.

152. A. Naba, K.R. Clauser, C.A. Whittaker, S.A. Carr, K.K. Tanabe, R.O. Hynes, Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver, BMC Cancer 14 (2014) 518.

153. J. Myllyharju, K.I. Kivirikko, Collagens and collagen-related diseases, Ann Med 33(1) (2001) 7-21.

154. L.M. Coussens, B. Fingleton, L.M. Matrisian, Matrix metalloproteinase inhibitors and cancer: trials and tribulations, Science 295(5564) (2002) 2387-92.

155. M. Kjaer, H. Langsberg, B.F. Miller, R. Boushel, R. Crameri, S. Koskinen, K. Heinemeier, J.L. Olesen, S. Dossing, M. Hansen, S.G. Pedersen, M.J. Rennie, P. Magnusson, Metabolic activity and collagen turnover in human tendin in response to physical activity, J musculoskelet neuronal interact 5(1) (2005) 41-52.

156. M. Kjaer, P. Magnusson, M. Krogsgaard, J. Boysen Moller, J. Olesen, K. Heinemeier, M. Hansen, B. Haraldsson, S. Koskinen, B. Esmarck, H. Langberg, Extracellular matrix adaptation of tendon and skeletal muscle to exercise, J Anat 208(4) (2006) 445-50.

157. R.T. Aimes, J.P. Quigley, Matrix metalloproteinase-2 is an interstitial collagenase. Inhibitor-free enzyme catalyzes the cleavage of collagen fibrils and soluble native type I collagen generating the specific 3/4- and 1/4-length fragments, J Biol Chem 270(11) (1995) 5872-6.

158. E.Y. Zhen, I.J. Brittain, D.A. Laska, P.G. Mitchell, E.U. Sumer, M.A. Karsdal, K.L. Duffin, Characterization of metalloprotease cleavage products of human articular cartilage, Arthritis Rheum 58(8) (2008) 2420-31.

159. J.P.G. Sluijter, D.P.V. Kleijn de, G. Pasterkamp, Vascular remodeling and protease inhibition–bench to bedside, Cardiovasc Res 69(3) (2006) 595-603.

160. P.A.M. Snoek van Beurden, J.W. Hoff von den, Zymographic techniques for the analysis of matrix metalloproteinases and their inhibitors, BioTechniques 38(1) (2005) 73-83.

161. M.D. Sternlicht, Z. Werb, How matrix metalloproteinases regulate cell behaviour., Annu Rev Cell Dev Biol 17 (2001) 463-516.

162. R. Ala-aho, V.M. Kähäri, Collagenases in cancer, Biochimie 87 (2005) 273-286.

163. S.W. Manka, F. Carafoli, R. Visse, D. Bihan, N. Raynal, R.W. Farndale, G. Murphy, J.J. Enghild, E. Hohenester, H. Nagase, Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1, Proc Natl Acad Sci USA 109(31) (2012) 12461-6.

164. J.L. Lauer-Fields, M.J. Chalmers, S.A. Busby, D. Minond, P.R. Griffin, G.B. Fields, Identification of specific hemopexin-like domain residues that facilitate matrix metalloproteinase collagenolytic activity, J Biol Chem 284(36) (2009) 24017-24.

R

165. N. Verzijl, J. DeGroot, S.R. Thorpe, R.A. Bank, J.N. Shaw, T.J. Lyons, J.W. Bijlsma, F.P. Lafeber, J.W. Baynes, J.M. TeKoppele, Effect of collagen turnover on the accumulation of advanced glycation end products, J Biol Chem 275(50) (2000) 39027-31.

166. S.S. Sivan, E. Wachtel, E. Tsitron, N. Sakkee, F. van der Ham, J. Degroot, S. Roberts, A. Maroudas, Collagen turnover in normal and degenerate human intervertebral discs as determined by the racemization of aspartic acid, J Biol Chem 283(14) (2008) 8796-801.

167. G.J. Rucklidge, G. Milne, B.A. McGaw, E. Milne, S.P. Robins, Turnover rates of different collagen types measured by isotope ratio mass spectrometry, Biochim Biophys Acta 1156(1) (1992) 57-61.

168. K.M. Heinemeier, P. Schjerling, J. Heinemeier, S.P. Magnusson, M. Kjaer, Lack of tissue renewal in human adult Achilles tendon is revealed by nuclear bomb (14)C, FASEB J 27(5) (2013) 2074-9.

169. K.M. Heinemeier, P. Schjerling, J. Heinemeier, M.B. Moller, M.R. Krogsgaard, T. Grum-Schwensen, M.M. Petersen, M. Kjaer, Radiocarbon dating reveals minimal collagen turnover in both healthy and osteoarthritic human cartilage, Sci Transl Med 8(346) (2016) 346ra90.

170. D.J. Wilkinson, M.V. Franchi, M.S. Brook, M.V. Narici, J.P. Williams, W.K. Mitchell, N.J. Szewczyk, P.L. Greenhaff, P.J. Atherton, K. Smith, A validation of the application of D(2)O stable isotope tracer techniques for monitoring day-to-day changes in muscle protein subfraction synthesis in humans, Am J Physiol Endocrinol Metab 306(5) (2014) E571-9.

171. M.A. Karsdal, F. Genovese, E.A. Madsen, T. Manon-Jensen, D. Schuppan, Collagen and tissue turnover as a function of age: Implications for fibrosis, J Hepatol 64(1) (2016) 103-9.

172. B.A. Wright, Low-angle X-ray diffraction pattern of collagen, Nature 162(4105) (1948) 23.

173. G.N. Ramachandran, G. Kartha, Structure of collagen, Nature 174(4423) (1954) 269-70.

174. J. Gross, F.O. Schmitt, The structure of human skin collagen as studied with the electron microscope, J Exp Med 88(5) (1948) 555-68.

175. T. Starborg, N.S. Kalson, Y. Lu, A. Mironov, T.F. Cootes, D.F. Holmes, K.E. Kadler, Using transmission electron microscopy and 3View to determine collagen fibril size and three-dimensional organization, Nat Protoc 8(7) (2013) 1433-48.

176. R. Cicchi, N. Vogler, D. Kapsokalyvas, B. Dietzek, J. Popp, F.S. Pavone, From molecular structure to tissue architecture: collagen organization probed by SHG microscopy, J Biophotonics 6(2) (2013) 129-42.

177. G. Cox, E. Kable, A. Jones, I. Fraser, F. Manconi, M.D. Gorrell, 3-dimensional imaging of collagen using second harmonic generation, J Struct Biol 141(1) (2003) 53-62.

178. H. Nystrom, B. Tavelin, M. Bjorklund, P. Naredi, M. Sund, Improved tumour marker sensitivity in detecting colorectal liver metastases by combined type IV collagen and CEA measurement, Tumour Biol 36(12) (2015) 9839-47.

179. A.J. Bailey, C.M. Peach, L.J. Fowler, Chemistry of the collagen cross-links. Isolation and characterization of two intermediate intermolecular cross-links in collagen, Biochem J 117(5) (1970) 819-31.

180. J.B. Fenn, Nobel Lecture. NobelPrize.org. https://www.nobelprize.org/prizes/chemistry/2002/fenn/lecture/, (2002).

181. K. Tanaka, Nobel Lecture. NobelPrize.org. https://www.nobelprize.org/prizes/chemistry/2002/tanaka/lecture/, (2002).

182. X. Han, A. Aslanian, J.R. Yates, 3rd, Mass spectrometry for proteomics, Curr Opin Chem Biol 12(5) (2008) 483-90.

183. B. Domon, R. Aebersold, Mass spectrometry and protein analysis, Science 312(5771) (2006) 212-7.

184. S. Mehmood, T.M. Allison, C.V. Robinson, Mass spectrometry of protein complexes: from origins to applications, Annu Rev Phys Chem 66 (2015) 453-74.

185. M.A. Baldwin, Protein identification by mass spectrometry: issues to be considered, Mol Cell Proteomics 3(1) (2004) 1-9.

186. Z. Szabo, T. Janaky, Challenges and developments in protein identification using mass spectrometry, Trac-Trend Anal Chem 69 (2015) 76-87.

187. F. Meier, P.E. Geyer, S. Virreira Winter, J. Cox, M. Mann, BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes, Nat Methods 15(6) (2018) 440-448.

188. S.Y. Ritter, J. Collins, B. Krastins, D. Sarracino, M. Lopez, E. Losina, A.O. Aliprantis, Mass spectrometry assays of plasma biomarkers to predict radiographic progression of knee osteoarthritis, Arthritis Res Ther 16(5) (2014) 456.

189. M. Holtta, H. Zetterberg, E. Mirgorodskaya, N. Mattsson, K. Blennow, J. Gobom, Peptidome analysis of cerebrospinal fluid by LC-MALDI MS, PLoS One 7(8) (2012) e42555.

190. F. Welker, M.J. Collins, J.A. Thomas, M. Wadsley, S. Brace, E. Cappellini, S.T. Turvey, M. Reguero, J.N. Gelfo, A. Kramarz, J. Burger, J. Thomas-Oates, D.A. Ashford, P.D. Ashton, K. Rowsell, D.M. Porter, B. Kessler, R. Fischer, C. Baessmann, S. Kaspar, J.V. Olsen, P. Kiley, J.A. Elliott, C.D. Kelstrup, V. Mullin, M. Hofreiter, E. Willerslev, J.J. Hublin, L. Orlando, I. Barnes, R.D. MacPhee, Ancient proteins resolve the evolutionary history of Darwin's South American ungulates, Nature 522(7554) (2015) 81-4.

191. K. Dreisewerd, A. Rohlfing, B. Spottke, C. Urbanke, W. Henkel, Characterization of whole fibril-forming collagen proteins of types I, III, and V from fetal calf skin by infrared matrix-assisted laser desorption ionization mass spectrometry, Anal Chem 76(13) (2004) 3482-91.

192. M. Gessel, J.M. Spraggins, P. Voziyan, B.G. Hudson, R.M. Caprioli, Decellularization of intact tissue enables MALDI imaging mass spectrometry analysis of the extracellular matrix, J Mass Spectrom 50(11) (2015) 1288-93.

193. S. Pataridis, A. Eckhardt, K. Mikulikova, P. Sedlakova, I. Miksik, Identification of collagen types in tissues using HPLC-MS/MS, J Sep Sci 31(20) (2008) 3483-8.

194. B. Zhang, J. Wang, X. Wang, J. Zhu, Q. Liu, Z. Shi, M.C. Chambers, L.J. Zimmerman, K.F. Shaddox, S. Kim, S.R. Davies, S. Wang, P. Wang, C.R. Kinsinger, R.C. Rivers, H. Rodriguez, R.R. Townsend, M.J. Ellis, S.A. Carr, D.L. Tabb, R.J. Coffey, R.J. Slebos, D.C. Liebler, C. Nci, Proteogenomic characterization of human colon and rectal cancer, Nature 513(7518) (2014) 382-7.

195. D. Pflieger, S. Chabane, O. Gaillard, B.A. Bernard, P. Ducoroy, J. Rossier, J. Vinh, Comparative proteomic analysis of extracellular matrix proteins secreted by two types of skin fibroblasts, Proteomics 6(21) (2006) 5868-79.

196. M. Buckley, M. Collins, J. Thomas-Oates, J.C. Wilson, Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, Rapid Commun Mass Spectrom 23(23) (2009) 3843-54.

197. Y. Taga, M. Kusubata, K. Ogawa-Goto, S. Hattori, Stable isotope-labeled collagen: a novel and versatile tool for quantitative collagen analyses using mass spectrometry, J Proteome Res 13(8) (2014) 3671-8.

198. L.V. Turoverova, M.G. Khotin, N.M. Iudintseva, K.E. Magnusson, M.I. Blinova, G.P. Pinaev, D.G. Tentler, Analysis of extracellular matrix proteins produced by cultured cells, Tsitologiia 51(8) (2009) 691-6.

199. A.J. Taylor, B.D. Ratner, L.D. Buttery, M.R. Alexander, Revealing cytokine-induced changes in the extracellular matrix with secondary ion mass spectrometry, Acta Biomater 14 (2015) 70-83.

200. S.E. Gilpin, J.P. Guyette, G. Gonzalez, X. Ren, J.M.

Asara, D.J. Mathisen, J.P. Vacanti, H.C. Ott, Perfusion decellularization of human and porcine lungs: bringing the matrix to clinical scale, J Heart Lung Transplant 33(3) (2014) 298-308.

201. M.M.A. Verstegen, J. Willemse, S. van den Hoek, G.J. Kremers, T.M. Luider, N.A. van Huizen, F. Willemssen, H.J. Metselaar, I.J. JNM, L.J.W. van der Laan, J. de Jonge, Decellularization of Whole Human Liver Grafts Using Controlled Perfusion for Transplantable Organ Bioscaffolds, Stem Cells Dev 26(18) (2017) 1304-1315.

202. A. Naba, K.R. Clauser, S. Hoersch, H. Liu, S.A. Carr, R.O. Hynes, The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices, Mol Cell Proteomics 11(4) (2012) M111 014647.

203. P.M. Vaysse, R.M.A. Heeren, T. Porta, B. Balluff, Mass spectrometry imaging for clinical research - latest developments, applications, and current limitations, Analyst 142(15) (2017) 2690-2712.

204. I. Kaya, D. Brinet, W. Michno, M. Baskurt, H. Zetterberg, K. Blenow, J. Hanrieder, Novel Trimodal MALDI Imaging Mass Spectrometry (IMS3) at 10 mum Reveals Spatial Lipid and Peptide Correlates Implicated in Abeta Plaque Pathology in Alzheimer's Disease, ACS Chem Neurosci 8(12) (2017) 2778-2790.

205. A. Takaoka, N. Babar, J. Hogan, M. Kim, M.O. Price, F.W. Price, Jr., S.L. Trokel, D.C. Paik, An Evaluation of Lysyl Oxidase-Derived Cross-Linking in Keratoconus by Liquid Chromatography/Mass Spectrometry, Invest Ophthalmol Vis Sci 57(1) (2016) 126-36.

206. N.M. Zork, K.M. Myers, K. Yoshida, S. Cremers, H. Jiang, C.V. Ananth, R.J. Wapner, J. Kitajewski, J. Vink, A systematic evaluation of collagen cross-links in the human cervix, Am J Obstet Gynecol 212(3) (2015) 321 e1-8.

207. D.M. Maahs, J. Siwy, A. Argiles, M. Cerna, C. Delles, A.F. Dominiczak, N. Gayrard, A. Iphofer, L. Jansch, G. Jerums, K. Medek, H. Mischak, G.J. Navis, J.M. Roob, K. Rossing, P. Rossing, I. Rychlik, E. Schiffer, R.E. Schmieder, T.C. Wascher, B.M. Winklhofer-Roob, L.U. Zimmerli, P. Zurbig, J.K. Snell-Bergeon, Urinary collagen fragments are significantly altered in diabetes: a link to pathophysiology, PLoS One 5(9) (2010) e13051.

208. M.E.E. Bröker, Z.S. Lalmahomed, H.P. Roest, N.A. van Huizen, L.J.M. Dekker, W. Calame, C. Verhoef, J.N.M. IJzermans, T.M. Luider, Collagen Peptides in Urine: A New Promising Biomarker for the Detection of Colorectal Liver Metastases, PlosOne 8(8) (2013).

209. Z.S. Lalmahomed, M.E. Broker, N.A. van Huizen, R.R. Coebergh van den Braak, L.J. Dekker, D. Rizopoulos, C. Verhoef, E.W. Steyerberg, T.M. Luider, I.J. JN, Hydroxylated collagen peptide in urine as biomarker for detecting colorectal liver metastases, Am J Cancer Res 6(2) (2016) 321-30.

210. M. Chalmers, J. Huckle, N.J. Cotton, Development of a capillary electrophoresis method for the characterization of collagens in cartilage tissue, J Chromatogr Sci 37(11) (1999) 443-7.

211. P. Chang, S. Kuan, G. Eberlein, D. Burke, R. Jones, Characterization of bovine collagens using capillary electrophoresis, J Pharm Biomed Anal 22(6) (2000) 957-66.

212. J. Novotna, Z. Deyl, I. Miksik, Capillary zone electrophoresis of collagen type I CNBr peptides in acid buffers, J Chromatogr B Biomed Appl 681(1) (1996) 77-82.

213. J. Metzger, C. Chatzikyrkou, V. Broecker, E. Schiffer, L. Jaensch, A. Iphoefer, M. Mengel, W. Mullen, H. Mischak, H. Haller, W. Gwinner, Diagnosis of subclinical and clinical acute T-cell-mediated rejection in renal transplant patients by urinary proteome analysis, Proteomics Clin Appl 5(5-6) (2011) 322-33.

214. A.D. Kistler, H. Mischak, D. Poster, M. Dakna, R.P. Wuthrich, A.L. Serra, Identification of a unique

urinary biomarker profile in patients with autosomal dominant polycystic kidney disease, Kidney Int 76(1) (2009) 89-96.

215. J. Drube, E. Schiffer, H. Mischak, M.J. Kemper, T. Neuhaus, L. Pape, R. Lichtinghagen, J.H. Ehrich, Urinary proteome pattern in children with renal Fanconi syndrome, Nephrol Dial Transplant 24(7) (2009) 2161-9.

216. L. Molin, R. Seraglia, A. Lapolla, E. Ragazzi, J. Gonzalez, A. Vlahou, J.P. Schanstra, A. Albalat, M. Dakna, J. Siwy, J. Jankowski, V. Bitsika, H. Mischak, P. Zurbig, P. Traldi, A comparison between MALDI-MS and CE-MS data for biomarker assessment in chronic kidney diseases, J Proteomics 75(18) (2012) 5888-97.

217. P. Onnerfjord, A. Khabut, F.P. Reinholt, O. Svensson, D. Heinegard, Quantitative proteomic analysis of eight cartilaginous tissues reveals characteristic differences as well as similarities between subgroups, J Biol Chem 287(23) (2012) 18913-24.

218. G. Uechi, Z. Sun, E.M. Schreiber, W. Halfter, M. Balasubramani, Proteomic View of Basement Membranes from Human Retinal Blood Vessels, Inner Limiting Membranes, and Lens Capsules, J Proteome Res (2014).

219. M.T. Clementz, New insight from old bones: stable isotope analysis of fossil mammals, J Mammal 93(2) (2012) 368-380.

220. M. Moini, C.M. Rollman, C.A. France, Dating human bone: is racemization dating species-specific?, Anal Chem 85(23) (2013) 11211-5.

221. M.L.S. Jorkov, J. Heinerneier, N. Lynnerup, Evaluating bone collagen extraction methods for stable isotope analysis in dietary studies, J Archaeol Sci 34(11) (2007) 1824-1829.

222. N. Piotrowska, T. Goslar, Preparation of bone samples in the Gliwice Radiocarbon Laboratory for AMS radiocarbon dating, Isotopes Environ Health Stud 38(4) (2002) 267-75.

223. S. Cersoy, A. Zazzo, M. Lebon, J. Rofes, S. Zirah, Collagen Extraction and Stable Isotope Analysis of Small Vertebrate Bones: A Comparative Approach, Radiocarbon 59(3) (2017) 679-694.

224. W. Kutschera, Applications of accelerator mass spectrometry, Int J Mass Spectrom 349 (2013) 203-218.

225. J.J. Hublin, S. Talamo, M. Julien, F. David, N. Connet, P. Bodu, B. Vandermeersch, M.P. Richards, Radiocarbon dates from the Grotte du Renne and Saint-Cesaire support a Neandertal origin for the Chatelperronian, Proc Natl Acad Sci U S A 109(46) (2012) 18743-8.

226. R.E. Wood, C.B. Ramsey, T.F.G. Higham, Refining Background Corrections for Radiocarbon Dating of Bone Collagen at Orau, Radiocarbon 52(2) (2010) 600-611.

227. M. Buckley, S.W. Kansa, S. Howard, S. Campbell, J. Thomas-Oates, M. Collins, Distinguishing between archaeological sheep and goat bones using a single collagen peptide, J Archaeol Sci 37(1) (2010) 13-20.

228. G.F. Zhang, T. Liu, Q. Wang, L. Chen, J.D. Lei, J. Luo, G.H. Ma, Z.G. Su, Mass spectrometric detection of marker peptides in tryptic digests of gelatin: A new method to differentiate between bovine and porcine gelatin, Food Hydrocolloids 23(7) (2009) 2001-2007.

229. H.H. Grundy, P. Reece, M. Buckley, C.M. Solazzo, A.A. Dowle, D. Ashford, A.J. Charlton, M.K. Wadsley, M.J. Collins, A mass spectrometry method for the determination of the species of origin of gelatine in foods and pharmaceutical products, Food Chem 190 (2016) 276-284.

230. Y. Kumazawa, Y. Taga, K. Iwai, Y. Koyama, A Rapid and Simple LC-MS Method Using Collagen Marker Peptides for Identification of the Animal Source of Leather, J Agric Food Chem 64(30) (2016) 6051-7.

231. K. Mizuno, S.P. Boudko, J. Engel, H.P. Bachinger, Kinetic hysteresis in collagen folding, Biophys J 98(12) (2010) 3004-14.

232. A.F. Woodlock, B.S. Harrap, The effects of salts on

the stability of the collagen helix under acidic conditions, Aust J Biol Sci 21(4) (1968) 821-6.

233. K. Inouye, S. Sakakibara, D.J. Prockop, Effects of the stereo-configuration of the hydroxyl group in 4-hydroxyproline on the triple-helical structures formed by homogenous peptides resembling collagen., Biochim. Biophys. Acta 420(1) (1976) 133-141.

234. V.C. Chan, J.A. Ramshaw, A. Kirkpatrick, K. Beck, B. Brodsky, Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix, J Biol Chem 272(50) (1997) 31441-6.

235. W. Yang, V.C. Chan, A. Kirkpatrick, J.A. Ramshaw, B. Brodsky, Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides, J Biol Chem 272(46) (1997) 28837-40.

236. C.M. Stultz, Localized unfolding of collagen explains collagenase cleavage near imino-poor sites, J Mol Biol 319(5) (2002) 997-1003.

237. E.Y. Jones, A. Miller, Analysis of structural design features in collagen, J Mol Biol 218(1) (1991) 209-19.

238. H.P. Bächinger, J.M. Davis, Sequence specific thermal stability of the collagen triple helix, Int J Biol Macromol 13(3) (1991) 152-156.

239. R. Dölz, E. Heidemann, Influence of different tripeptides on the stability of the collagen triple-helix. I. Analysis of the collagen sequence and identification of typical tripeptides, Biopolymers 25(6) (1986) 1069-1080.

240. G.R. Masson, M.L. Jenkins, J.E. Burke, An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery, Expert Opin Drug Discov 12(10) (2017) 981-994.

241. S.W. Chang, B.P. Flynn, J.W. Ruberti, M.J. Buehler, Molecular mechanism of force induced stabilization of collagen against enzymatic breakdown, Biomaterials 33(15) (2012) 3852-9.

242. S.W. Chang, S.J. Shefelbine, M.J. Buehler, Structural and mechanical differences between collagen homo- and heterotrimers: relevance for the molecular origin of brittle bone disease, Biophys J 102(3) (2012) 640-8.

243. N.V. Kuznetsova, D.J. McBride, S. Leikin, Changes in thermal stability and microunfolding pattern of collagen helix resulting from the loss of alpha2(I) chain in osteogenesis imperfecta murine, J Mol Biol 331(1) (2003) 191-200.

244. L.C. Junqueira, G. Bignolas, R.R. Brentani, Picrosirius staining plus polarization microscopy, a specific method for collagen detection in tissue sections, Histochem J 11(4) (1979) 447-55.

245. A. Lopez-De Leon, M. Rojkind, A simple micromethod for collagen and total protein determination in formalin-fixed paraffin-embedded sections, J Histochem Cytochem 33(8) (1985) 737-43.

246. D. Dayan, Y. Hiss, A. Hirshberg, J.J. Bubis, M. Wolman, Are the polarization colors of picrosirius red-stained collagen determined only by the diameter of the fibers?, Histochemistry 93(1) (1989) 27-29.

247. K.A. Wegner, A. Keikhosravi, K.W. Eliceiri, C.M. Vezina, Fluorescence of Picrosirius Red Multiplexed With Immunohistochemistry for the Quantitative Assessment of Collagen in Tissue Sections, J Histochem Cytochem 65(8) (2017) 479-490.

248. G.S. Karagiannis, C. Petraki, I. Prassas, P. Saraon, N. Musrap, A. Dimitromanolakis, E.P. Diamandis, Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis, Oncotarget 3(3) (2012) 267-85.

249. G. Clair, P.D. Piehowski, T. Nicola, J.A. Kitzmiller, E.L. Huang, E.M. Zink, R.L. Sontag, D.J. Orton, R.J. Moore, J.P. Carson, R.D. Smith, J.A. Whitsett, R.A. Corley, N. Ambalavanan, C. Ansong, Spatially-Resolved Proteomics: Rapid Quantitative Analysis of Laser Capture Microdissected Alveolar Tissue Samples, Sci Rep 6 (2016) 39223.

250. C. Guzel, N.I. Govorukhina, G.B.A. Wisman, C. Stingl, L.J.M. Dekker, H.G. Klip, H. Hollema, V. Guryev, P.L. Horvatovich, A.G.J. van der Zee, R. Bischoff, T.M. Luider, Proteomic alterations in early stage cervical cancer, Oncotarget 9(26) (2018) 18128-18147.

251. D.J. Ryan, J.M. Spraggins, R.M. Caprioli, Protein identification strategies in MALDI imaging mass spectrometry: a brief review, Curr Opin Chem Biol 48 (2019) 64-72.

252. S.K. Maier, H. Hahne, A.M. Gholami, B. Balluff, S. Meding, C. Schoene, A.K. Walch, B. Kuster, Comprehensive identification of proteins from MALDI imaging, Mol Cell Proteomics 12(10) (2013) 2901-10.

253. L.H. Cazares, D.A. Troyer, B. Wang, R.R. Drake, O.J. Semmes, MALDI tissue imaging: from biomarker discovery to clinical applications, Anal Bioanal Chem 401(1) (2011) 17-27.

254. L. Mourino-Alvarez, I. Iloro, F. de la Cuesta, M. Azkargorta, T. Sastre-Oliva, I. Escobes, L.F. Lopez-Almodovar, P.L. Sanchez, H. Urreta, F. Fernandez-Aviles, A. Pinto, L.R. Padial, F. Akerstrom, F. Elortza, M.G. Barderas, MALDI-Imaging Mass Spectrometry: a step forward in the anatomopathological characterization of stenotic aortic valve tissue, Sci Rep 6 (2016) 27106.

255. M. Dilillo, D. Pellegrini, R. Ait-Belkacem, E.L. de Graaf, M. Caleo, L.A. McDonnell, Mass Spectrometry Imaging, Laser Capture Microdissection, and LC-MS/MS of the Same Tissue Section, J Proteome Res 16(8) (2017) 2993-3001.

256. M. Aichler, A. Walch, MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice, Lab Invest 95(4) (2015) 422-31.

257. R.G. de Vega, M.L.F. Sanchez, N. Eiro, F.J. Vizoso, M. Sperling, U. Karst, A.S. Medel, Multimodal laser ablation/desorption imaging analysis of Zn and MMP-11 in breast tissues, Anal Bioanal Chem 410(3) (2017) 913-922.

258. D. Luptakova, L. Baciak, T. Pluhacek, A. Skriba, B. Sediva, V. Havlicek, I. Juranek, Membrane depolarization and aberrant lipid distributions in the neonatal rat brain following hypoxic-ischaemic insult, Sci Rep 8(1) (2018) 6952.

259. P. Magangane, R. Sookhayi, D. Govender, R. Naidoo, Determining protein biomarkers for DLBCL using FFPE tissues from HIV negative and HIV positive patients, J Mol Histol 47(6) (2016) 565-577.

260. P.M. Angel, H.S. Baldwin, D. Gottlieb Sen, Y.R. Su, J.E. Mayer, D. Bichell, R.R. Drake, Advances in MALDI imaging mass spectrometry of proteins in cardiac tissue, including the heart valve, Biochim Biophys Acta Proteins Proteom 1865(7) (2017) 927-935.

261. Integraal Kankercentrum Nederland, 2018. https://www.cijfersoverkanker.nl.

262. M. Riihimaki, A. Hemminki, J. Sundquist, K. Hemminki, Patterns of metastasis in colon and rectal cancer, Sci Rep 6 (2016) 29765.

263. P.D. Piehowski, V.A. Petyuk, D.J. Orton, F. Xie, R.J. Moore, M. Ramirez-Restrepo, A. Engel, A.P. Lieberman, R.L. Albin, D.G. Camp, R.D. Smith, A.J. Myers, Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis, J Proteome Res 12(5) (2013) 2128-37.

264. S.A. Carr, S.E. Abbatiello, B.L. Ackermann, C. Borchers, B. Domon, E.W. Deutsch, R.P. Grant, A.N. Hoofnagle, R. Huttenhain, J.M. Koomen, D.C. Liebler, T. Liu, B. MacLean, D.R. Mani, E. Mansfield, H. Neubert, A.G. Paulovich, L. Reiter, O. Vitek, R. Aebersold, L. Anderson, R. Bethem, J. Blonder, E. Boja, J. Botelho, M. Boyne, R.A. Bradshaw, A.L. Burlingame, D. Chan, H. Keshishian, E. Kuhn, C. Kinsinger, J.S. Lee, S.W. Lee, R. Moritz, J. Oses-Prieto, N. Rifai, J. Ritchie, H. Rodriguez, P.R. Srinivas, R.R. Townsend, J. Van Eyk, G. Whiteley, A. Wiita, S. Weintraub, Targeted peptide measurements

in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach, Mol Cell Proteomics 13(3) (2014) 907-17.

265. B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics 26(7) (2010) 966-8.

266. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/, 2016.

267. A. White, L. Ironmonger, R.J.C. Steele, N. Ormiston-Smith, C. Crawford, A. Seims, A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK, BMC Cancer 18(1) (2018) 906.

268. R.A. Gjaltema, R.A. Bank, Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease, Crit Rev Biochem Mol Biol 52(1) (2017) 74-95.

269. M.J. Bissel, W.C. Hines, Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression, Nature Medicine 17(3) (2011) 320-329.

270. S. Gout, J. Huot, Role of cancer microenvironment in metastasis: focus on colon cancer, Cancer Microenviron 1(1) (2008) 69-83.

271. D.F. Quail, J.A. Joyce, Microenvironmental regulation of tumor progression and metastasis, Nat Med 19(11) (2013) 1423-37.

272. P.P. Provenzano, D.R. Inman, K.W. Eliceiri, J.G. Knittel, L. Yan, C.T. Rueden, J.G. White, P.J. Keely, Collagen density promotes mammary tumor initiation and progression, BMC Med 6 (2008) 11.

273. H. Nystrom, P. Naredi, A. Berglund, R. Palmqvist, B. Tavelin, M. Sund, Liver-metastatic potential of colorectal cancer is related to the stromal composition of the tumour, Anticancer Res 32(12) (2012) 5183-91.

274. Z.S. Lalmahomed, M.E. Broker, N.A. van Huizen, R.R. Coebergh van den Braak, L.J. Dekker, D. Rizopoulos, C. Verhoef, E.W. Steyerberg, T.M. Luider, J.N.M. IJzermans, Hydroxylated collagen peptide in urine as biomarker for detecting colorectal liver metastases, Am J Cancer Res 6(2) (2016) 321-30.

275. C. Kahlert, M. Pecqueux, N. Halama, H. Dienemann, T. Muley, J. Pfannschmidt, F. Lasitschka, F. Klupp, T. Schmidt, N. Rahbari, C. Reissfelder, C. Kunz, A. Benner, C. Falk, J. Weitz, M. Koch, Tumour-site-dependent expression profile of angiogenic factors in tumour-associated stroma of primary colorectal cancer and metastases, Br J Cancer 110(2) (2014) 441-9.

276. Z.S. Zeng, J.G. Guillem, Unique activation of matrix metalloproteinase-9 within human liver metastasis from colorectal cancer, Br J Cancer 78(3) (1998) 349-53.

277. S. Zucker, J. Vacirca, Role of matrix metalloproteinases (MMPs) in colorectal cancer, Cancer Metastasis Rev 23(1-2) (2004) 101-17.

278. A. Fabregat, K. Sidiropoulos, P. Garapati, M. Gillespie, K. Hausmann, R. Haw, B. Jassal, S. Jupe, F. Korninger, S. McKay, L. Matthews, B. May, M. Milacic, K. Rothfels, V. Shamovsky, M. Webber, J. Weiser, M. Williams, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The Reactome pathway Knowledgebase, Nucleic Acids Res 44(D1) (2016) D481-7.

279. D. Croft, A.F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M.R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, P. D'Eustachio, The Reactome pathway knowledgebase, Nucleic Acids Res 42(Database issue) (2014) D472-7.

280. N. Fukai, L. Eklund, A.G. Marneros, S.P. Oh, D.R. Keene,

L. Tamarkin, M. Niemela, M. Ilves, E. Li, T. Pihlajaniemi, B.R. Olsen, Lack of collagen XVIII/endostatin results in eye abnormalities, EMBO J 21(7) (2002) 1535-44.

281. J. Saarela, M. Rehn, A. Oikarinen, H. Autio-Harmainen, T. Pihlajaniemi, The short and long forms of type XVIII collagen show clear tissue specificities in their expression and location in basement membrane zones in humans, Am J Pathol 153(2) (1998) 611-26.

282. O. Musso, M. Rehn, J. Saarela, N. Theret, J. Lietard, Hintikka, D. Lotrian, J.P. Campion, T. Pihlajaniemi, B. Clement, Collagen XVIII is localized in sinusoids and basement membrane zones and expressed by hepatocytes and activated stellate cells in fibrotic human liver, Hepatology 28(1) (1998) 98-107.

283. O. Musso, N. Theret, R. Heljasvaara, M. Rehn, B. Turlin, J.P. Campion, T. Pihlajaniemi, B. Clement, Tumor hepatocytes and basement membrane-Producing cells specifically express two different forms of the endostatin precursor, collagen XVIII, in human liver cancers, Hepatology 33(4) (2001) 868-76.

284. D.M. Gilkes, P. Chaturvedi, S. Bajpai, C.C. Wong, H. Wei, S. Pitcairn, M.E. Hubbi, D. Wirtz, G.L. Semenza, Collagen prolyl hydroxylases are essential for breast cancer metastasis, Cancer Res 73(11) (2013) 3285-96.

285. R. Afik, E. Zigmond, M. Vugman, M. Klepfish, E. Shimshoni, M. Pasmanik-Chor, A. Shenoy, E. Bassat, Z. Halpern, T. Geiger, I. Sagi, C. Varol, Tumor macrophages are pivotal constructors of tumor collagenous matrix, J Exp Med 213(11) (2016) 2315-2331.

286. H. Du, M. Pang, X. Hou, S. Yuan, L. Sun, PLOD2 in cancer research, Biomed Pharmacother 90 (2017) 670-676.

287. K. Mori, Y. Toiyama, K. Otake, H. Fujikawa, S. Saigusa, J. Hiro, M. Kobayashi, M. Ohi, K. Tanaka, Y. Inoue, Y. Kobayashi, I. Kobayashi, Y. Mohri, A. Goel, M. Kusunoki, Proteomics analysis of differential protein expression identifies heat shock protein 47 as a predictive marker for lymph node metastasis in patients with colorectal cancer, Int J Cancer 140(6) (2017) 1425-1435.

288. B. Seubert, B. Grunwald, J. Kobuch, H. Cui, F. Schelter, S. Schaten, J.T. Siveke, N.H. Lim, H. Nagase, N. Simonavicius, M. Heikenwalder, T. Reinheckel, J.P. Sleeman, K.P. Janssen, P.A. Knolle, A. Kruger, Tissue inhibitor of metalloproteinases (TIMP)-1 creates a premetastatic niche in the liver through SDF-1/CXCR4-dependent neutrophil recruitment in mice, Hepatology 61(1) (2015) 238-48.

289. H. Peinado, H. Zhang, I.R. Matei, B. Costa-Silva, A. Hoshino, G. Rodrigues, B. Psaila, R.N. Kaplan, J.F. Bromberg, Y. Kang, M.J. Bissell, T.R. Cox, A.J. Giaccia, J.T. Erler, S. Hiratsuka, C.M. Ghajar, D. Lyden, Pre-metastatic niches: organ-specific homes for metastases, Nat Rev Cancer 17(5) (2017) 302-317.

290. R.N. Kaplan, S. Rafii, D. Lyden, Preparing the "soil": the premetastatic niche, Cancer Res 66(23) (2006) 11089-93.

291. H. Peinado, M. Aleckovic, S. Lavotshkin, I. Matei, B. Costa-Silva, G. Moreno-Bueno, M. Hergueta-Redondo, C. Williams, G. Garcia-Santos, C. Ghajar, A. Nitadori-Hoshino, C. Hoffman, K. Badal, B.A. Garcia, M.K. Callahan, J. Yuan, V.R. Martins, J. Skog, R.N. Kaplan, M.S. Brady, J.D. Wolchok, P.B. Chapman, Y. Kang, J. Bromberg, D. Lyden, Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET, Nat Med 18(6) (2012) 883-91.

292. X. Sole, M. Crous-Bou, D. Cordero, D. Olivares, E. Guino, R. Sanz-Pamplona, F. Rodriguez-Moranta, X. Sanjuan, J. de Oca, R. Salazar, V. Moreno, Discovery and validation of new potential biomarkers for early detection of colon cancer, PLoS One 9(9) (2014) e106748.

293. M. Cerna, L. Holubec, Jr., M. Pesta, S. Kormunda, O. Topolcan, R. Cerny, Quantitative estimation of CEA and CK20 expression in tumour tissue of colorectal

cancer and its liver metastases with reverse transcription and real-time PCR, Anticancer Res 26(1B) (2006) 803-8.

294. B.K. Gupta, D.M. Maher, M.C. Ebeling, V. Sundram, M.D. Koch, D.W. Lynch, T. Bohlmeyer, A. Watanabe, H. Aburatani, S.E. Puumala, M. Jaggi, S.C. Chauhan, Increased expression and aberrant localization of mucin 13 in metastatic colon cancer, J Histochem Cytochem 60(11) (2012) 822-31.

295. R.D. Blumenthal, E. Leon, H.J. Hansen, D.M. Goldenberg, Expression patterns of CEACAM5 and CEACAM6 in primary and metastatic cancers, BMC Cancer 7 (2007) 2.

296. A.M. Scott, F.T. Lee, R. Jones, W. Hopkins, D. MacGregor, J.S. Cebon, A. Hannah, G. Chong, P. U, A. Papenfuss, A. Rigopoulos, S. Sturrock, R. Murphy, V. Wirth, C. Murone, F.E. Smyth, S. Knight, S. Welt, G. Ritter, E. Richards, E.C. Nice, A.W. Burgess, L.J. Old, A phase I trial of humanized monoclonal antibody A33 in patients with colorectal carcinoma: biodistribution, pharmacokinetics, and quantitative tumor uptake, Clin Cancer Res 11(13) (2005) 4810-7.

297. L. Belov, J. Zhou, R.I. Christopherson, Cell surface markers in colorectal cancer prognosis, Int J Mol Sci 12(1) (2010) 78-113.

298. R.A. Bartolome, R. Barderas, S. Torres, M.J. Fernandez-Acenero, M. Mendes, J. Garcia-Foncillas, M. Lopez-Lucendo, J.I. Casal, Cadherin-17 interacts with alpha2beta1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis, Oncogene 33(13) (2014) 1658-69.

299. M. Kopljar, L. Patrlj, D. Korolija-Marinic, M. Horzic, K. Cupurdija, B. Bakota, High Expression of DARPP-32 in Colorectal Cancer Is Associated With Liver Metastases and Predicts Survival for Dukes A and B Patients: Results of a Pilot Study, Int Surg 100(2) (2015) 213-20.

300. X.Z. Wu, F. Ma, X.L. Wang, Serological diagnostic factors for liver metastasis in patients with colorectal cancer, World J Gastroenterol 16(32) (2010) 4084-8.

301. Y. Kawashima, Y. Kodera, A. Singh, M. Matsumoto, H. Matsumoto, Efficient extraction of proteins from formalin-fixed paraffin-embedded tissues requires higher concentration of tris(hydroxymethyl) aminomethane, Clin Proteomics 11(1) (2014) 4.

302. J.V. Olsen, S.E. Ong, M. Mann, Trypsin cleaves exclusively C-terminal to arginine and lysine residues, Mol Cell Proteomics 3(6) (2004) 608-14.

303. V. Singh, E.D. van Pelt, M.P. Stoop, C. Stingl, I.A. Ketelslegers, R.F. Neuteboom, C.E. Catsman-Berrevoets, T.M. Luider, R.Q. Hintzen, Gray matter-related proteins are associated with childhood-onset multiple sclerosis, Neurol Neuroimmunol Neuroinflamm 2(5) (2015) e155.

304. K.L. Gorres, R.T. Raines, Prolyl 4-hydroxylase, Crit Rev Biochem Mol Biol 45(2) (2010) 106-124.

305. P. Tianen, A. Pasanen, R. Sormunen, J. Myllyharju, Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV, J Biol Chem 283(28) (2008) 19432-19439.

306. C.L. Jenkins, L.E. Bretscher, A. Guzei, R.T. Raines, Effect of 3-hydroxyproline residues on collagen stability, J Am Chem Soc 125(21) (2003) 6422-6427.

307. M.A. Weis, D.M. Hudson, L. Kim, M. Scott, J.J. Wu, D.R. Eyre, Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly, J Biol Chem 285(4) (2010) 2580-2590.

308. D.R. Eyre, M.A. Weis, D.M. Hudson, J.J. Wu, L. Kim, A novel 3-hydroxyproline (3Hyp)-rich motif marks the triple-helical C terminus of tendon type I collagen, J Biol Chem 286(10) (2011) 7732-7736.

309. M. Yamauchi, M. Shiiba, Posttranslational modifications of proteins, Humana Press2002.

310. J. Myllyharju, k.I. Kivirikko, Collagens, modifying

enzymes and their mutations in humand, flies and worms, Trends Genet 20(1) (2004) 33-43.

311. K. Mizuno, S.P. Buodko, J. Engel, H.P. Bächinger, Kinetic hysteresis in collagen folding, Biophys J 98(12) (2010) 3004-3014.

312. J.A.M. Ramshaw, N.K. Shah, B. Brodsky, Gly-X-Y tripeptide frequencies in collagen: a context for host-guest tiple-helical peptides, J Struct Biol 122(1-2) (1998) 86-91.

313. E. Leikina, M.V. Mertts, N. Kuznetsova, S. Leikin, Type I collagen is thermally unstable at body temperature, Proc Natl Acad Sci USA 99(3) (2002) 1314-1318.

314. J.C. Utting, S.P. Robins, A. Brandao-Burch, I.R. Orriss, J. Behar, T.R. Arnett, Hypoxia inhibits the growth, differentiation and bone-forming capacity of rat osteoblasts, Exp Cell Res 312(10) (2006) 1693-1702.

315. N.I. Nissen, M. Karsdal, N. Willumsen, Collagens and Cancer associated fibroblasts in the reactive stroma and its relation to Cancer biology, J Exp Clin Cancer Res 38(1) (2019) 115.

316. M.S. Molony, C. Quan, M.G. Mulkerrin, R.J. Harris, Hydroxylation of Lys residues reduces their susceptibility to digestion by trypsin and lysyl endopeptidase, Anal Biochem 258(1) (1998) 136-7.

317. T. Slechtova, M. Gilar, K. Kalikova, E. Tesarova, Insight into Trypsin Miscleavage: Comparison of Kinetic Constants of Problematic Peptide Sequences, Anal Chem 87(15) (2015) 7636-43.

318. J.A. Siepen, E.J. Keevil, D. Knight, S.J. Hubbard, Prediction of missed cleavage sites in tryptic peptides aids protein identification in proteomics, J Proteome Res 6(1) (2007) 399-408.

319. S.R. McKeown, Defining normoxia, physoxia and hypoxia in tumours-implications for treatment response, Br J Radiol 87(1035) (2014) 20130676.

320. C. Yang, A.C. Park, N.A. Davis, J.D. Russel, B. Kim, D.D. Brand, M.J. Lawrence, Y. Ge, M.S. Westphall, J.J. Coon, D.S. Greenspan, Comprehensive mass spectrometric mapping of the hydroxylated amino acid residues of the a1(V) collagen chain, J Biol Chem 287(48) (2012) 40598-40610.

321. M.J. Frisch, M. Headgordon, J.A. Pople, A Direct Mp2 Gradient-Method, Chem Phys Lett 166(3) (1990) 275-280.

322. S. Hayakawa, M. Hashimoto, H. Matsubara, F. Turecek, Dissecting the proline effect: dissociations of proline radicals formed by electron transfer to protonated Pro-Gly and Gly-Pro dipeptides in the gas phase, J Am Chem Soc 129(25) (2007) 7936-49.

323. C. Spahr, K. Gunasekaran, K.W. Walker, S.D. Shi, High-resolution mass spectrometry confirms the presence of a hydroxyproline (Hyp) post-translational modification in the GGGGP linker of an Fc-fusion protein, MAbs 9(5) (2017) 812-819.

324. T. Zhou, L. Erber, B. Liu, Y. Gao, H.B. Ruan, Y. Chen, Proteomic analysis reveals diverse proline hydroxylation-mediated oxygen-sensing cellular pathways in cancer cells, Oncotarget 7(48) (2016) 79154-79169.

325. E.Y. Jones, A. Miller, Analysis of structural design features in collagen, J Mol Biol 218(1) (1991) 209-219.

326. L.E. Bretcher, C.L. Jenkins, K.M. Taylor, M.L. DeRider, R.T. Raines, Conformational stability of collagen relies on a stereoelectronic effect, J Am Chem Soc 123(4) (2001) 777-778.

327. F.A. Venning, L. Wullkopf, J.T. Erler, Targeting ECM Disrupts Cancer Progression, Front Oncol 5 (2015) 224.

328. J. Sapudom, T. Pompe, Biomimetic tumor microenvironments based on collagen matrices, Biomater Sci 6(8) (2018) 2009-2024.

329. B.F. Matte, A. Kumar, J.K. Placone, V.G. Zanella, M.D. Martins, A.J. Engler, M.L. Lamers, Matrix stiffness mechanically conditions EMT and migratory behavior of oral squamous cell carcinoma, J Cell Sci 132(1) (2019).

330. B.K. Brisson, E.A. Mauldin, W. Lei, L.K. Vogel, A.M. Power, A. Lo, D. Dopkin, C. Khanna, R.G. Wells, E. Pure, S.W. Volk, Type III Collagen Directs Stromal Organization and Limits Metastasis in a Murine Model of Breast Cancer, Am J Pathol 185(5) (2015) 1471-86.

331. E. Makareeva, S. Han, J.C. Vera, D.L. Sackett, K. Holmbeck, C.L. Phillips, R. Visse, H. Nagase, S. Leikin, Carcinomas contain a matrix metalloproteinase-resistant isoform of type I collagen exerting selective support to invasion, Cancer Res 70(11) (2010) 4366-74.

332. Y. Oka, I. Naito, K. Manabe, Y. Sado, H. Matsushima, Y. Ninomiya, M. Mizuno, T. Tsuji, Distribution of collagen type IV alpha1-6 chains in human normal colorectum and colorectal cancer demonstrated by immunofluorescence staining using chain-specific epitope-defined monoclonal antibodies, J Gastroenterol Hepatol 17(9) (2002) 980-6.

333. P.J. van Dam, S. Daelemans, E. Ross, Y. Waumans, S. Van Laere, E. Latacz, R. Van Steen, C. De Pooter, M. Kockx, L. Dirix, P.B. Vermeulen, Histopathological growth patterns as a candidate biomarker for immunomodulatory therapy, Semin Cancer Biol 52(Pt 2) (2018) 86-93.

334. H. Nystrom, P. Naredi, L. Hafstrom, M. Sund, Type IV collagen as a tumour marker for colorectal liver metastases, Eur J Surg Oncol 37(7) (2011) 611-7.

335. S.N. Kehlet, R. Sanz-Pamplona, S. Brix, D.J. Leeming, M.A. Karsdal, V. Moreno, Excessive collagen turnover products are released during colorectal cancer progression and elevated in serum from metastatic colorectal cancer patients, Sci Rep 6 (2016) 30599.

336. L.D. Maxim, R. Niebo, M.J. Utell, Screening tests: a review with examples, Inhal Toxicol 26(13) (2014) 811-28.

337. J. Haga, M. Shimazu, G. Wakabayashi, M. Tanabe, S. Kawachi, Y. Fuchimoto, K. Hoshino, Y. Morikawa, M. Kitajima, Y. Kitagawa, Liver regeneration in donors and adult recipients after living donor liver transplantation, Liver Transpl 14(12) (2008) 1718-24.

338. G. Broughton II, R.J. Rohrich, Wounds and Scars, SRPS 10(7) (2005) 1-56.

339. T. Yoshimoto, Y. Morine, S. Imura, T. Ikemoto, S. Iwahashi, Y.U. Saito, S. Yamada, D. Ishikawa, H. Teraoku, M. Yoshikawa, J. Higashijima, C. Takasu, M. Shimada, Maximum Diameter and Number of Tumors as a New Prognostic Indicator of Colorectal Liver Metastases, In Vivo 31(3) (2017) 419-423.

340. K. Ichida, K. Suzuki, Y. Muto, T. Fukui, Y. Takayama, K. Futsuhara, S. Tsujinaka, Y. Miyakura, H. Noda, T. Rikiyama, Significance of the difference in size of liver tumors in management of patients with colorectal liver metastases, Annals of Oncology 27 (2016).

341. A. Cucchetti, A. Ferrero, M. Cescon, M. Donadon, N. Russolillo, G. Ercolani, G. Stacchini, F. Mazzotti, G. Torzilli, A.D. Pinna, Cure model survival analysis after hepatic resection for colorectal liver metastases, Ann Surg Oncol 22(6) (2015) 1908-14.

APPENDICES

Acknowledgements
List of publications
PhD portfolio
Biography

ACKNOWLEDGEMENTS

Beste Prof. **IJzermans**, Als eerste wil ik u bedanken voor de mogelijkheid om mijn PhD op de afdeling Heelkunde van het Erasmus MC te hebben mogen uitvoeren. Ik heb altijd genoten van onze 2-wekelijkse werkbesprekingen. U heeft altijd scherpe vragen, en als ik soms iets afdwaalde m.b.t. de data analyse wilde u altijd weten hoe dit de patiënt nu precies ten goede komt. Ik kijk uit naar onze verdere samenwerking in de toekomst.

Beste **Theo**, ik wil je enorm bedanken voor alle mogelijkheden die je mij hebt geboden om onderzoek te doen. Ik ben eind augustus 2010 binnen gekomen voor een derdejaars bachelor stage en eigenlijk nooit meer weg geweest. Ik heb enorm veel geleerd tijdens deze periode, en hoop nog veel nieuwe dingen te mogen leren en ontdekken tijdens onze verdere samenwerking.

Beste Prof. **Sillevis Smitt** graag wil ik u ook bedanken voor de mogelijkheid geboden om mijn PhD traject op de afdeling neurologie van het Erasmus MC uit te voeren.

Ik wil graag de **overige leden van de commissie** bedanken dat ze in de oppositie hebben plaatsgenomen voor de verdediging van dit proefschrift.

The statistics performed in this thesis have been mostly performed in R. I would like to thank the very active **R-community**. I was able to find my way in R by just scrolling through endless posts on fora and many tutorial websites. I was also able to pick-up statistics, which was very helpful.

Peter, beiden zijn we van oorsprong analytisch chemici en, misschien wel daardoor, zochten we elkaar vaak op. Hier en daar heb ik je kunnen helpen met projecten die ver van mijn project aflagen, maar die wat mij betreft niet minder interessant zijn! Ik vind het heel erg leuk dat je naast mij wilt staan als paranimf.

Lennard, Ik zal mijn eerste trip naar Amerika nooit vergeten, wat een prachtige plekken hebben we bezocht. Daarnaast sta je altijd klaar als ik vragen heb of voor een kletspraatje. Het is voor mij daarom ook bijna vanzelfsprekend dat je naast mij staat als paranimf.

Christoph, je enorme kennis wat betreft massaspectrometrie, vloeistofchromatografie, en bijbehorende software is indrukwekkend en ik heb veel van je kunnen leren.

A

Coskun, ik kan altijd binnen lopen op je kamer voor praatje of een serieuze wetenschappelijke discussie waar dat ook in het in gebouw mocht zijn. Ik heb dat altijd erg gewaardeerd. We zitten gelijktijdig in de laatste fase van ons PhD-traject, ik wens je veel succes met je verdediging.

Martijn, dankzij jouw heb ik R leren kennen wat me enorm veel werk in Excel heeft gescheeld. Jouw en **Eric's** "technical support" wat betreft computers en aanverwante zaken hebben me vaker geholpen dan ik zou willen toegeven.

Lona, je grote inzet zorgde ervoor dat toen ik op het lab kwam je vele protocollen had klaar liggen die ik kon gebruiken, ik heb veel van je kunnen leren. Ik vrees dat zonder jou ons lab ook niet zo netjes zou zijn.

Ik ben ooit aan dit project begonnen als Research analist, en stond er gelukkig niet alleen voor. **Mirelle** en **Zarina**, jullie waren op dat moment PhD-studenten op dit research veld en later opgevolgd door **Robert**. Dankzij jullie organiseer- en verzameltalent, staat er bij ons op het lab nu een vriezer vol met urine samples waar ik dankbaar gebruik van heb gemaakt en verder kan maken.

Michael Doukas, je hebt altijd stapels met coupes liggen om te bestuderen die volgens mij ook vaak meer prioriteit hebben dan mijn coupes. Desondanks maakte je altijd tijd vrij om met mij naar coupes te kijken en mij het één en ander uit te leggen. Dank je wel.

Verder wil ik nog graag bedanken **Diana**, **Gero**, **Henk**, **Ingrid**, **Marcel**, **Roland**, **Sadaf, Somayya**, **Yesim,** en **alle andere (oud) collega's en stagiaires** voor de fijne tijd, lunches, en alle discussies en leuke gesprekken.

Mama, **Papa**, jullie hebben mij altijd de kans en de mogelijkheid gegeven om te studeren, zonder die steun had ik nooit zover gekomen. **Mike**, als jij me niet geholpen had met het leren van de structuren van alle aminozuren was er vrees ik niets van terecht gekomen. Daarnaast kan ik bij jou altijd terecht voor een ontspannende game.

Marina, I remember clearly when you started working in the Erasmus MC. You have grown from an intern to a PhD student and I am very proud of you, soon you will also have to defend your thesis. I have learned a lot from you not only scientifically, but in all fields of life. Thank you for your tremendous amount of support and look forward to enjoy a life enduring collaboration.

LIST OF PUBLICATIONS

N.A. van Huizen, J.L. Holmes, and P.C. Burgers, One electron less or one proton more: how do they differ? J Mass Spectrom, 2019, Accepted.

N.A. van Huizen, J.N.M. IJzermans, P.C. Burgers, and T.M.Luider, Collagen analysis with mass spectrometry, Mass Spectrom Rev, 2019, Epub ahead of print

N.A. van Huizen, P.C. Burgers, F. Saintmont, P. Brocorens, P. Gerbaux, C. Stingl, L.J.M. Dekker, J.N.M. IJzermans and T.M. Luider, Identification of 4-Hydroxyproline at the Xaa Position in Collagen by Mass Spectrometry. J Proteome Res, 2019. 18(5): p. 2045-2051.

Q. Duez, **N.A. van Huizen**, V. Lemaur, J. de Winter, J. Cornil, P.C. Burgers, and P. Gerbaux, Silver ion induced folding of alkylamines observed by ion mobility experiments. Int J Mass Spectrom, 2019, 436: p. 34-41.

N.A. van Huizen, R. Coebergh van den Braak, M. Doukas, L.J. Dekker, J.N.M. IJzermans and T.M. Luider, Up-regulation of collagen proteins in colorectal liver metastasis compared with normal liver tissue. J Biol Chem, 2018. 294(1): p. 281-289.

J.L. Holmes, **N.A. van Huizen** and P.C. Burgers, Proton affinities and ion enthalpies. Eur J Mass Spectrom, 2017. 23(6): p. 341-350.

M.M.A. Verstegen, J. Willemse, S. van den Hoek, G.J. Kremers, T.M. Luider, **N.A. van Huizen**, F. Willemssen, H.J. Metselaar, J.N.M. IJzermans, L.J.W. van der Laan and J. de Jonge, Decellularization of Whole Human Liver Grafts Using Controlled Perfusion for Transplantable Organ Bioscaffolds. Stem Cells Dev, 2017. 26(18): p. 1304-1315.

Z.S. Lalmahomed, M.E.E. Broker, **N.A. van Huizen**, R. van den Braak, L.J.M. Dekker, D. Rizopoulos, C. Verhoef, E.W. Steyerberg, T.M. Luider and J.N.M. IJzermans, Hydroxylated collagen peptide in urine as biomarker for detecting colorectal liver metastases. Am J Cancer Res, 2016. 6(2): p. 321-330.

N.A. van Huizen, T.M. Luider, K.J. Jobst, J.K. Terlouw, J.L. Holmes and P.C. Burgers, Interaction of metal cations with functionalised hydrocarbons in the gas phase: further experimental evidence for solvation of metal ions by the hydrocarbon chain. Eur J Mass Spectrom, 2016. 22(2): p. 61-70.

K.J. Jobst, J.K. Terlouw, T. Luider, **N.A. van Huizen** and P.C. Burgers, Interaction of metal cations with alkylnitriles in the gas phase: solvation of metal ions by the hydrocarbon chain. Eur J Mass Spectrom (Chichester), 2015. 21(3): p. 579-87.

M.M.A. Verstegen, S. van den Hoek, G.J. Kremers, T.M. Luider, **N.A. van Huizen**, J.N.M. IJzermans, L.J.W. van der Laan and J. de Jonge, Repopulation and Revascularization of Human Decellularized Liver Grafts. Transplantation, 2015. 99: p. 116-116.

M.E.E. Broker, Z.S. Lalmahomed, H.P. Roest, **N.A. van Huizen**, L.J.M. Dekker, W. Calame, C. Verhoef, J.N.M. IJzermans and T.M. Luider, Collagen Peptides in Urine: A New Promising Biomarker for the Detection of Colorectal Liver Metastases. Plos One, 2013. 8(8): p. e70918.

R.J.W. Meesters, J. Zhang, **N.A. van Huizen**, G.P. Hooff, R.A. Gruters and T.M. Luider, Dried matrix on paper disks: the next generation DBS microsampling technique for managing the hematocrit effect in DBS analysis. Bioanalysis, 2012. 4(16): p. 2027-2035.

G.P. Hooff, R.J.W. Meesters, J.J.A. van Kampen, **N.A. van Huizen**, B. Koch, A.F.Y. Al Hadithy, T. van Gelder, A. Osterhaus, R.A. Gruters and T.M. Luider, Dried blood spot UHPLC-MS/MS analysis of oseltamivir and oseltamivircarboxylate-a validated assay for the clinic. Analytical and Bioanalytical Chemistry, 2011. 400(10): p. 3473-3479.

G.P. Hooff, **N.A. van Huizen**, R.J.W. Meesters, E.E. Zijlstra, M. Abdelraheem, W. Abdelraheem, M. Hamdouk, J. Lindemans and T.M. Luider, Analytical Investigations of Toxic p-Phenylenediamine (PPD) Levels in Clinical Urine Samples with Special Focus on MALDI-MS/MS. Plos One, 2011. 6(8).

R.J.W. Meesters, G.P. Hooff, **N.A. van Huizen**, R.A. Gruters and T.M. Luider, Impact of internal standard addition on dried blood spot analysis in bioanalytical method development. Bioanalysis, 2011. 3(20): p. 2357-2364.

PHD PORTFOLIO

Name PhD student:	Nick Arnold van Huizen
Erasmus MC department:	Surgery
PhD period:	Aug 2015 – Jan 2020
Promotors:	Prof. dr. J.N.M. IJzermans
	Prof. dr. P.A.E. Sillevis Smitt
Co-promotor:	Dr. T.M. Luider

CONFERENCES (16.4 ECTS)

Year	Topic	ECTs
2016	4th Daniel den Hoed day + oral	1.3
2017	European Bioinformatics Community + poster	2.4
2017	5th Daniel den Hoed day + oral	1.3
2017	American Society of Mass Spectrometry + poster	2.7
2017	Liver metastasis research network + oral	1.6
2017	Validation of Biomarkers	0.3
2017	The Erasmus MC Cancer Institute Research Day	0.3
2018	22th MolMed day + oral + poster	2.3
2018	American Society of Mass Spectrometry + poster	2.1
2018	Clinical update in DCD liver transplantation	0.1
2016-2019	JNI lectures + oral	2.0

COURSES (10.6 ECTS)

Year	Topic	ECTs
2016	Biomedical English writing course for MSc and PhD students	2.0
2016	Biomedical research techniques XV	1.5
2017	Microscope image analysis	0.8
2017	ESP2017 Multiple linear regression	1.4
2017	Research integrity	0.3
2017	Masterclass logframe	0.3
2017	JASP	0.3
2018	ASMS, Peptides and Proteins in Mass Spectrometry	0.6
2018	Workshop career development for PhD candidates	0.2
2018	Supervising Students	0.1
2019	Data analysis in Python	1.7
2019	ESP2019 Logistic regression	1.4

TEACHING (8 ECTS)

Year	Topic	ECTs
2016-2019	Laboratory tours	2.0
2016-2019	General training of students	2.0
2019	Supervision Bachelor internship	2.0

A

BIOGRAPHY

Nick Arnold van Huizen was born on the 31st of May, 1991 in Rotterdam, the Netherlands, and grew up in Numansdorp. In 2008 he graduated with a HAVO diploma from the Willem van Oranje, Oud-Beijerland, and started a bachelor Chemistry (minor: Analytical Chemisty), at the Hogeschool Rotterdam, which he graduated Cum Laude in 2012. During his bachelor, he did his 3rd year internship entitled: "Development and application of an ultra-fast MALDI-triple quadrupole tandem mass spectrometry assay for the investigation of toxic p-phenylenediamine concentrations in human urine" at the group of dr. T.M. Luider at the Erasmus MC, this internship would severely shape his future career. His 4th year graduation internship entitled: "The Radiolabelling of Micelles" was in the group of dr. A. Denkova at the TU Delft. During his bachelor's he had an extra-job as technical engineer at Inspectorate, and as research technician at the Erasmus MC. After his bachelor he graduated in 2015 with a master's degree in Chemistry (track: Analytical Sciences) at the University of Amsterdam. During his master's he worked part time as a research technician under the supervision of dr. T.M. Luider and prof J.N.M. IJzermans in the Erasmus MC. After finalizing his master's, he started a PhD entitled: "Collagen in colorectal cancer – A mass spectrometry analysis" under supervision of dr. T.M. Luider and prof J.N.M. IJzermans. His PhD project was a continuation of the collagen project he worked on as research technician. A grant from the European Union has been received, therefore he will continue as a Postdoc at the research group of dr. T.M. Luider and Prof IJzermans. During his PhD period he met Marina Zajec with whom he is living happily together. At the moment of writing they are expecting a baby girl.

A