

# **New Insights into Structure and Function of Type I Collagen**

Der Fakultät Energie-, Verfahrens- und Biotechnik  
der Universität Stuttgart  
zur Erlangung der Würde eines  
Doktor der Naturwissenschaften (Dr. rer. nat)  
Genehmigte Abhandlung

Vorgelegt von

**Xin Xiong**

aus V. R. China

Hauptberichter: Prof. Dr. H. Brunner

Mitberichter: Prof. Dr. R. Ghosh

Tag der mündlichen Prüfung: 29. August. 2008

Institut für Grenzflächenverfahrenstechnik der Universität Stuttgart

2008

# **New Insights into Structure and Function of Type I Collagen**

Der Fakultät Energie-, Verfahrens- und Biotechnik  
der Universität Stuttgart  
zur Erlangung der Würde eines  
Doktor der Naturwissenschaften (Dr. rer. nat)  
Genehmigte Abhandlung

Vorgelegt von  
**Xin Xiong**  
aus V. R. China

Hauptberichter: Prof. Dr. H. Brunner  
Mitberichter: Prof. Dr. R. Ghosh  
Tag der mündlichen Prüfung: 29. August. 2008

# Contents

<b>Contents.....</b>	<b>I</b>
<b>Zusammenfassung.....</b>	<b>IV</b>
<b>Abstract.....</b>	<b>VI</b>
<b>Abbreviations.....</b>	<b>VIII</b>
<b>1. Introduction.....</b>	<b>1</b>
<b>1.1 Collagen.....</b>	<b>1</b>
1.1.1 Major component of extracellular matrix.....	1
1.1.2 Type I collagen and its structure.....	3
1.1.3 Post-translational modifications of collagen and fibril formation.....	6
1.1.4 Self-assembly of collagen Type I and its aggregation.....	7
1.1.5 Cell-ECM interaction and role of collagen.....	10
1.1.6 Collagen as 3D-biomaterial/biomatrix for tissue engineering.....	14
1.1.7 Isolation and purification of collagens.....	17
1.1.8 Rat tail tendon and Type I collagen.....	18
1.1.9 Recombinant collagen.....	19
1.1.10 Applications of Type I collagen.....	20
<b>1.2 Mass spectrometry.....</b>	<b>21</b>
1.2.1 Introduction to mass spectrometry (MS).....	21
1.2.2 Tandem MS and protein identification.....	21
1.2.3 Mass spectrometry of collagen.....	24
<b>1.3 Goals of this study.....</b>	<b>25</b>
<b>2. Materials and Methods.....</b>	<b>26</b>
<b>2.1 Chemicals.....</b>	<b>26</b>
<b>2.2 Buffers, solutions and Media.....</b>	<b>26</b>
2.2.1 Commonly used buffers: Phosphate buffered saline (PBS), Tris HCl buffer..	26
2.2.2 Laemmli electrophoresis buffer.....	26
2.2.3 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).....	27
2.2.4 Western blot buffers.....	27
<b>2.3 Protein isolation from rat tail tendons (RTT).....</b>	<b>27</b>
2.3.1 Isolation and gel-filtration.....	27
2.3.2 Productivity determination.....	28
<b>2.4 Gel electrophoresis and Western blot.....</b>	<b>28</b>
2.4.1 SDS-PAGE analysis.....	28

## Contents

2.4.2 Gel staining .....	29
2.4.3 Western blot.....	29
<b>2.5 Chemical modifications of the collagen.....</b>	<b>30</b>
2.5.1 Biotinylation.....	30
2.5.2 Purification and concentration of biotinylated protein .....	30
2.5.3 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) determination of free thiols in collagen.....	30
<b>2.6 Cleavage of protein .....</b>	<b>31</b>
2.6.1 Cyanogen bromide cleavage.....	31
2.6.2 Trypsin digestion .....	31
2.6.3 Chymotrypsin and chymotrypsin/trypsin digestion.....	32
2.6.4 Carboxypeptidase Y digestion.....	32
<b>2.7 Mass determination.....</b>	<b>32</b>
<b>2.8 Tandem Mass spectrometric (MS/MS) analysis.....</b>	<b>33</b>
2.8.1 Electrospray ionization MS/MS.....	33
2.8.2 MALDI-TOF/TOF .....	33
<b>2.9 Spectroscopic analysis.....</b>	<b>34</b>
2.9.1 UV-spectroscopy.....	34
2.9.2 UV-CD-spectroscopy.....	34
2.9.3 Gaussian analysis of UV-VIS spectra .....	34
2.9.4 Histidine determination using UV-spectroscopy .....	34
<b>2.10 Reverse-phase-HPLC (RP-HPLC) .....</b>	<b>35</b>
<b>2.11 Atomic force microscopy. ....</b>	<b>35</b>
<b>2.12 Scanning electron microscopy (SEM) .....</b>	<b>35</b>
2.12.1 SEM of collagen.....	35
2.12.2 SEM of cells grown on different materials.....	35
<b>2.13 Cell culture.....</b>	<b>36</b>
<b>2.14. Real Time-PCR (RT-PCR).....</b>	<b>36</b>
<b>2.15 Type I collagen expression.....</b>	<b>38</b>
2.15.1 Cell culture and transfection .....	38
2.15.2 Expression and purification .....	39
<b>3. Results .....</b>	<b>40</b>
3.1 Collagen extraction and SDS-PAGE analysis.....	40

## Contents

<b>3.2 Superose 12 gel filtration and SDS-PAGE analysis .....</b>	<b>41</b>
<b>3.3 Mass determination and tandem MS/MS .....</b>	<b>43</b>
<b>3.4 UV-CD spectroscopy .....</b>	<b>50</b>
<b>3.5 UV-VIS spectroscopy of urea-extracted collagen.....</b>	<b>52</b>
<b>3.6 Carboxypeptidase and chymotrypsin digestion .....</b>	<b>53</b>
<b>3.7 Biotinylation of collagen .....</b>	<b>55</b>
<b>3.8 AFM and SEM of collagen .....</b>	<b>56</b>
<b>3.9 DTNB reaction and histidine measurement.....</b>	<b>57</b>
<b>3.10 Chromatographic analysis of cleaved collagen.....</b>	<b>58</b>
<b>3.11 Phenotypic comparison of AC and UC upon cell cultures .....</b>	<b>60</b>
<b>3.12 Real time PCR and regulation at transcriptional level.....</b>	<b>63</b>
<b>3.13 Type I Collagen expression.....</b>	<b>64</b>
<b>4. Discussion.....</b>	<b>66</b>
<b>REFERENCE .....</b>	<b>80</b>
<b>Figure index .....</b>	<b>89</b>
<b>Table index.....</b>	<b>90</b>
<b>Protein sequence alignment of Type I collagen from bovine, human and rat..</b>	<b>91</b>
<b>Acknowledgements.....</b>	<b>96</b>
<b>Curriculum vitae .....</b>	<b>98</b>
<b>Publications.....</b>	<b>99</b>
<b>Declaration.....</b>	<b>100</b>

## Zusammenfassung

### Zusammenfassung

Kollagen ist eines der abundantesten Proteine in Säugetieren und sehr stark konserviert. Es bildet ein Drittel des menschlichen Proteomes und macht Dreiviertel des Trockengewichts menschlicher Haut aus. Es ist weiter bekannt als wichtigster struktureller Bestandteil in Säugetieren wie z.B. in Knochen, Knorpel und Haut. Außer der strukturellen Funktion, hat Kollagen viele zelluläre Funktionen und ist in mehreren zellulären Prozessen wie Zelldifferenzierung, Zellbewegung, Zellkommunikation und Apoptose beteiligt. Seine einzigartige Triple-Helix-Struktur verleiht ihm hohe Stabilität. Wegen der leichten Zugänglichkeit, sind Typ I und Typ III Kollagen am besten erforscht und werden als biokompatible Materialien in Zellkultur, Tissue Engineering und Medizintechnik angewandt.

Bis heute ist das Verständnis der molekularen Mechanismen für die Faltung und Anordnung des Kollagens im Gewebe von großem medizinischem und auch biotechnologischen Interesse. In dieser Arbeit wurde ein Verfahren zur Isolierung und Aufreinigung von Typ I Kollagen aus Rattenschwanzsehnen etabliert. Dabei wird ein zweistufiger Reinigungsprozess mittels Extraktion mit 9 M Harnstoff und anschließender Superose 12 Chromatographie durchgeführt. Diese einfache Aufreinigung von Typ I Kollagen lässt sich die industrielle Produktion auch in größerem Maßstab aufbauen. Das Produkt kann lyophilisiert und dadurch gut lagerbar gemacht werden. AFM und SEM Aufnahmen zeigten eine natur-ähnliche Struktur des harnstoffisolierten Kollagens.

Dieses Kollagen wurde mittels verschiedenen biochemischen, physikalischen und zellbiologischen Methoden charakterisiert.

Massenspektrometrie hat nur Typ I Kollagen-Peptide identifiziert, damit die Reinheit des isolierten Kollagens nachgewiesen. Der Vergleich zwischen Harnstoff-extrahiertem (UC) Kollagen und Essigsäure-extrahiertem Kollagen (AC) zeigt signifikante Unterschiede. So hat das AC eine deutlich niedrigere Masse als UC. Dies kann auf partielle Hydrolyse oder enzymatischem Abbau während der Extraktion zurückzuführen sein. Tandem MS Analysen haben differentielle post-transnationale Modifikationen identifiziert, was einen neuen Einblick in Kollagenstruktur in vivo ermöglichen kann.

Das in 8 M Harnstoff gereinigte Kollagen wurde durch Dialyse gegen Wasser quantitativ renaturiert. Diese Renaturierung wurde mittels UV-Zirkularem Dichroismus verfolgt. Dabei wurde die Bildung einer Triple-Helix bestätigt. Das in 8 M Harnstoff aufgelöste Kollagen zeigt ein reversibles Aggregationsphänomen. UV-Zirkulare Dichroismus-Analyse zeigt, dass Kollagen bei Konzentrationen kleiner als 4 M Harnstoff eine Triple-Helix Struktur bildet. Diese Struktur ist auch vergleichbar zur der Struktur von synthetischen Kollagen-Peptiden.

## Zusammenfassung

Harnstoffextrahiertes Kollagen wurde auf seine Eignung für die Zellkultur untersucht. Dabei konnte gezeigt werden, dass die 3T3 Fibroblasten auf UC Kollagen eine höhere Motilität besitzen, als auf AC Kollagen. Die Analyse mittels real time-PCR zeigt die Unterschiede auf transkriptionale Ebene. Gene, die zuständig für Reaktion auf Stress sind, werden in 3T3 Fibroblasten auf AC im Vergleich zu UC induziert. Dies deutet auf ein erhöhtes Stress-Level bei AC Kollagen hin. Diese Ergebnisse können von Nutzen für biotechnologische/biomedizinische Anwendungen sein.

Harnstoffextrahiertes Kollagen zeigt eine einzigartige reversible Aggregation in 8 M Harnstoff während der Gelfiltration. Dabei konnten keine kovalenten Bindungen und Quervernetzungen intra- und intermolekular festgestellt werden, selbst unter reduzierenden Bedingungen. Diese Daten und weitergehende massenspektrometrische Analysen führten uns zum Entwurf eines Modells für die Reversibilität der Assoziation der drei Kollagenketten ( $2\alpha_1$ ,  $1\alpha_2$ ) sowie der Faltung an sich. Zusätzlich wurde eine Hypothese über den Exportmechanismus des Kollagens gewagt.

Die vorliegenden Ergebnisse haben zu einer Patentmeldung geführt (Patent wurde am 30.04.2008 erteilt). Eine Publikation wurde in FEBS J eingereicht.

## Abstract

### Abstract

Collagen is one of the most abundant proteins in mammals and strongly conserved throughout evolution. It constitutes one third of the human proteome and comprises three-quarters of the dry weight of human skin. It is widely accepted as a major structural component in animal body such as in bones, cartilage and skins. More and more studies have shown that, in addition to the structural function, collagens can induce or regulate many cellular functions and processes such as cell differentiation, cell motion, cell communication and apoptosis. Furthermore, its unique triple helix structure gained more attention since it is responsible for its high stability and biological function. Due to the high accessibility, Type I and Type III collagen are widely studied and frequently used as biocompatible materials in cell culture, tissue engineering and medical technology.

Until now the understanding of the molecular mechanisms for collagen assembly is of great medical and also biotechnological importance. Here, large amounts of highly purified homogeneous Type I collagen have been obtained from rat tail tendon by a simple two-step purification involving extraction with 9 M urea followed by Superose 12 chromatography. This simple two step purification of Type I collagen is up-scalable. The yield is up to 95%. The product could be easily lyophilized and stored. AFM and SEM images showed a structure similar to natural collagen.

This collagen was extensively characterized by different biochemical, physical and cell biological methods. Mass spectrometry identified only collagen Type I peptides indicating that the extracted collagen was homogeneous. The comparison between urea-extracted (UC) collagen and acetic acid-extracted collagen (AC) showed significant differences whereby the UC was not degraded or hydrolyzed as in acetic acid. Furthermore, tandem MS analyses showed some interesting post-translational modifications, which will result in new insights into collagen structure *in vivo*.

The purified collagen was renatured quantitatively by dialysis against water to form triple-helices, as judged by UV-circular dichroism. The collagen dissolved in 8M urea exhibits a unique reversible aggregation behavior which is not affected by the presence of reducing agents. UV-circular dichroism analysis shows that collagen initiates triple helix formation at 4 M urea or below. This triple helix structure is comparable to that observed with synthetic collagen peptides.

Cultures of a 3T3 mouse embryonic fibroblast cells incubated with urea-extracted collagen showed a higher motility than those grown with acetic acid-extracted collagen as judged by light microscopy and scanning electron microscopy. The real time PCR showed significant



## **Abstract**

difference on transcriptional level and showed clearly up regulation of the genes involved in response to mechanical stress in AC but not in UC and reference culture in medium. All these results indicate a benefit of UC for biotechnological/biomedical applications.

The urea-extracted collagen exhibits a unique reversible-aggregational phenomenon during gel filtration in 8 M urea. We could show that covalent bonds and cross-linkings are not involved. This observation and subsequently extensive mass spectrometric analyses led to a new model of triple-helical assembly and a hypothesis about the collagen export, which may offer some new insights into understanding of collagen structure and transport from cytosol to extracellular space.

The results presented here have led to an industrial patent (patented on 04. 30<sup>th</sup>. 2008). A manuscript was submitted to FEBS J for publication.

## Abbreviations

### Abbreviations

1D	One-dimensional
2D	Two-dimensional
AC	Acetic acid-extracted collagen
ACN	Acetonitrile
AFM	Atomic force microscopy
APS	Ammonium persulfate
$\beta$ -ME	$\beta$ -mercaptoethanol
CBB	Coomassie Brilliant Blue
CD	Circular dichroism
CHO	Chinese hamster ovary
ChyTrp	Chymotrypsin
CLP	Collagen-like protein
CNBr	Cyanogen bromide
Da	Dalton
D-MEM	Dulbecco's modified Eagle medium
DTNB	5,5'-dithiobis-2-nitrobenzoic acid
DTT	Dithiothreitol
ECM	Extracellular matrix
EDTA	Ethylene diamine tetracetic acid
ESI	Electrospray ionization
FCS	Fetal calf serum
FTMS	Fourier transform mass spectrometry
G3P	Glyceraldehyde-3-phosphate dehydrogenase
HPLC	High performance liquid chromatography
Hyl	Hydroxylysine
Hyp	Hydroxyproline
IAA	Iodacetamide
kDa	Kilodalton
M (mM)	mol/l (mmol/l)
MALDI	Matrix-assisted laser desorption/ionization
MBP	Maltose binding protein
MMP	Matrix metalloproteinase

## Abbreviations

MS	Mass spectrometry
PAGE	Polyacrylamide gel electrophoresis
PBS/PBST	Phosphate buffered saline/PBS+0.05% Tween 20
PCR	Polymerase chain reaction
POD	Peroxidase
PTM	Post-translational modifications
PVDF	Polyvinylidene difluoride
PMSF	Phenylmethylsulfonylfluoride
rpm	Revolutions per minute
RT-PCR	Real-time PCR
RTT	Rat tail tendon
SDS	Sodium dodecyl sulfate
SEM	Scanning electron microscopy
TEMED	N, N, N', N'-tetramethylethylenediamine
TFA	Trifluoroacetic acid
TOF	Time-of-flight
TWEEN 20	Polyoxyethylene (20) sorbitan monolaurate
UC	Urea-extracted collagen
UPL	Universal probe library
UV	Ultraviolet
w/v	weight/volume
v/v	volume/volume

# Introduction

## 1. Introduction

### 1.1 Collagen

#### 1.1.1 Major component of extracellular matrix

The extracellular matrix (ECM) may be looked as a glue of different macromolecules within the extracellular spaces which give a tissue its specific form and maintain its integrity [1, 2]. Apart from its structural functions the ECM regulate a variety of cell behaviors such as differentiation, motion and communication [3, 4]. Collagens are the most abundant proteins in all animals ECM and highly conserved throughout evolution [1, 3]. It constitutes one third of the human proteome and comprises three-quarters of the dry weight of human skin [5]. Increasingly, this family of proteins is gaining attention from scientists in different fields, ranging from biochemistry to regenerative medicine.

Collagen was first found and defined as gelatin by extractive cooking of bones at the beginning of the 1800's. This family of proteins was studied by scientists in different areas, such as in biochemistry, cell biology etc [3, 6-9]. The research on collagens is present with more than 100,000 publications, which include structural, chemical and cell biological aspects [3]. Despite of this intensity, the understanding of collagens is still limited and the numbers of collagen types or collagen like proteins (CLP) are increasing continuously. Until now, 28 types of collagen have been identified in vertebrates and numbered I-XXVIII [3]. Many CLPs such as acetyl cholinesterase, adiponectin and C1q are also classified [10]. The 28 types of different collagens consist of at least 42 distinct  $\alpha$  chains. Some collagens, such as Type I collagen, are heterotrimer containing two identical  $\alpha_1$  chains and a  $\alpha_2$  chain. However, most other collagens are homotrimers containing only identical  $\alpha_1$  chains, such as Type II collagen which consists of  $[\alpha_1]_3$ . In general,  $\alpha_1$  chains from one type of collagen differ in their supramolecular structure to those of another type collagen [3, 11].

Collagens are known as triple helical proteins which are widespread throughout the body and are important for a broad range of functions, including tissue scaffolding, cell adhesion, cell migration, cancer, angiogenesis, tissue morphogenesis and tissue repair [3]. Generally the vertebrate collagens are categorized in 7 sub-groups according to their structure and function: fibril-forming collagens, which include Type I, II, III, V, XI, XXIV, XXVII and compose 90% of collagen [2] thus they are the major protein in the ECM. Fibril-associated collagen with interrupted triple helix (FACITs) including Type IX, XII, XIV, XVI, XIX, XX, XXI, XXII and XXVI are widely spread with different functions. Network forming collagens including Type IV, VIII, X make basement membranes and Descemet's membrane.

## Introduction

Transmembrane collagens including XIII, XVII, XXIII and XXV also have different localizations and numerous functions, however the mechanism of this assembly of these groups is still unknown. Endostatin-producing collagens including XV and XVIII are closely homologous to each other and localize either closely to the basement membranes or associate with basement membranes. Anchoring fibrils and beaded-filament-forming collagens, also known as microfibrillar collagens [2, 3] including VI, VII, XXVI and XXVIII, are widely spread in the body [3, 11]. Only the Type I-XI have been identified and studied biochemically on a protein level, whereas all other types of collagens have only been identified by cDNA clones.

All collagens have the characteristic Gly-X-Y repeats whereby normally the X position is proline and the Y position is a 4-hydroxylated proline. While every third residue throughout the triple-helical region of the fibrillar collagens is Gly, the non-fibrillar collagens contain triple-helical domains with varying lengths. Therefore in all of these non-fibrillar collagens the Gly-X-Y pattern contains interruptions [12]. So far 354 interruptions have been found in the non-fibrillar collagens and the most common case is one residue between the two Gly-P-Hyp repeats, rather than two, three or more. Interestingly if only one residue interrupts the Gly-P-Hyp sequences in more than 70% of these cases the amino acid is found to be hydrophobic. In several well-studied non-fibrillar collagens these interruptions were reported to have different biological functions such as the binding site for tumor integrins or being the site for MMP cleavage [12].

An important reason for studying collagen is to understand the diseases caused by genetic defects of collagen encoding genes and abnormal collagen modifications. Abnormal degradation, cross-linking, or reactivity of collagen have been observed in many diseases such as arthritis, diabetes and cancer. Apart from these common diseases there are also some hereditary diseases caused by disorder of collagens. Since collagen is the major structural component in all organs, giving them mechanical support as well as participating in the organ development, collagen diseases can lead to severe pathological symptoms as well as death [13-15]. More than 200 mutations have been determined just for human Type I collagen coding genes which lead to osteogenesis imperfecta, osteoporosis and Ehlers-Danlos syndrome type VII A and B. These disorders weaken bones, skin and other collagen Type I containing organs, and in severe cases, death occurs shortly after the birth. Even in non-lethal forms, the patients suffer from permanent defects [14]. The collagen-related diseases are affected by the location of the mutation, dependent upon which chain is affected and if these sites are involved in the association to other matrix components or serve as binding sites for

## Introduction

the cell-collagen binding. To combat these diseases, it is meaningful to understand this family of proteins in detail.

### 1.1.2 Type I collagen and its structure

Collagen nomenclature is confusing because no strict criteria have been defined. Collagens should contain at least one triple helical domain, should be able to form supramolecular aggregates and should be deposited within the extracellular matrix. From the structural point of view, collagens consist of three chains, all in polyproline like-II conformation and supercoiled around the same axis. This structure is relatively rarely in other proteins and this triple helix differs basically from other helical structures, such as the coiled-coiled structure or DNA-double helix. Type I collagen is the most abundant protein in the animal body and diversely located in all ECM of connective tissues like bone, skin, tendon and blood vessels. The primary function of Type I collagen is to give mechanical support for structural organization of the ECM of connective tissues, therefore giving the tissues their characteristic form and maintaining their cellular integrity in the ECM. Possibly there are still many unknown functions to be found. Type I collagen represents the characteristic structure very well and has the characteristic large domains comprised of peptides with varied length of [Gly-X-Y] repeats of up to 300 nm in length and folded into unique triple-helical structure [14]. Every chain from the three polypeptide-chains  $[\alpha_1]_2[\alpha_2]$  form a left-handed helix then fold together to form a right-handed triple helix. The Gly residues in the trimer repeats are essential for folding and are packed into the center of the helix. No other amino acid larger than Gly would fit into this position. The next two amino acids at the X and Y position are normally Pro and Hyp residues which limit the rotation of the triple helix. Furthermore, Hyp residues form hydrogen bonds and water bridges which together stabilize the folded triple helix with the hydrophobic and charged side chains [14, 16]. Collagen assembly and folding has been determined as an entropy-driven self-assembly process. However, the rate of assembly between 37 °C and 41 °C is limited by the reversible micro-unfolding process [17]. Studies have also shown that Type I collagen in physiological solution is unstable and that the thermodynamically preferred structure of Type I collagen at body temperature (37 °C) should be random coil rather than the triple helix [18]. Leikina *et al.* have concluded that this thermal instability of Type I collagen has a real physiological significance. Thus the body may control the collagen fibril intrinsically unstable rather than stable and the chaperon may regulate the unfolding of the collagen fibrils to prevent the fibrillogenesis [18].

## Introduction

Classically, the proline and Hyp residues have the decisive role in stabilization of the triple helix through the hydrogen bonds. Both residues make 21% of the total sequence from the triple helix of Type I collagen (rat Type I collagen). Tertiary amides have dichotomous consequences for conformational stability whereby the pyrrolidine-rings constrain both residues and therefore stabilize the triple helix by this rigidity [19]. Especially the Hyp residues increase the stability strongly. Only 4(*R*)-hydroxyproline but not 4(*S*)-hydroxyproline shows a stabilizing effect leading to the conclusion that the –OH group contributes to this effect by either water-bridges between –OH groups and backbone groups or inductive effects. The latter hypothesis is much more accepted within the community.

Also *cis/trans* isomerisation of the peptide bonds plays a role in the stabilization of the collagen triple helix. Normally, collagen molecules in the native state contain only *trans* peptide bonds, therefore the *cis-trans* ratio is probably very important during the folding process, however this transition is still a controversially debated field.

Type I collagen is synthesized as procollagen Type I then processed in intracellular and extracellular spaces to collagen. Furthermore, in mammals, Type I collagen forms fibrils with typical axial periods of 67 nm, also defined as D-periods [20]. Although the triple helix is regarded as a rigid structure it shows dynamic motion dependent on the sequence. The [Gly-X-Y] motif is the most rigid part and most important for stabilizing the structure. Other amino acids contribute more to the biological functions such as receptor and ligand for cell-matrix contact [21]. Type I collagen has two non-helical regions at both termini, whereby the N-terminal domain is more stable than the C-terminal domain. Thus the C-terminal domain is more susceptible to enzymatic cleavage especially by collagenases.

There are two major methods to study collagen: using isolated natural collagen [17] and synthetic model peptide/protein [5, 22]. Both methods have their advantages and drawbacks. Most model peptides studies have focused on the trimer repeats of [Gly-X-Y] and have been mostly limited to [Gly-pro-Hyp]<sub>10</sub> [23], which is only a short peptide much shorter than any natural collagen (≈300nm). Although these model peptides yield much information about the dynamic process of the assembly and triple helix building *in vivo*, the information is just part of that necessary for the understanding of natural collagen. Studies have shown that polymerization of these model peptides yield the triple helix form but with a high polydispersity [5, 24]. Nevertheless, these model peptides have led to large progress in the clarification of the assembly of triple helix [21] by the application of X-ray crystallography, nuclear magnetic resonance (NMR), and other biophysical techniques. In the last two decades, many different spectroscopic techniques have been applied, such as optical rotary dispersion,

## Introduction

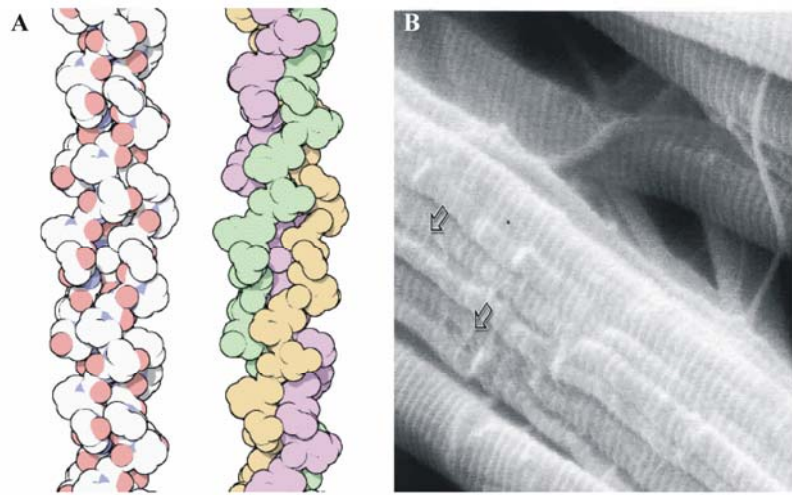
circular dichroism spectroscopy, and infrared spectroscopy, to establish relationships between amino acid sequence and triple helix stability and conformation [12]. Modern biophysical methods including X-ray crystallography and NMR were effectively applied to study the triple helical peptides but not the native collagen, because of the inability to obtain crystals of the whole collagen molecule, and the fact that the collagen molecule is unsuitable for multidimensional NMR. For these reasons, the structural and interaction studies on collagen still strongly rely on model peptides, particularly upon peptides with Gly-X-Y repeats. X-ray crystallography has also indicated that the rigidity of the triple helix varies in dependence of the amino acid sequences of the  $\alpha$  chains. Amino acid sequences containing only Gly-Pro-Hyp or Gly-Pro-Pro show  $7/2$  symmetry and less rigidity than the sequences flanked only by Gly-Pro-Hyp of the same length. The latter sequences show a  $10/3$  symmetry [25, 26]. These model peptides reveal that triple-helical packing of collagen molecules may be complex, whereby domains which contain GPP repeats are tighter than those domains with fewer Pro and Hyp residues. To resolve the detailed structure of real collagen molecules, it will be important to understand their folding, assembly and stability. Furthermore, it must be taken into account, the  $\alpha_1$  and  $\alpha_2$  chains, contain about 1000 amino acids each, showing only 43 and 28 [Gly-Pro-Hyp] (rat collagen Type I, sp:02454 and 02466) repeats respectively. For this reason, the understanding of the collagen protein as a whole should be increased [27].

For the isolation of natural collagen there are many complications, in particular contaminations with other proteins and impurities during the purification. Also degradation effects following the commonly used acetic acid extraction or neutral salts extraction were observed [28-31]. This will be also discussed in detail later.

An important feature of fibrillar collagen such as the Type I collagen, is the formation of a large fibril or the arrangement of the molecules into fibrils. The fibrils are arranged in longitudinally staggered arrays of molecules of a length that is a non-integer multiple of the stagger between neighboring molecules. The result is a gap occurring sequentially between neighboring molecules fibrils. This gap overlap structure is defined as the D-period. This D-Period is characteristic and tissue specific, e.g. in rat tail tendon the distance is 67 nm and in human dermis is 64 nm [32].



## Introduction



**Figure 1. Structure model of collagen Type I and SEM of RTT collagen Type I fibrils.** (A) is the model from PDB data base [33]. (B) is the SEM image of Type I collagen fibrils from rat tail tendons [34].

### 1.1.3 Post-translational modifications of collagen and fibril formation

After the synthesis of collagen, it is post-translationally modified by different enzymes, including some modifications (particularly hydroxylation) which are only observed in collagen and CLPs. Hydroxylation is a process catalyzed by hydroxylases to introduce a hydroxyl ( $-OH$ ) group into prolyl or lysyl residues, thereby oxidizing them to 4-hydroxyproline (4-Hyp), 3-hydroxyproline (3-Hyp) and hydroxylysine (Hyl). The function of 4-Hyp is generally accepted to stabilize the collagen triple helix structure while the function of 3-Hyp is still not clearly known [35]. Hydroxyproline can be determined by chemical methods, as described by [36]. Hydroxylation of proline is the characteristic property of the collagen and CLPs and has an important role in the stabilization of the triple helix structure, although the significance of its role remains still debated [3, 37]. Many studies suggested that the 4-Hyps enhance the stability of collagen strongly through stereoelectronic effects. In contrast to 4-Hyp the 3-Hyp was suggested to exhibit a destabilization function [35]. Furthermore the 4-Hyp on the Y position accepts a hydrogen bond from the neighboring Gly, therefore contributing to the conformational stability of the triple helix. Reports also indicate that the Hyp is rarely found to be at the X position thus lending more stability to the structure due to the lack of steric hindrance.

Hyl has two functions: it serves as an attachment site for carbohydrate units by O-glycosylation and also stabilizes the collagen fibril by intermolecular cross-links which form normally during the maturation process. These cross-links are limited to the N- and C-termini [38]. Cross-links are formed via the lysine- and the Hyl-aldehyde pathways [38], which lead to the products of allysine and hydroxyallysine respectively. In both pathways, a single

## Introduction

copper metalloenzyme, Lysyl oxidase, catalyze the reactions in the presence of pyridoxal phosphate as cofactor. Final cross-link products of Hyl cross-linking are hydroxylysylpyridinoline (HP) and lysylpyridinoline (LP), both are non-reducible cross-links of mature collagen. HP is present in virtually all mature tissues and LP is found principally in dentin and bone [39-43]. Chemical composition of the cross-links are more tissue specific than collagen type specific, e.g. rat tail tendon contains only allysine [44]. It is also well known that hydroxylated lysines can only be oxidized to hydroxyallysine. Furthermore, the glycosylated hydroxylysine is not a substrate for lysyl oxidase. Some researchers have hypothesized that the glycosylation of Hyl is important to control the localization of aldehyde cross-links [45].

There are four loci for the cross links in collagen Type I, II and III. Two are aldehyde sites in both telopeptides and two are hydroxylysines symmetrically placed at about 90 residues from each end of the molecule. During the fibril formation of the collagen molecules, the two Hyl-residues react with the two in telopeptides from other molecules so that the characteristic 4D-periodic stagger can be formed [44].

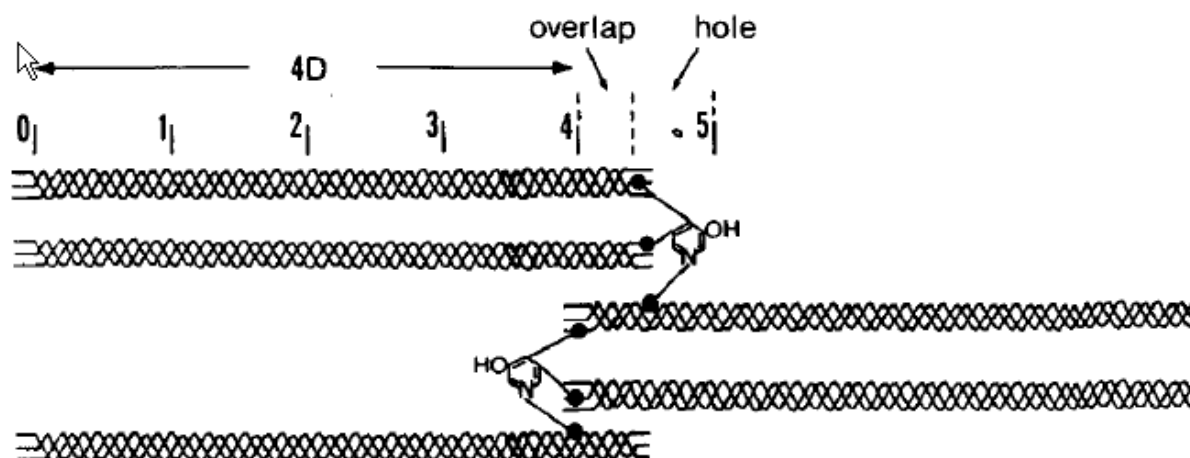


Figure 2. Cross-linking of hydroxylysine residues and formation of collagen fibril [44].

### 1.1.4 Self-assembly of collagen Type I and its aggregation

Collagen fibril formation is a self-assembly process that has been investigated in vitro for over four decades. It is generally similar to other protein self-assembly systems with an entropy-driven process whereby the surface/volume ratio is minimized through the loss of solvent molecules from surface of proteins. Although it is clearly defined that the major driving force of the fibril building is from the amino acids residing in the sequence, many studies showed other factors could effect or participate in this process [46, 47]. Not only the kinetics of fibril formation but also the diameter of the fibril is controlled by these factors. Studies have shown that proteoglycan is involved in control of the shape and size of Type I

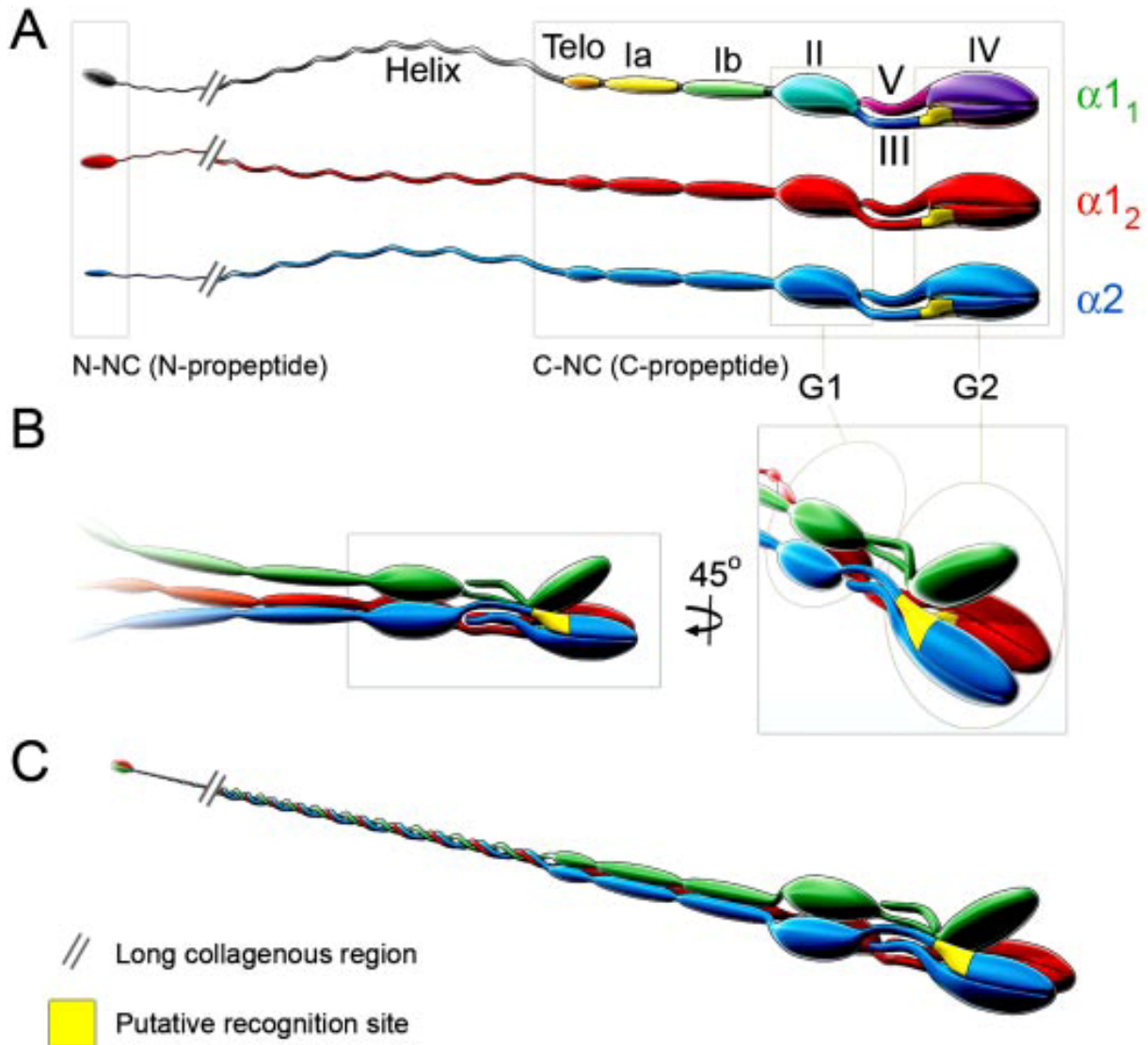
## Introduction

collagen in echinoderm. Decorin is also suggested to influence the fibril fusion [46]. Although different assembly forms were observed, the aggregation of collagen molecules into unbranched fibrils has been conserved in evolution. The fibril formation *in vivo* in vertebrates is either bipolar or unipolar, thus the spatial arrangement of the fibrils determine the mechanical and physical properties of the tissues.

Fibril-forming collagens including Type I-III, V, XI, XXIV, and XXVII are all similar in size and share the same basic sequence in that they contain a large uninterrupted triple-helical region (collagenous region). The collagenous region is flanked by N- and C-terminal non-collagenous region and both propeptides [48]. They are first synthesized in the cell as procollagen, the precursor of collagen, and then processed to collagen in the extracellular space including the cleavage of both N- and C-terminal propeptides. However, the assembly to triple helical structure begins already in the ER, with collagen single chains forming homo- or heterotrimers, even though the homotrimer of  $(\alpha_1)_3$  is the thermodynamic preferred form of Type I collagen assembly. *In vitro* studies suggest that the collagen trimerization is controlled by the non-collagenous (NC) regions [18, 49]. It has been shown that the homotrimer of  $\alpha_1$  is the default form which assembles in absence of the  $\alpha_2$  chain, whereas the homotrimer of  $\alpha_2$  is only formed if the C-terminal non-collagenous domain is replaced with synthetic peptides. Thus the C-NC-domain of  $\alpha_2$  is recognized as a key domain for the heterotrimer formation (Fig. 3). Early studies have shown that the disulfide bridges in the C-NC domain of the  $\alpha_2$  are crucial for heterotrimer assembly and folding [50]. Energy minimization and molecular dynamic simulations indicated that the trimerization of both  $\alpha_1$  C-NC domains with the  $\alpha_2$  C-NC domains is key step in folding, docking and assembly of the three heterotrimer chains [22]. Folding of collagen is initiated from the C-termini by association of  $\alpha_2$  and  $(\alpha_1)_2$  and then proceeds sequentially toward the N-termini. Trimerization begins after the interchain disulfide bonds have been formed between  $\alpha_2$  and  $(\alpha_1)_2$  C-NC domains.

At the same time, new recognition sites, encoded by the sequence information, allow formation of oligomers and fibril according to the tissue specific requirements [48]. These recognition sites determine the enzymatic cleavage sites, connections of collagen molecules, lateral association and interaction with other macromolecules from cells such as integrin and proteoglycans in the ECM.

## Introduction



**Figure 3. Protomer assembly of collagen I.** Assembly is initiated by specific interactions of the C-NC domains of the  $\alpha_1$  and  $\alpha_2$  chains, after which folding of the triple helical domain proceeds toward the amino terminus forming a heterotrimeric protomer. Subdomains dock through interactions involving recognition sites [48].

Although the understanding of the mechanisms for the trimerization and self-assembly has been increased significantly in last decades, several features are still open for future studies, e.g. the recognition motif for trimerization, differential regulation of different C-NC domains. These could be very useful for engineering proteins with a superstructure which is more suitable for application as biomaterials. Furthermore, the understanding of N-NC domain is still limited [51]. What is known is that, the N-NC peptides in procollagen  $\alpha_1$  are involved in numerous processes, including the maturation of collagen and function of Type I collagen. However, the deletion of these peptides has no significant phenotype in the mouse. Interestingly, these N-NC domains are cysteine-rich peptides (10 Cys in N-NC of  $\alpha_1$ ) and are very well conserved in almost all known procollagen types and all animals. No other amino

## Introduction

acids in this N-NC domain are seen as conserved so strongly as cysteine. The function of the disulfide bonds in this region is still an open question [51].

Whether the purified collagen can refold to triple-helical structure after the denaturation or unfolding is still unclear. Classical results in the last 40 years have indicated that unfolded collagen can not assemble to form native triple helix. In these studies, although UV-CD-spectroscopy showed a “significant” positive peak around 225 nm, which is characteristic for triple helix, the electron microscopic images revealed that only misaligned helices were present. Some studies have also shown so-called segment-long-spacing aggregations. This conclusion is supported by the above theory that the C-NC domain is a key regulator for triple helix folding. Surprisingly, all studies so far have indicated that collagen at body temperature (37°C) is instable [18]. Putatively, proteoglycans, glycosylation, heat shock proteins (HSP) and Hyps stabilize the collagen structure.

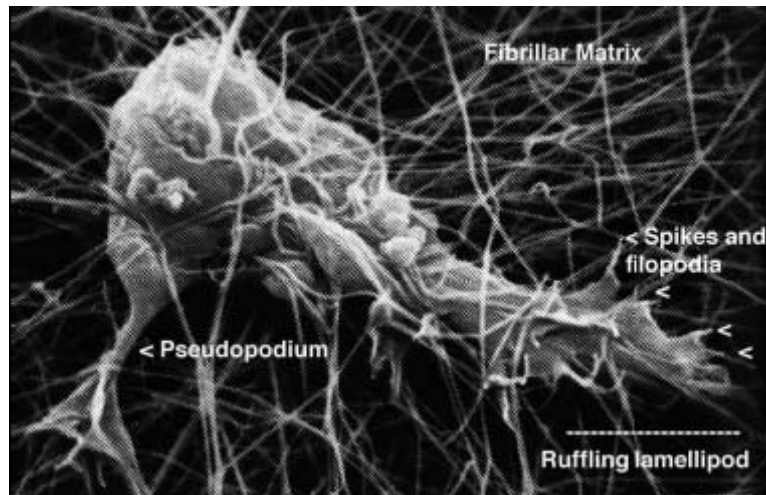
Recent studies have introduced some new results to understand the folding of the collagen which are based upon model peptides. Interestingly, these model peptides are from the collagenous region i.e. consisting only of [GPP] repeats. These researchers observed that the Hyp is not essential in stabilization of the collagen model *in vitro* [23], what is contradictory to the other studies. The question how collagen molecules are stabilized *in vivo* is still unresolved.

### 1.1.5 Cell-ECM interaction and role of collagen

Cell-extracellular matrix contact structures are spatially restricted sites on the cell surface at which specific adhesion receptors bind extracellular matrix (ECM) and link intracellularly to cytoskeletal components. Extracellular matrix supports adhesion of cells and transmits signals through cell-surface adhesion receptors. The ECM contains collagens, glycoproteins and proteoglycans. In tumor-tissue, ECM has several specific components such as tenascin, fibronectin and variant isoforms of laminin which may stimulate cancer progression [52].

Cell-matrix contacts serve several major purposes: to anchor elements of the ECM at the cell surface; to form the continuous physical linkage between ECM and elements of the cytoskeleton that is needed for cell adhesion and locomotion; to act as localized sites for transmission of mechanical force and elastic recoil between cells and extracellular matrix, and to act as sites for localized activity of signaling molecules. They are thus key components in the integration and organization of cellular and acellular elements within tissues. Currently, characterizations of cell-ECM contacts are built on morphological and biochemical criteria [53].

## Introduction



**Figure 4. SEM image of a motile fibroblast in three-dimensional matrix.** This image illustrates the complex relationships between transient matrix contact structures formed within a fibrillar matrix. Bar, 2.5  $\mu\text{m}$ . [53].

Cells may be completely surrounded by ECM, as is the case for chondrocytes, or may contact the ECM only at one surface, as exemplified by epithelial and endothelial cells. In some tissues only a proportion of the cells are exposed to ECM: for example, in stratified epithelia. The ECM offers structural support for cells, and can also act as a physical barrier or selective filter to soluble molecules [54].

It has become increasingly clear that cell-matrix interactions critically determine biological functions in all tissues, and that consequently matrix gene mutations also can underlie functional failures of internal organs, the nervous system, and blood vessels [13].

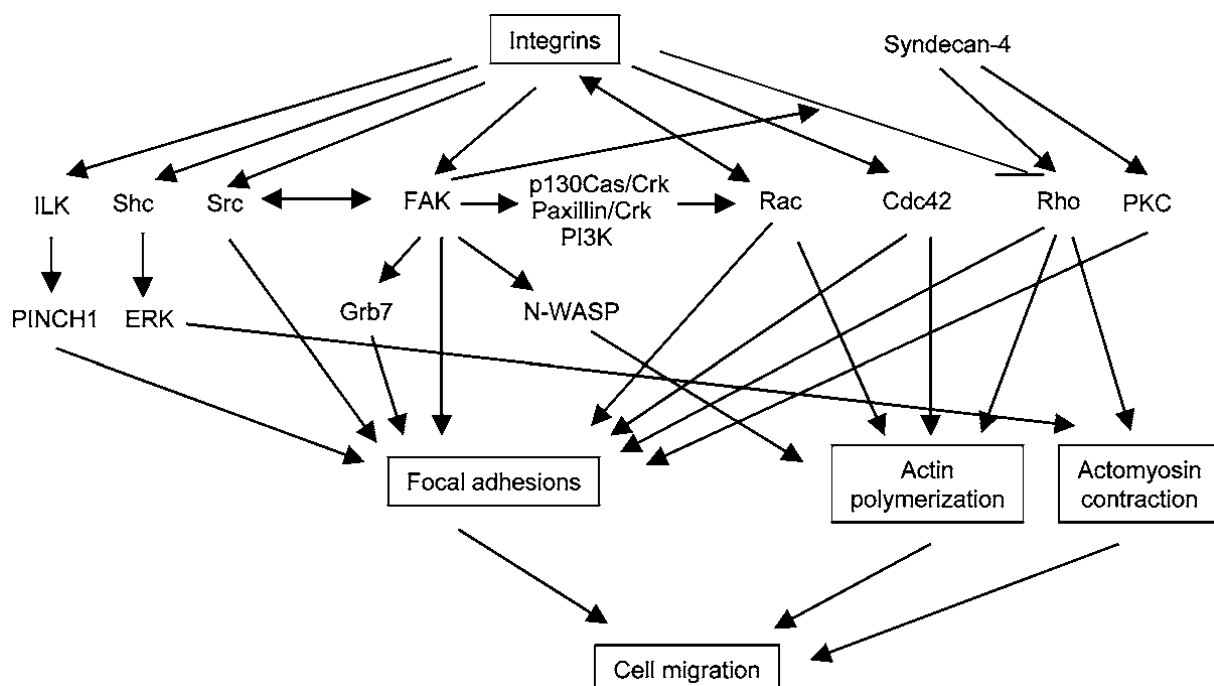
Besides their structural function, collagens interact with numerous molecules and are crucial for development and homeostasis of connective tissue. Much studied are the binding sites for numerous integrins, which are ubiquitous cellular receptors for extracellular matrix proteins. Binding of collagen to the integrins and other cellular receptors mediates cell adhesion thus regulates the cell motility. Moreover, the interaction with proteoglycans and other components in ECM can regulate the mechanical properties of different tissues [21, 55]. Cell behavior can be affected by the changes of the ECM, in particular the fibrous collagen. Changes in the compression and alignments of the fibrils of collagen can lead to cell signals showing positive or negative feedbacks. These signal-feedback cycles can give rise to the formation of a regular geometric system in the tissue so that the tissue gains its characteristic structure [56].

Self-assembly of procollagen and its processing occurs extracellularly and ECM/collagen are continuously disassembled, degraded and newly synthesized in the tissues. This is a dynamic system where the cells and ECM, particularly collagens found in tendons communicate with

## Introduction

each other. Collagen degradation and synthesis are catalyzed by a family of zinc-dependent endopeptidases known as MMPs, these include more than 21 known members [52]. Matrix remodeling, particularly collagen remodeling influences many cellular behaviors such as motion and proliferation.

It is known cell-ECM contact is important for the cell adhesion, motility, matrix remodeling and proliferation. Several pathways are found to control these cellular behaviors especially adhesion and cell motion. Different signals from the ECM can be sensed by a cellular receptor which then initiates the cell migratory response. Several features are involved in cell migration: cell protrusion, cell adhesion and migration [57]. The key step of this response is the formation of new focal adhesions points (FAs) [57]. FAs are formed to anchor the actin filaments and the cell to the substrate. Subsequently, the actin cytoskeleton and FAs exert traction force on the substrate, which results in a counter-force on the cell to promote cell migration. The contraction of actomyosin filaments pulls the cell body toward the leading edge by releasing of FAs at the rear to allow the tail to retract and the cell to move forward. ECM induces the cell signaling of cell motion through integrin activation. Integrin can modulate the Rho GTPase family and FAs such as FAK, Src, integrin-linked kinase (ILK), and Shc hence regulate the migration of cells. Cell adhesion may regulate the local actin organization through Rho GTPases in addition to providing anchors for actin filaments. FAK plays very important role in the regulation of cell motion/adhesion whereby it can regulate several different pathways. The more active the FAK the faster the cell can migrate. Rho regulates generation the force on ECM as well as Rac. It has been shown that in rapidly migrating cells the expression of FAK is strongly up regulated [57].



## Introduction

**Figure 5. Schematic diagram showing the cell-ECM adhesion-mediated signaling in cell migration [57].**

Inhibition of FAK proteins leads to a significant decrease of cell motility of fibroblasts and endothelial cells. For cell motion, stable attachments with the ECM have to be formed at the leading edge of a migrating cell, accompanied by disassembly of focal contacts at the rear edge of the cell, to allow the cell to move forward. Upon cell adhesion, FAK is co-activated with FA-formation. In order to regulate the migration, FAK can associate with numerous factors to regulate the downstream signaling pathways, including Src, p130Cas, paxillin, ERK, and MLCK, which are critical for cell migration.

In motility regulation, integrin can also crosstalk to other molecules such as growth factors to modulate the motility, including the vascular endothelial growth factor (VEGF) receptor and PDGF receptor, whereby FAK plays again the role as a linker between growth factors and integrin. After the adhesion, the next step is the traction. Traction force is transmitted through the FAs and also modulates the assembly of FAs.

Apart from cellular internal factors, external factors can also affect cell motility. Soluble molecules can lead to chemotaxis. ECM regulates cell motility by its chemical and physical properties. Moreover, the ECM molecules can even determine the speed of cell motion and its direction. One variant is the so-called haptotaxis, whereby the cells move according to the surface density of ECM molecules such as collagen. It has been shown that the inhibition of Rac and Cdc42 decreases microvascular endothelial cells migration towards collagen.

A second variant is the effect of topography and rigidity of the ECM. Cells can sense the topography of the ECM (e.g., aligned fibril) with dimensions from a few nanometers to hundreds of micrometers. The migration is then regulated by topographic guidance. This is meaningful for tissue engineering and cell culture techniques in order to control organization and structure of engineered tissue could be simply controlled. The ECM rigidity can regulate the direction of migration and studies have shown that fibroblasts and epithelial cells on flexible substrates have reduced spreading, lesser phosphotyrosine at FAs, and greater motility in comparison with rigid substrates. FAK is also involved in sensing ECM rigidity. A rigid matrix allows the cells to better interact with ECM molecules, whereas in a flexible matrix the cells tend to move towards the rigid matrix.

The third property of the ECM is to induce a mechanical force which causes the cells to move away from the ECM region. This property is also regulated by FAK. Generally, integrin-mediated signaling, cell membrane tension, and cytoskeleton tension regulates actin



## Introduction

polymerization and cell motility. Thus the adhesion receptor plays a central role in the response to mechanical force.

Migration of most adherent cells in the three dimensional matrix requires proteolysis of ECM components such as collagen, which is catalyzed by MMPs. It was shown that the expression of collagens in cultivated human fibroblast in 3D-matrix is down regulated whereas the MMPs are up regulated. Adherent fibroblasts in a 3D-matrix have a characteristic spindle or elongated form [57].

Mutations in Type I collagen result in fragility of bone and other tissues, while mutations in Type II collagen result in cartilage disorders. The phenotype of each mutation depends on its effect on the structural integrity of the protein and the extent to which the abnormal collagen chains are incorporated into the extracellular matrix [54, 58]. Viral integration within the  $\alpha 1$  Type I collagen chain gene was found to block its transcription and cause perinatal death of homozygous embryos [59] but surprisingly, the gene knock-out mouse strains are phenotypically normal [60].

### 1.1.6 Collagen as 3D-biomaterial/biomatrix for tissue engineering

Despite enormous advances in the fields of materials science and cell biology, major challenges remain for engineering materials that control and regulate cellular behavior and generation of tissues that can substitute for at least some functions of human organs. Because cells isolated from organs can not spontaneously reassemble into functional tissues by themselves, in most approaches of tissue engineering, synthetic materials are used to help the cells to get properly organized.

The first generation of cellular scaffolds has been used successfully as a substitute for the skin. For these skin equivalents, there were at least two different types: in one skin equivalent that recently obtained FDA approval (Dermagraft™), dermal fibroblasts are suspended in a polymer mesh [61]; In another product (Apligraf™), fibroblasts are seeded in a collagen gel that is then coated with a layer of human epidermal cells [62-64].

To date, the major challenge for tissue engineering is still the substitution of more complex organs, whereby the organization and functions of tissues should be completely substituted. *In vivo*, numerous cell types co-exist in an organ and assemble to a specific architecture, which gives the characteristic form and functions of the organs. Recent studies in cell biology, nanotechnology, and computation have given more new insights especially the understanding that physical cues in addition to chemical cues can regulate cell signaling and gene expression. These insights have revealed that previous approaches to engineer cell-surface interactions are

## Introduction

too simplistic and neglect critical feedback mechanisms by which cells sense and respond to their environments [65].

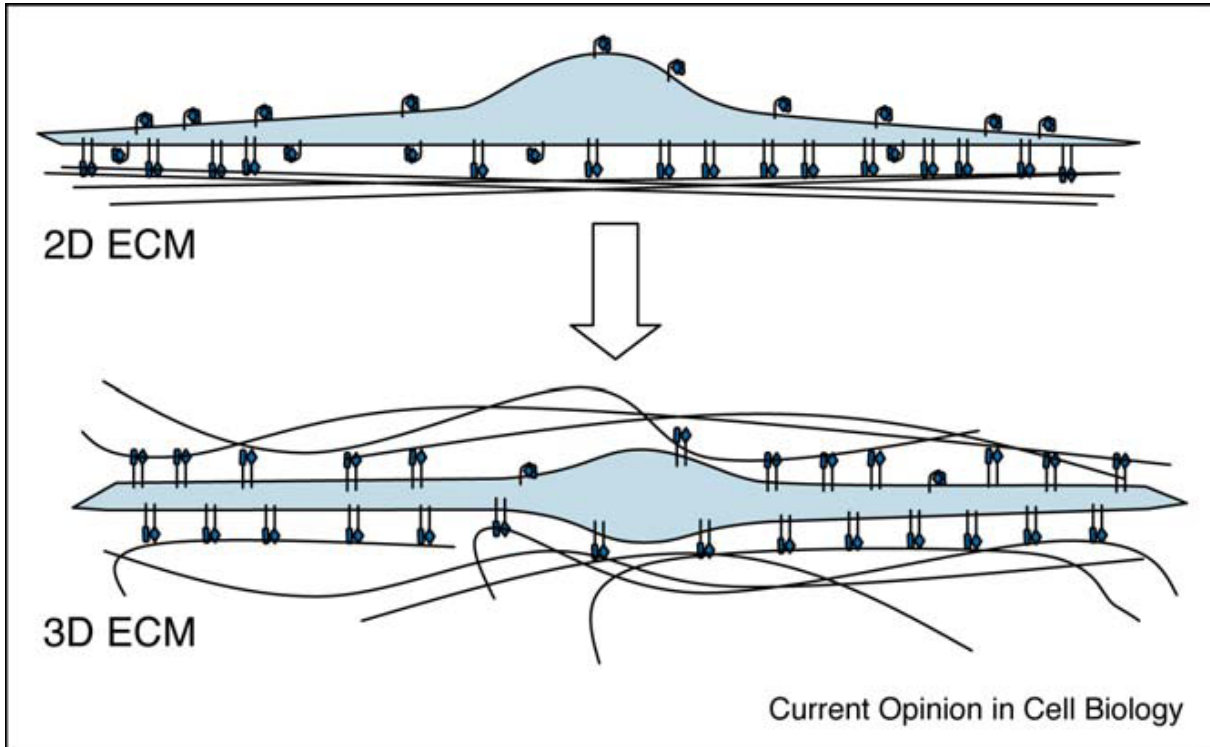
Cells show clearly different behavior whether cultured in 2-D or 3-D matrix. Other factors to which cells respond include the nano- and microscale surface topography [66-68], spatial organization of cell recognition sites [69, 70], and nutrient flow [71-73]. Because biological tissues and organs consist of living cells organized within the complex structural and functional framework of 3-D extracellular matrices, it is not surprising that significant differences have been found when cells are cultured in 3-D matrices rather than on planar 2-D surfaces. In 3-D matrices, cells assume different shapes, and functional activities such as cell migration and proliferation are enhanced [74-76].

Collagen-derived materials are among the most popular bio-derived 3-D tissue scaffolds. Its biocompatibility, biodegradability and low antigenicity makes collagen an ideal material for tissue engineering [77]. However, many studies showed that collagen has very weak mechanical properties so that during cultivation with incorporated cells the collagen is strongly remodeled and contracts with time resulting in a reduction of volume by 10% within 14 days [78, 79]. To get rid of this undesirable effect many researchers have tried to modify collagen by chemical, or enzymatic cross-linking, co-polymerization with other natural or synthetic polymers to form hydrogels with increased rheological properties [77]. Except for their mechanical weakness, studies show that the collagen matrices which are commonly made from collagen Type I and II can only bind  $\alpha_1\beta_1$ ,  $\alpha_2\beta_1$ ,  $\alpha_{10}\beta_1$ , and  $\alpha_{11}\beta_1$  integrins [80] but not  $\alpha_5\beta_1$ ,  $\alpha_v\beta_8$ . Both  $\alpha_5\beta_1$ ,  $\alpha_v\beta_8$  integrins are major factors in cell adhesion and motility. Thus many researchers have questioned the usage of collagen as a bio-scaffold. Studies comparing collagen- and fibronectin-containing matrices (2D matrices) have indicated that adhesion was reduced to one sixth of that on 3D extracellular matrices, proliferation was reduced by two thirds, and cells were significantly delayed in developing an *in vivo* spindle-shaped morphology [74]. Although collagen possesses an RGD-motive integrin  $\alpha_5$  cannot bind. One thing should be mentioned here that the studies have not yet made the connection between the collagen native structure and the integrin binding sites.

Despite of these open questions, fibroblasts cultures in 3D collagen gels have been beneficial for studying how cells interact with the chemical and physical signals of the surrounding matrix. When fibroblasts are cultured in a floating collagen gel they become quiescent and take on a dendritic shape similar to that seen in resting connective tissue. Interestingly, fibroblasts in fixed gels continue proliferating and do not form organized structures. However, it is unknown why fibroblasts behave differently when cultured on a 2D substrate compared

## Introduction

to a 3D floating or a 3D restrained collagen gel [81]. Fibroblasts cultured in a 3D cell derived matrix take on a spindle-shaped morphology, similar to *in vivo* fibroblast morphology. These cells also increase proliferation, migration, and adhesion compared to fibroblasts cultured on a 2D flattened cell-derived matrix or on fibronectin alone [82].



**Figure 6. Different cell morphology whether cultivated on 2D- or 3D matrices.** On a 2D substrate *in vitro*, fibroblast cells are polarized such that only the ventral surface contacts the substrate. In this state, cells are typically well-spread and only a subset of integrins is activated: bent conformation, inactive integrin; extended conformation, activated integrin. Upon plating within a 3D matrix or following treatment with exogenous stimulators, integrins undergo a conformational change and are activated, leading to enhanced integrin–ECM interactions and increased ECM production. As the matrix assumes of a more 3D character, cell morphology becomes more bipolar or stellate, at least partly as a result of increased engagement of receptors on the dorsal cell surface [83].

In 2-D migration models depending on adhesiveness, an almost complete lack of biophysical resistance may result in greatly varying cell morphologies, which for fibroblasts ranges from bi-polar to multi-polar. After transfer from a culture dish into 3-D matrices, a previously spread-out morphology rapidly changes towards a spindle-like elongated shape as was shown for embryonic fibroblasts, endothelial and epithelial cells, and many tumor cell lines [84]. The transition from initially round (non-motile) to polarized and spindle-shaped implicates the development of a tensile force between the cell edges. These changes in shape require cytoskeletal action and are accompanied by the redistribution of cell surface receptors including integrins, signal induction, cell spreading and/or contraction, and the induction of

## Introduction

gene expression manifested by both degradation and synthesis of extracellular macromolecules [84-88]. In 3-D collagen lattices, migrating fibroblasts and tumor cells frequently develop a length axis of up to 100  $\mu\text{m}$  and the mean maximal diameter of the cell body may surpass 20  $\mu\text{m}$  [89]. The leading edge consisting of one or several pseudopods which protrudes for the creation of new attachments followed by forward traction of the cell body, putatively provided by myosin motors. In 3-D collagen lattices, morphologically large cells, as determined by the longitudinal axis after polarization and by confocal cross-sectional area of the cell body, migrate at low speed [84].

### 1.1.7 Isolation and purification of collagens

To isolate collagen, the first step is the processing of the respective tissues. Because of the extreme diversity of tissues and types of collagen in given tissues it is normally difficult to develop a standard method for all types of collagen from different tissues. Type I collagen is extensively studied as the most abundant protein [90] in the animal body and the major fibrillar protein beside Type II and III collagen. It exists natively as a triple helix formed by two  $\alpha_1$  and one  $\alpha_2$  chains ( $[\alpha_1]_2\alpha_2$ ). The helices have the typical trimer repeats of Gly-X-Y for collagen. It is the major structural component in connective tissues such as tendon, skin and blood vessels. Because of its importance, Type I collagen has been extensively investigated as a potential candidate for use as a natural scaffold for tissue engineering and reconstructive medicine. In most tendons, Type I collagen typically forms fibrils with a length of 300 nm and a fibrillar diameter of up to 1000 nm [27]. The fibrils show a periodicity along the axis with an X-ray diffraction of 67 nm which is defined as D-period.

It has been known for a long time that collagen can be isolated by extraction in neutral salt or low ionic strength acidic solutions. In neutral salt solutions, e.g. 1 M NaCl, 0.05 M Tris, pH 7.5, collagen can be solubilized in solution. The efficiency of the extraction is also increased with increased salt concentration. However, in normal tissues the proportion of neutral salt-soluble collagen is very small so that the final yield is very low. Moreover, a variety of tissue proteinases may also be present and active in this solubilizing system. To minimize enzymatic cleavage, numerous proteinase inhibitors (such as EDTA and DFP) should be present during the extraction [91].

Another widely used method is solubilization with diluted organic acid e.g. 0.5 M acetic or citric acid, pH 3. It has also been described that lowering the pH to 2.5 in the presence of EDTA and PMSF effectively inhibits degradation. Clearly, this method has a higher capacity to solubilize collagen than neutral salt but is still limited to younger non cross-linked tissues.

## Introduction

Thus an acid-based approach was developed which contains a pepsin digestion as the first extraction step. However, the amount of pepsin sufficient for collagen solubilization is tissue dependent. Unfortunately, pepsin can also cause cleavage of collagen.

Nevertheless, the method using acetic acid to extract collagen from tissues is well established and the most widely used in research and in industrial production of collagen. Although this extraction was standardized more than 40 years ago it has still two major problems. First, the definition of collagen solubility is still ill-defined due to cross-linking mediated aggregation, so that the reproducibility of the collagen preparations is poor. Secondly, the collagen peptides especially the short non-helical regions of collagen are susceptible to proteolysis/hydrolysis during the isolation [30].

In addition to these two problems, the duration required to solubilize collagen from tissues is normally between 1-3 weeks, with high protein loss and partial degradation of the collagen peptides [5, 31]. For this reason the utility of the acidic-extracted collagen is limited, as the isolated material must be stored in cold acetic acid solution or dried. The maximal concentration of collagen obtainable is also limited to 10 mg/ml as estimated by wet weight determination and also by determination of amino acid composition. Unfortunately, the protein determination is also limited by the non-applicability of common methods such as Lowry or Bradford. Finally, the neutral salt and acetic acid extractions require that optimally younger animals are fed  $\beta$ -aminopropionitrile to inhibit cross-linking thus enhancing the solubility in the solution [30, 91].

To overcome these disadvantages, some researchers have tried to extract collagen using 8-10 M urea followed by centrifugations and different chromatographic steps using carboxymethyl-cellulose (CM-cellulose) or similar ion exchange materials [28, 92]. However, the attempts which focused on the isolation of  $\alpha_1$ -chain from Type I collagen or procollagen have reported extensive precipitation, irreversible denaturation and enzymatic cleavage during the isolation even in 8 M urea [92, 93]. All of these reports emphasized the quality of the collagen obtained was poor. This has resulted the infrequent use of the urea-extraction procedure.

### 1.1.8 Rat tail tendon and Type I collagen

Our source of collagen in this study is the rat tail tendon. Collagen Type I is the predominate component of the animal tendons and forms the mechanical network giving the tissue its elastic properties and is also responsible for the storage and transmission of mechanical power [94]. Normal tendons are made up from collagen Type I (>80%), proteoglycan and water. Although collagen present in tendon was initially considered to form a rigid structure, recent

## **Introduction**

studies showed that the flexibility of collagen Type I and its fibrils in tendon are correlated with fewer Pro and Hyp residues in the non-helical regions. This flexibility affects both the tendon stiffness and the mechanical properties. In mature tendon, several small molecular weight proteoglycan such as decorin as well as the high molecular weight hyaluronan are attached to the collagen fibrils. Genetic studies of animal models containing different mutations have shown that the lack of proteoglycans leads to larger diameter fibrils with a high tensile strength indicating that proteoglycan is required for the fibrillogenesis of the collagen. The proteoglycans also assist the alignment of the collagen fibrils as well as facilitating sliding during mechanical deformation [94]. In other tendons such as in the leg, mineralization occurs during maturation.

### **1.1.9 Recombinant collagen**

Collagen and its partially degraded product, gelatin, are considered to be used as matrices equally suitable for tissue engineering and drug delivery [95, 96]. Thus the production of purified collagen of high quality with regard to its natural structure and function are increasingly important for its application.

For the consideration of collagen as a biomaterial, several points must be taken into account. The properties of isolated collagen vary according to the tissue source, tissue maturation and collagen modifications, aging and cross-linkings in collagen. Gelatin derived from denatured and degraded collagen during the isolation and purification procedure, is a mixture of fragments of different length, structure integrity and aggregational state. These differences cause different gelling effects as well as changing the physical and mechanical properties of gelatin [97, 98]. These effects complicate the usages of isolated collagen and gelatin from natural sources.

In the last decades, rat tail tendons have been a preferred source of Type I collagen due to its high accessibility and homogeneity [99, 100]. However this source is limited to the laboratory scale. For industrial production, bovine and porcine skins are generally used. Although the production process is standardized, there still remain some ethic and epidemic questions concerning the application of animal tissues, like transmission of animal diseases to human (such as Creutzfeldt-Jakob disease (CJD)) as well as the difficulty in reproducing or maintaining the quality of the material [96-98, 101]. Worries about the transmission of diseases such as CJD, are particularly intense, as at least 100 deaths due to CJD following collagen transplantation have been reported [101].

## Introduction

For these reasons, it is meaningful to generate a pure and safe collagen source for medical applications. Naturally, human collagen would be the best choice for usage in human medicine. Some groups have made attempts to generate recombinant collagen in acceptable organisms as a production source for medical purposes, such as in *Escherichia coli* [102], yeast [103], silk worms [104], insects [105], mice [106] and mammalian cells [107, 108]. Nevertheless, the cost and complexity of the production of recombinant collagen is extremely high, with a milligram price of up to 3000 US \$. To generate recombinant collagen with native like structure and modifications, hydroxylases must be co-cloned into the host strain. Even then only a part of the natural PTMs can be achieved in recombinant collagens.

### 1.1.10 Applications of Type I collagen

As mentioned above collagen is a major component of all tissues and has a variety of structural functions, so that this protein is regarded to be one of the best candidates for a wide range of medical applications. There is a long history for the usage of collagen for numerous applications including drug delivery systems [109], scaffold in tissue engineering and regenerative medicine [110, 111]. In comparison to other biomaterials, naturally derived collagen shows excellent biocompatibility and safety due to its high abundance in all vertebrate animals and high biodegradability [112]. In addition, collagen also exhibits very low antigenicity [113].

In the last two decades, many researchers have shown that collagen has not only structural function but is also involved in many other cellular processes, including regulation of cell motion, cell proliferation and cell apoptosis [11]. Also in some biological research areas Type I collagen has been applied as a coating for culture dishes or as a scaffold for microbiological adherence and invasion test systems [114].

Many studies have shown that a 3D-culture matrix of collagen can act as a support for cells. Although results have indicated that the 3D-matrix can mimic the real ECM, many studies have demonstrated the shortcomings of the commonly used acetic-extracted collagen product. The main problem is that the porous structure of collagen spheres show pores about 200-300  $\mu\text{m}$  which are closer to those of a 2D-matrix rather than the size observed for a typical 3D-matrix in situ [115].

Collagen also shows an important potential in protein engineering. Increasingly, peptides and proteins in different species have been shown to contain collagen like protein (CLP) with characteristic GPP repeats. Thus, understanding of collagen structure could be a key area for lead to newly engineered proteins with specific properties such as high tensile strength and high stability.

## Introduction

### 1.2 Mass spectrometry

#### 1.2.1 Introduction to mass spectrometry (MS)

Mass spectrometry (MS) has recently become one of the most informative methods for studying proteins. MS-analysis of peptide fragments can lead to the identification and partial sequencing of a protein and identification of modifications. Such strategies are referred to as bottom-up sequencing, and are the most common techniques used in MS. An alternative defined as Top-Down sequencing, is starting to emerge where intact proteins are fragmented directly in the mass spectrometer.

All mass spectrometers contain three main parts: ion source, mass analyzer and detector. Analyte ions are produced in the ion source and most commonly used for proteomics are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). The ions are then transferred to the mass analyzer where they are separated according to their mass-to-charge ratio ( $m/z$ ). Ion sources can be combined with different mass analyzers such as time of flight (TOF), ion trap (IT) and Fourier transform ion cyclotron resonance (FTICR or FTMS). The physical entity measured by mass analyzers is the  $m/z$  value of the ions. The output, which is recorded at the detector, is the ion intensity at different  $m/z$  values. The result is visualized by an  $m/z$  vs. intensity plot or a mass spectrum [116].

#### 1.2.2 Tandem MS and protein identification

Common tandem MS-experiments begin with purifying a mixture of proteins from tissues or cells through different methods using a complex set of techniques. Finally, the proteins are cleaved into peptides by proteolysis and the mixture of peptides is then analyzed using MS. Every mass peak-separated is further fragmented and measured by tandem MS. Thus the parent proteins are identified from the tandem mass spectra [117]. The spectra are then submitted to software for searching of the matched peptides. There are three types basically different software for MS-identification: database searching, *de novo* sequencing and hybrid approaches [118]. At moment, the database searching is still the mostly common choice for the tandem MS analysis. However, most proteins are modified either enzymatically *in vivo* or chemically during the preparation. Therefore, the fragments containing modifications have different masses which can not be identified by data base search.

In MALDI, the analyte is mixed with a compound that absorbs the energy from the laser, thus the analyte is co-crystallized with an excess amount of the matrix. Several different matrix-analyte preparations have been developed. The most common method is the dried droplet method [119, 120].



## Introduction

In MS-TOF analysis, ions derived from MALDI are accelerated by an electrical potential and pass through a field free region where every ion flies due to its characteristic  $m/z$  value. The TOF is then measured at the end of the field. Peptide ions by MALDI normally are in a low-charge state and mostly only the singly charged state will be measured. In contrast to peptide ions, proteins with higher charged states can also be detected [121]. By spectra-analysis, several points should also be taken in account: enzymatic mis-cleavage, modifications and contaminations from other proteins. Especially PTMs are difficulty for the protein identification using MS. However, MALDI-TOF/TOF still plays a major role in proteomics due to its high sensitivity. Using post source decay (PSD) the sequence information is strongly improved. Furthermore the collision-induced dissociation (CID) also generates low- and high-energy fragmentation, which can provide more significant data and extend the MALDI method for more complex mixtures.

ESI-MS/MS is another method widely used in proteomics where ions are formed at atmospheric pressure is followed by droplet evaporation. The initial drops shrink during evaporation thus increasing the charge density and leading to a Coulomb explosion of big droplets into smaller ones, until the droplets are small enough to desorb analyte ions into the gas phase. Generally, the longer the peptide or protein the higher the charge state because more groups can be ionized. Therefore it is common to observe multiple charged fragments [122]. ESI-MS/MS spectra are mostly in the positive ionization mode but the negative mode is also used to detect phosphorylation [123]. So far, the ESI is generally performed at atmospheric pressure so that a coupling with high performance liquid chromatography (HPLC) or capillary electrophoresis (CE) is possible, thereby facilitating the determination of a complex mixture. The coupling of nano-LC to ESI-MS has strongly increased the sensitivity and efficiency of ESI-MS [124, 125].

In tandem MS, the first mass analyzer separates and isolates the precursor ions and then transmits ions of interest to the collision cell to be fragmented. The fragments are transmitted and analyzed in the second mass analyzer. The resulting spectrum is called a MS/MS spectrum. Numerous programs have been developed to interpret the MS/MS spectra but all have limitations. Therefore, the combination of different programs and different MS-methods is meaningful for many cases.

To describe the ions generated during the fragmentation, the general nomenclature was established more than 20 years ago. a-, b-, and c-ions are charged fragments containing the N-terminal part and x-, y-, and z the C-terminal part. Mass differences between neighboring  $b_n$  and  $b_{n+1}$ -ion peaks correspond to the residue mass of the C-terminal residue of the  $b_{n+1}$ -ion.

## Introduction

Similarly, the mass difference between  $y_n$  and  $y_{n+1}$  reveals the mass of the N-terminal residue of the  $y_{n+1}$ -ion [126].

The computational analysis of mass spectra begins with the identification of the spectra. The most efficient methods are based on protein data base searches and widely used commercial software are the SEQUEST and MASCOT. For the assignment of the peptide sequences to the MS/MS spectra there are 3 categories of approaches. First, peptide sequences are extracted directly from the spectra without referring to a database (*de novo*) [127-129]. Second, protein identification is performed by correlating the spectra with theoretical spectra predicted for each peptide contained in the database [130-138]. Third, a hybrid approach, which is based on the extraction of short sequence tags, normally 3-5 residues, followed by “error tolerant” database searching [139-142]. For protein identification, database searching is commonly used.

In this study, three programs have been applied and compared, including OMSSA, MASCOT and PEAKs. The Open Mass Spectrometry Search Algorithm (OMSSA) is an efficient search engine for identifying MS/MS peptide spectra by searching libraries of known protein sequences. OMSSA scores hits with a probability score developed using classical hypothesis testing, the same statistical method used in BLAST. It is freely available from NCBI. Similar to OMSSA, MASCOT is also based on searching in primary sequence databases but the unknown sequences are also considered for further homology searching. PEAKs is a relative new program and provides another possibility combining database searching and *de novo* sequencing. Furthermore, it can perform database searching, *de novo* sequencing, and homologue searching during one running process, thus yielding more insights for unknown proteins with limited primary sequence information in the database [118, 143].

For database searching, several parameters should be taken into account: (1). The type of fragment ions, which are specific for each type of mass spectrometer, containing both C- and N-terminal ions. Under low energy CID conditions, fragmentation results commonly in the b- and y-ions; (2). In high-resolution mass spectrometers the measured mass is more likely to be the monoisotopic mass whereas the average mass should be chosen for low-resolution instruments; (3). The peptide ion charge state is also an important parameter for reliable mass determination and database searching; (4). The parent ion mass tolerance defines the range of the measured peptide mass which are selected from the database and scored against the experimental spectrum; (5). The enzymatic digestion profiles, e.g. for tryptic digestion, the resulting peptides should have K or R at the C-terminus. Non-specific digestion due to

## Introduction

posttranslational modifications or self digestion of the enzymes can occur. Databases used for data base searching are generally NCBI and SWISSPROT.

In most studies, the identification of a protein is more interesting than identification of peptides. Therefore, identified peptides should be grouped to a corresponding protein and the score should be calculated newly [144].

### 1.2.3 Mass spectrometry of collagen

Although the collagens have been studied intensively for more than 50 years, most of the amino acid sequences are generated from the mRNA and partially Edman sequencing which was obtained in 1960's. Also for the most abundant protein Type I collagen, more than half of its sequence has been annotated from mRNA sources. The poor quality of collagen sequence available strongly limited the analysis of collagen structure and function as well as expression as recombinant protein. Additional complications are due to the sequence ambiguities arising from GPP repeats and variety of PTMs. Thus the modelings of collagen like or similar peptides can only yield partial information about collagen.

Traditionally, the analysis of collagen triple helix, dimers of  $\alpha$  chains or CNBr-cleaved collagen peptides have employed SDS-PAGE or different chromatographic procedures including ion exchange, gel filtration and capillary electromigration [100, 145]. However, as mentioned above, the motility of collagen, its dimer and oligomers vary strongly in the SDS-PAGE behavior so that the determination of the real molecular weight of the molecules is not possible using SDS-PAGE. Even after collagen peptide separation, the determination of collagen type is also difficult due to the high homology between the collagen types and the high abundance of several types of collagens in parallel. Although there are many commercial antibodies which are available to distinguish collagen types using histochemical staining or immunoblotting [146, 147], the accuracy of these determinations is limited by the cross reactivity between collagens. Above all, for these conventional analyses, the collagen must be available in large amounts and usually solubilized by limited enzymatic digestion [147], which results in decreased sensitivity and accuracy of these methods. .

Despite the fact that mass spectrometry/tandem MS have become routine techniques in proteomics and protein analysis due to its high accuracy and sensitivity, the MS-analysis of collagen has remained relatively undeveloped and rarely applied. It has been reported that collagen is difficult to be analyzed using MS [148, 149]. Major challenges for applying MS technique to collagen is still its complexity, large molecular weight, complex stable structure and a large amount of PTMs [148]. To overcome these problems several research groups have applied the IR-MS [149], isotopic labeling using  $^{13}\text{C}$  [150], C-terminal MS sequencing [151],

## **Introduction**

which is giving new insights into the analysis of collagen mutations, PTMs and structures. However, these methods are still too complicated to be used routinely in any laboratory so that these methods are not widely applied. Nevertheless, MS technology is developing quickly, especially in proteomics area making it likely that the MS-analysis of collagen will soon be easier than in past years.

Another improvement is the development of the database and MS-data analysis programs so that data collection and handling have been strongly improved. As mentioned above, some programs have improved the sensitivity and quality of the analysis, thus yielding more sequence and PTM information. This can in turn supplement and even correct the existing database derived mostly from mRNA sequence annotations.

### **1.3 Goals of this study**

Collagens, particularly the Type I collagen has been studied extensively in last 40 years. The application of Type I collagen is increasing continuously, such as in tissue engineering as natural matrix. However, these applications are increasingly questioned with regard to different non-understood features of collagen, including its assembly, folding, post-translational modifications, export and cellular function. This understanding is particularly hampered by the isolation process, which commonly uses acetic acid-extraction. Moreover, most structural studies have been carried out using model peptides leading to limited understanding of the collagen structure and assembly.

Due to the extreme importance of biocompatible matrices for tissue engineering and application in medical technology, the availability of native collagen should be studied by refining the extraction procedure of collagens.

One of the goals of this study was to reexamine the quality of Type I collagen after extraction with urea. We expected the results would also enable a commercial patent to be successfully approved by respective authorities.

The high quality of urea-extracted collagen (UC) led to a further goal of detailed characterization of the physical state of the protein. This may further provide interesting new insight in aspect of protein design and engineering of artificial collagen-like proteins.

## Materials and Methods

### 2. Materials and Methods

#### 2.1 Chemicals

1. Roth GmbH & Co. KG, Karlsruhe:

Bovines Serum Albumin (BSA), bromphenolblue, EDTA, ethanol, glycerin, glycin, HCl, isopropanol, KCl, KH<sub>2</sub>PO<sub>4</sub>, MgCl<sub>2</sub>, methanol, Na<sub>2</sub>CO<sub>3</sub>, NaCl, NaOH, sodiumacetate, sodium azid (NaN<sub>3</sub>), SDS, and Tris.

2. Molecular Probes, Eugene, Oregon:

Streptavidin-POD, Sulfo-NHS-LC-Biotin.

3. New England Biolabs:

Restriction enzymes: EcoRI, EcoRV, BamHI, PstI, HindIII, NotI.

Modification enzyme: T4-DNA-Ligase, alkaline phosphatase (AP, EC.3.1.3.1).

4. Amersham-Pharmacia Biotech Europe GmbH, Freiburg i. Breisgau:

ECL-Plus Westernblot detection Kit.

5. Pierce:

Immobilized TCEP disulfide reducing gel, EZ-link-Maleimide PEO<sub>2</sub>-Biotin.

6. Roche Diagnostics:

Carboxypeptidase Y (EC 3.4.16.1, sequencing grade), chymotrypsin (EC 3.4.21.1, sequencing grade), proteinase K (EC 3.4.21.64, PCR grade); LightCycler 480 Probes Master, Universal probe library.

7. Qiagen

RNeasy kit, QuantiTect Multiplex PCR kit, QuantiTect reverse transcription kit.

#### 2.2 Buffers, solutions and Media

##### 2.2.1 Commonly used buffers: Phosphate buffered saline (PBS), Tris HCl buffer

##### 2.2.2 Laemmli electrophoresis buffer [152]:

- Ten-fold concentrated electrophoresis buffer: Tris 25 mM, glycin 192 mM and Sodium dodecyl sulfate (SDS) 0.1% (w/v) in water. Shortly before the electrophoresis, it is diluted to one fold solution.
- Stacking gel buffer: Tris-HCl pH 6.8 0.5 M and SDS 0.4% (w/v) in water.
- Resolving gel buffer: Tris-HCl pH 8.8 1.5 M and SDS 0.4% (w/v) in water.
- Acrylamide-/bisacrylamide (N,N'-Methylenebisacrylamide) solution (ABS): 40% solution, 37.5:1 Bio-Rad Laboratories GmbH, München.
- APS: 10% Ammonium persulfate in water.

## Materials and Methods

- Protein sample buffer 4 fold: Tris-HCl pH 8.0 40 mM, EDTA 4 mM, SDS 4% (w/v), glycerol 40%, 0.02% bromophenol blue in water (deduced buffer contains 20%  $\beta$ -mercaptoethanol).

### 2.2.3 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

[152] Stacking gel: water 3.3 ml, stacking gel buffer 1.25 ml, ABS 665  $\mu$ l, APS 50  $\mu$ l and TEMED (N, N, N', N'-tetramethylethylenediamine) 5  $\mu$ l;

Resolving gel: 5%-16%.

	5%	8%	10%	12%	16%
Water	7 ml	5.8 ml	5ml	4.2 ml	2.6 ml
Buffer			3 ml		
ABS	2 ml	3.2 ml	4 ml	4.8 ml	5.6 ml
APS			100 $\mu$ l		
TEMED			14 $\mu$ l		

**Table 1. SDS-PAGE preparation.**

Gels were usually stained Coomassie Brilliant Blue R250 using standard protocol.

### 2.2.4 Western blot buffers

- Transfer buffer: Tris-HCl pH 8.0 25 mM, glycine 192 mM und 20% methanol in water, immediately mixed prior to application.
- PBST: PBS pH 7.4 und 0.5% Tween20.
- One fold Roti-block solution: 1:10 (v/v) diluted from Roti-block concentrate (Carl-Roth GmbH, Karlsruhe)
- Anti-collagen Type I rat antibody and anti-His tag antibody were dissolved 1:10000 (v/v) in one fold roti-block solution respectively. Secondary antibody was dissolved 1:2000 (v/v) in one fold Roti-block solution.
- Streptavidin-POD was dissolved 1:2000 (v/v) in one fold Roti-block solution.

## 2.3 Protein isolation from rat tail tendons (RTT)

### 2.3.1 Isolation and gel-filtration

Rat tail tendon, teased out from 3-month-old rats was washed extensively in phosphate saline buffer (PBS) pH 7.4 at 4 °C. The tendon was extracted with 9 M urea at 25 °C for 20 h, followed by a centrifugation at 4000xg for 15 min to sediment the insoluble components. The urea-soluble proteins were purified by gel-filtration using a Superose 12 column (GE-Healthcare, diameter 6 cm and height 60 cm) at a flow rate of 25 ml/h. The column was

## Materials and Methods

equilibrated with running buffer, containing 8 M urea, 50 mM TrisHCl pH 7.5, and 100 mM NaCl. Collected fractions were then dialyzed repeatedly against a 100-fold volume of distilled water at 4 °C for two days and then lyophilized. For further analysis the urea-soluble proteins were fractionated by using a SMART-system (Amersham Biosciences) through a smaller column with same media. The proteins were detected by monitoring the UV absorbance at 213 nm, 256 nm and 280 nm.

### 2.3.2 Productivity determination

The yield of collagen extraction was determined as described in following process:

- Tendon was washed twice in PBS pH 7.4, and then centrifuged by 4000xg for 3 min.
- Supernatant was dialyzed against water then lyophilized as neutral salt soluble fraction.
- Washed tendons were also lyophilized and the dry weight was set as input for the extraction using urea or acetic acid solution.
- Urea extraction after 20 h, acetic acid extraction after 7 days was centrifuged by 4000xg for 15 min. Supernatant was transferred to a new tube and laid as urea/acetic soluble fraction. Pellet was dialyzed against water and lyophilized as insoluble fraction.
- Urea/acetic soluble fractions were sufficiently dialyzed against water with repeated water changes, eventually lyophilized and weighted out as extracted collagen.
- An average collagen extraction yield was performed in triplicate using the

$$\text{calculation: } Y_{UE} = \frac{(W_{RTT} - W_{PBS} - W_I)}{W_{RTT}} \times 100\% .$$

$Y_{UE}$  is yield of urea-extraction.  $W_{RTT}$  is dry weight of rat tail tendon.  $W_{PBS}$  is the dry weight of components remaining in PBS by the wash steps.  $W_I$  is the dry weight of urea-insoluble fractions.

## 2.4 Gel electrophoresis and Western blot

### 2.4.1 SDS-PAGE analysis

For SDS-PAGE analysis, lyophilized fractions were dissolved in sample buffer, and then boiled at 100 °C for 5 min prior to gel loading. Generally, 0.75 mm thick gels were applied in

## Materials and Methods

the study. Resolving gel concentration varied from 4% to 16 % according to requirements. The gels were usually stained with Coomassie Brilliant Blue R250.

### 2.4.2 Gel staining

Coomassie staining:

Gels are first incubated in staining solution containing 10% acetic acid, 50% methanol and 0.1% Coomassie Brilliant Blue R250 over night. Then the gels were destained in solution containing 10% acetic acid and 30% methanol until the bands are significantly observable.

Silver staining:

The complete procedure was proceeded at room temperature by shaking on an orbital shaker by 50 rpm. For fixation of the protein/peptides in gel, the gels were incubated for 1 h in Ethanol/acetic acid/water (40:10:50) followed by dehydration step for 2 h in Ethanol/acetic acid/water (5:5:90). After a short incubation in water for 5 min, the gels were transferred into an aqueous solution containing 0.5 M sodium acetate and 1% glutaraldehyde and incubated for 1 h. Finally, the gels were washed three times in water (10 min/wash). In order to get a homogeneous staining, the gels were incubated twice in 0.05% naphthalene disulfonic acid solution (15 min/incubation). The excess of naphthalene disulfonic acid was removed through water washes with water changes every 15 min. Finally the gels were stained with silver nitrate/ammonium hydroxide solution. This silver solution was prepared in two steps, in which 0.32 g silver nitrate was dissolved in 3 ml water and pipetted slowly into ammonium hydroxide solution composed by mixing of 0.5 ml 25% ammonium hydroxide and 80  $\mu$ l 10 N NaOH to 8.5 ml water. This solution was then added to 40 ml and used for staining. The stained gels were washed in water with several changes for every 5 min to remove the excess of silver nitrate. Then the gels were shortly incubated in developer solution containing 0.1% formaldehyde and 0.01% sodium citrate. Since the signal was sufficient, the staining was stopped by transferring of the gels into aqueous solution with 5% Tris base and 2% acetic acid.

### 2.4.3 Western blot

For specific detection of collagen, His-tagged collagen peptides and biotinylated protein samples the SDS-PAGE gels were transferred onto a polyvinylidene difluoride (PVDF) membrane (Immobilin P, Millipore) in a semi-dry matter. The membrane was pre-equilibrated in transfer buffer. After the transfer, the membrane was blocked in Roti-block solution for 1h then incubated with anti-collagen, anti-His (1:10000 v/v diluted in Roti-block solution) and Streptavidin-POD (1:2000 v/v diluted in Roti-block solution) respectively for 1 h. Streptavidin-POD blots were directly reacted with the ECL-Plus Western blot detection Kit and



## Materials and Methods

the chemiluminescence signal was documented using a CCD-camera (LAS 1000, Fujifilm). In other cases,  $\alpha$ -goat-anti-rat/mouse-POD was used as secondary antibody (1:10000 v/v in Roti-block) and incubated for 30 min.

### 2.5 Chemical modifications of the collagen

#### 2.5.1 Biotinylation

For detection of cysteine residues and disulfide bridges, the collagen samples were reduced using immobilized TCEP (Tris[2-carboxyethyl] phosphine hydrochloride) disulfide reducing gel (Pierce). Collagen was first dissolved in PBS at pH 6.9 and a concentration of 1  $\mu\text{g}/\mu\text{l}$ . 100  $\mu\text{l}$  gel slurry was centrifuged at 1000xg for 30 sec and the gel was mixed with 100  $\mu\text{g}$  collagen sample and incubated at pH 6.9 and 37 °C for 30 min. After centrifugation at 10000 rpm for 1 min the reduced supernatant was recovered and biotinylated using EZ-link-Maleimide PEO<sub>2</sub>-Biotin (ME-Biotin, Pierce). The concentration of ME-Biotin stock was 10.5  $\mu\text{g}/\mu\text{l}$  (suggested by the manufacture). The biotinylation was carried out at room temperature for 1 min to 20 h. The biotinylation was stopped by dialysis using Slide-A-Lyzer Mini-Dialysis Unit (Membrane: 3500 MWCO, Pierce) with repeated water changes at 4 °C.

#### 2.5.2 Purification and concentration of biotinylated protein

In order to separate and concentrate the biotinylated collagen, an affinity chromatography was performed in following steps. 2 mg CNBr-activated Sepharose 4B (Amersham Pharmacia) was equilibrated using 1 mM HCl whereby the beads was washed with more than 500 ml 1 mM HCl. 5 mg Neutravidin was dissolved in binding buffer containing 0.1 M NaHCO<sub>3</sub> and 0.5 M NaCl. 0.3 g. Neutravidin was applied to the column with washed Sepharose beads for 2 h incubation. Excess of Neutravidin was washed out by repeated loading of the column with binding buffer. Then the column was washed alternatively for at least 3 times using sodium buffer containing 0.1 M sodium acetate and 0.5 M NaCl pH 4.0 and 0.1 M Tris buffer pH8.0 with 0.5 M NaCl.

The biotinylated collagen was applied to this column and incubated at 4 °C for 20 h. Excess protein or non-biotinylated proteins were removed by washing using PBS pH 7.4 containing 0.1 % SDS. The specifically bound proteins were then eluted using boiled 1% SDS solution by incubation at 100 °C for 5 min.

#### 2.5.3 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) determination of free thiols in collagen

To estimate the free thiols as well as disulfide bridges in the urea-extracted collagen, 25  $\mu\text{l}$  (correspond to 0.125  $\mu\text{g}$  weighed collagen) of a collagen sample was diluted to 200  $\mu\text{l}$  with 10

## Materials and Methods

mM sodium phosphate buffer pH 7.5, and 200  $\mu$ l 0.1 M DTNB was added to start the reaction. After incubation time (5 min) the absorption ( $A_{412}$ ) was measured at 412 nm (25 °C) [153, 154]. A calibration curve was constructed using known concentrations of cysteine, dissolved in the same buffer, as standard.

### 2.6 Cleavage of protein

#### 2.6.1 Cyanogen bromide cleavage

Lyophilized urea-extracted collagen (UC) from gel-filtration was dissolved in formic acid containing 30% ethanol (v/v) at a collagen concentration of 3 mM as determined by  $\Delta A_{220}$ . For the cleavage of methionine, cyanogen bromide (CNBr) 75-fold in excess of the His amount which can be estimated from the  $\Delta A_{220}$  was added to the solution in a Reacti-Vial (Pierce) and the reaction was performed for 6 h in the dark at room temperature under stirring. The reaction mix was centrifuged at 16,000xg for 10 min and the supernatant was transferred to an Eppendorf tube for drying in a Speed-Vac (Univapo-100H, Uniequip GmbH, Germany) for 2 h.

#### 2.6.2 Trypsin digestion

For chromatographic analysis of tryptic-digested collagen, the collagen was first dissolved in 50 mM  $\text{NH}_4\text{HCO}_3$  (pH 7.5) solution and reduced by adding 100 mM dithiothreitol (DTT) for 15 min by 55 °C. After cooling to room temperature, the reduced protein was alkylated by incubation with iodacetamide (IAA) at room temperature in dark for 15 min. Both of DTT and IAA were dissolved in 50 mM  $\text{NH}_4\text{HCO}_3$  (pH 7.5) solution. To stop the reaction and remove the excess of reagents, the samples were dialyzed against water using Slide-A-Lyzer Mini-Dialysis Unit. Eventually, the samples were concentrated by lyophilisation and redissolved in 25 mM  $\text{NH}_4\text{HCO}_3$  (pH 7.5) solution and mixed with trypsin (Promega), E/C ratio is 1/100. The digestion was carried out at 37 °C for over night.

For tandem MS analysis of the samples analyzed using SDS-PAGE gels, an in-gel digestion protocol was applied to this purpose [155]. Bands of interest were cut thoroughly from the washed gel and washed in 50 mM  $\text{NH}_4\text{HCO}_3$ /acetonitrile 1+1 (v/v) for 15 min. After removal of all the liquid, the gel was covered with 100% acetonitrile (ACN) until the gel shrunk together and looks white. Then the gel was rehydrated in 50 mM  $\text{NH}_4\text{HCO}_3$  for 5 min and given equal volume of ACN for another 5 min. Finally, the liquid was removed again and gels were incubated with 100% ACN then dried in the Speed-Vac. The gel was reduced and alkylated in the same matter as described above but the duration of every steps are doubled to

## Materials and Methods

ensure the diffusion into the gel. After the removal of IAA the gel was transferred to 50 mM  $\text{NH}_4\text{HCO}_3$ /acetonitrile 1+1 (v/v) for 15 min with at least two changes. The solution was again removed and the gel was transferred into 100% ACN. After the gel shrunk, it was dried using the Speed-Vac.

Freshly prepared trypsin solution in 25 mM  $\text{NH}_4\text{HCO}_3$  was applied to the dried gel in a volume sufficient just to cover the gel. The gel was incubated for 30 min so that the enzyme could permeate into the gel. Then enough 25 mM  $\text{NH}_4\text{HCO}_3$  solution was given to keep the gel wet. The digestion was taken over-night at 37 °C.

### 2.6.3 Chymotrypsin and chymotrypsin/trypsin digestion

The chymotrypsin (ChyTrp) digestion was carried out with UC samples with a E/C ratio for 1/50 at 20 °C for 1 min to 20 h. The digestion was stopped by the addition of PMSF or a 10% boiled SDS-solution.

The digestion with a mixture of chymotrypsin and trypsin (Trp) was carried out at 37 °C and the given ChyTrp/Trp ratio is 10/1. ChyTrp/collagen ratio is kept by 1/50.

### 2.6.4 Carboxypeptidase Y digestion

Dialyzed UC and AC samples were digested with yeast carboxypeptidase Y for 1 min to 20h and the enzyme/collagen ratio varied from 1/200 to 1/25 (w/w). The digestions were carried out at 37 °C and stopped by addition of PMSF.

## 2.7 Mass determination

To determine the mass of the collagen protein, a thin-layer preparation with a sinapic acid matrix was used in MALDI-TOF MS. 1  $\mu\text{l}$  sinapic acid (20 mg/ml in acetone) was spotted on a stainless steel MALDI target (Bruker Daltonik GmbH, Bremen, Germany) and dried. 1  $\mu\text{l}$  aqueous protein solution (2.5  $\mu\text{g}/\mu\text{l}$ ) was deposited on the matrix layer and also dried. Afterwards, the sample was covered with 1  $\mu\text{l}$  sinapic acid solution (20 mg/ml) dissolved in 0.1% aqueous trifluoroacetic acid (TFA) containing 40% acetonitrile and incubated to dryness. The average mass of the proteins was analyzed by MALDI-TOF MS using an Ultraflex II TOF/TOF 200 (Bruker). The mass spectrometer was set to scan over a mass range of 10-140 kDa. The mass spectra were acquired in linear mode and processed using FlexControl and FlexAnalysis Software (Bruker).  $\beta$ -galactosidase monomer (116,483 Da; Calbiochem) and glycogen phosphorylase (97,289 Da; Sigma-Aldrich) were used for external calibration.

## Materials and Methods

### 2.8 Tandem Mass spectrometric (MS/MS) analysis

#### 2.8.1 Electrospray ionization MS/MS

About 0.5  $\mu\text{g}/\mu\text{l}$  of protein was dissolved in a solution containing 8 M urea and 50 mM  $\text{NH}_4\text{HCO}_3$ , then reduced and alkylated using standard protocols. The sample was diluted and tryptic digestion was carried out at 37 °C over night. After digestion the solution was adjusted to a pH of 2-3 with concentrated formic acid.

Peptide mixtures were separated for nano-LC-ESI-MS/MS using a FAMOS auto-sampler (Dionex), an Ultimate inert HPLC (Dionex) and an Agilent HPLC 1100 pump connected to the nano-ESI-Source of a Finnigan LTQ-FT (Thermo Electron Corporation) for online mass detection. Peptides were first collected on a trap column (dimension 0.1 x 15 mm, self-made using Zorbax Eclipse XDB-C18, 5 $\mu\text{m}$ , Agilent Technology) for desalting and concentrating followed by separation on an analytical column which was self-made with loader kits SP035 from Proxeon Biosystems using Hydrosphere C18, 3  $\mu\text{m}$  (YMC) and fused silica emitters (0.075 x 105 mm, 6  $\mu\text{m}$ , Proxeon Biosystems). Peptides were eluted during a 60 min gradient using 97% water, 3% acetonitrile and 0.1% formic acid as solvent A and 80% acetonitrile, 20% water and 0.1% formic acid as solvent B at a flow rate of 0.15  $\mu\text{l}/\text{min}$ . Mass spectrometric detection consisted of FT full scans at a resolution of 25.000 followed by data dependent FT SIM scans at a resolution of 50.000 and low resolution IT MS/MS scans using a dynamic exclusion of parent ion masses for 60 sec.

The MS and MS/MS spectra were submitted to an in-house installation of the open mass spectrometry search algorithm (23) (OMSSA, version 2.1) and searched against NCBI non-redundant (nr) database (release Feb 8, 2007) and its subsets. Search results were filtered and sorted using in-house written software (F. Drepper, unpublished data). Peptide hits were considered significant if the precursor and product ion masses matched within 3 ppm and 0.5 rel. mass units, respectively, and if the E-value was below 0.01.

#### 2.8.2 MALDI-TOF/TOF

Digested peptides were purified using ZipTipC<sub>18</sub><sup>™</sup> (Millipore Corporation, Billerica, MA, USA) and concentrated to 10  $\mu\text{l}$  followed the instruction of manufacture. These were directly spotted onto the Prespotted AnchorChip-Target (Bruker Daltonik GmbH, Bremen, Germany). The monoisotopic mass of the peptides were determined using ultraflex II TOF/TOF 200 (Bruker) and the mass between 700 and 4000 Da were picked out in reflector modus using FlexControl and FlexAnalysis Software (Bruker). Furthermore, the picked peaks were fragmented in PSD matter and significant peaks were analyzed. The analysis was performed

## Materials and Methods

using MASCOT (Matrix Science, London, Great Britain) as data base searching program against nr database. The tolerances of the MASCOT search were set to 1.0 Da and 0.5 Da for peptide mass and MS/MS respectively. The match was only accepted if the p-value was smaller than 0.05 which indicated the probability of a random assignment.

### 2.9 Spectroscopic analysis

#### 2.9.1 UV-spectroscopy

UV-spectra were obtained using a JASCO V-560 spectrometer using a scanning speed of 200 nm/min and slit width of 2 nm. Generally, 2 mm path-length quartz cuvettes were used for the measurements.

#### 2.9.2 UV-CD-spectroscopy

UV-CD spectra were acquired with a JASCO J-715 CD spectrometer (scanning speed of 50 nm/min, 2 mm cuvette, slit width 2nm, averaging over 9 scans). Spectra were obtained in the range of 320 nm to 190 nm.

For the spectroscopic measurements, all buffers were degassed under vacuum prior to use.

#### 2.9.3 Gaussian analysis of UV-VIS spectra

Gaussian deconvolution of UV-VIS spectra was performed using an in-house program using the following algorithm: (1) the baseline absorbance due to turbidity at the red end (300 nm) of the spectrum was subtracted from the complete wavelength region; (2) a wavelength-dependent correction ( $K/\lambda^4$ , where K is an empirical constant) was then applied. The value of the K-constant was adjusted to reduce the blue end (260 nm) of the spectrum to those of reference solution spectra of tyrosine (Tyr) and tryptophan (Trp); (3) Gaussian curves were then fitted to the baseline-corrected spectrum until the sum of residuals (experimental-minus-calculated) was less than about 1%.

#### 2.9.4 Histidine determination using UV-spectroscopy

The spectra of His and urea are overlapping but distinguishable at wavelengths above 215 nm, thus allowing the amount of His to be quantified by comparison with difference spectra obtained with the same amount of urea and a known concentration of His. In the event, we showed that 1 mM His dissolved in 4 M urea yielded a differential absorbance value at 220 nm ( $\Delta A_{220}$ ) of 0.64.

Assuming a total of 30 His residues per mol (pro- $\alpha_1$ )<sub>2</sub>(pro- $\alpha_2$ ) procollagen (SWISSPROT accession nos. P02454 and P02466, respectively), of which only 13 are located in the “core”

## Materials and Methods

collagen sequences (2 in  $\alpha_1$  at position 256 and 1099; 9 His in  $\alpha_2$  at position 152, 188, 518, 950, 972, 981, 1031, 1044 and 1076). In contrast to core region, 7 His are in pro- $\alpha_1$  propeptides and only 3 His in pro- $\alpha_2$  propeptides region.

### 2.10 Reverse-phase-HPLC (RP-HPLC)

For separation and analysis of the CNBr-cleaved/tryptic digested peptides, RP-HPLC was performed with an SMART system, using a C18-column 30 nm column. Peptides were applied to the column using 0.1% trifluoroacetic acid (24) (v/v) in H<sub>2</sub>O and then eluted with a gradient containing 7-40% acetonitrile/ 0.1% TFA (v/v). The flow rate was 0.1 ml/min. Peptides were monitored with three different wavelengths (215 nm, 256 nm and 280 nm). Prior to loading to the column, the dried CNBr-cleaved/tryptic digested peptides were dissolved in aqueous solution containing 0.1% (v/v) TFA and 5% acetonitrile.

### 2.11 Atomic force microscopy.

The dialyzed urea extract was characterized by AFM. Images were captured using a CP microscope (Park Scientific Instrument Auto probe microscope CP/Veeco) in the tapping mode by using silicon-etched RTESPA cantilevers (Veeco Nanoprobes), which have a nominal tip radius smaller than 10 nm and a spring constant of 20-80 N/m. The drive frequency was set at 200-400 kHz with integral and proportional gains of 0.5-1.5. The scan rate was 1.0 Hz and the scan size was 2-5 and 20-50  $\mu\text{m}$  for collection of heights. All images were flattened and zoomed off-line with ProScan software.

### 2.12 Scanning electron microscopy (SEM)

#### 2.12.1 SEM of collagen

The urea-extracted proteins were gelled during the dialysis against water and then cut to 1-2 mm thick pieces. These were fixed in 1% glutaraldehyde for 2 h at 4 °C followed by several washing steps. Then, the fixed samples were dried at critical point. The SEM-imaging was performed by using a field emission electron microscope (FEEM) (Model 1530 VP, Leo Electron Microscopy Ltd. US.).

#### 2.12.2 SEM of cells grown on different materials

NIH 3T3 fibroblasts were cultivated routinely as described (25). For morphology studies, 5000 cells were seeded to the collagen-coated glass slides and cultivated for one week. Medium was removed; samples were washed twice in PBS buffer and fixed for 30 min at

## Materials and Methods

4 °C with 2.5% glutaraldehyde in PBS buffer, pH 7.4). Following fixation, samples were treated for 30 min with a 2% osmium tetroxide solution in PBS buffer. Samples were then dehydrated with graded ethanol (from 50% to 100%) and eventually were sputtered according to standard protocols. The SEM-imaging was performed by using a field emission electron microscope (FEEM) (Model 1530 VP, Leo Electron Microscopy Ltd. US.).

### 2.13 Cell culture

To characterize the urea-extracted collagen and compare the properties of both UC and AC, different cells were grown on these collagens.

The fibroblast 3T3 cell line was grown in Dulbecco's modified Eagle medium (DMEM) (GIBCO) in tissue culture flask (Greiner) at 37 °C, 5% CO<sub>2</sub>, until reaching confluence. This DMEM contains 10% fetal calf serum (FCS) (Clonetics), 2 mM L-glutamine, 0.1% gentamicin, 0.45% glucose and 0.1% penicillin-streptomycin.

The same amount of UC and AC were dissolved in DMEM (1 mg/ml) and dropped into 24 well-plate with inserts for expression analysis and into 6 well-plate with cover slips for phenotypical studies by SEM. The collagen solutions were put under clean bench for 2 h so that the collagen could gel. Then 10<sup>4</sup> and 2.5x10<sup>4</sup> cells in 100 µl DMEM were dropped onto the collagen surface in the inserts. For the collagen-coated cover slips only 5000 cells were given [156]. For both studies the cells were cultivated for 14 days with medium changes every 2 days.

Cell vitality was judged by the live/dead cytotoxicity test. The samples were washed twice in phosphate-buffered saline (PBS) and stained for 60 min with 200 µl calcein/ethidium homodimer-1 solution. Subsequently, the samples were washed in PBS and fixed in 4% glutaraldehyde.

### 2.14. Real Time-PCR (RT-PCR)

To study the difference in cell behavior on different collagens, specific genes were chosen to be compared at a transcriptional level using real time polymerase chain reactions (RT-PCR). Three culture conditions were compared cells growing on a plastic surface used as reference culture. Cells grew on urea-extracted (UC) and acetic acid-extracted collagens (AC) were compared to the reference culture.

#### 1. RNA isolation

Trypsin was given to every well and incubated at 37 °C for 5 min. Cell suspension or gel form collagen cultures were dropped into liquid nitrogen for generating cell pearls. Cell pearls were then disrupted mechanically by grinding in a pre-cooled Retsch-mill (Retsch, Haan, Germany)

## Materials and Methods

and RNA was further purified using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following instructions of the manufacturer.

### 2. cDNA synthesis

Purified RNA was reverse transcribed to cDNA using QuantiTect reverse transcription kit following instructions of the manufacturer.

### 3. RT-PCR

At first, standard curves of selected genes were generated, whereby a dilution series of cDNA from reference culture were amplified in duplicate and detected using Universal Probe Library Probes (UPL probes). The glyceraldehyde-3-phosphate dehydrogenase (G3P) was chose as a reference gene with the assumption that the transcript amount of this gene in any culture under any condition is approximately equal. The cDNAs from UC and AC under different culture conditions were amplified and detected using Universal Probe Library Probes. Eventually, the transcript amounts from genes of interest were compared to these in reference culture normalized by comparison of the transcript amounts of the reference gene in the samples.

Primer set and UPL probes were given in table 2. All the primers were synthesized and purified using HPLC (TIB MOLBIO, Berlin, Germany).

Gene	UPL Nr.	Forward Primer	Reverse primer
G3P	29	gagccaaacgggtcatca	catatttctcgtggtcacacc
Actin $\beta$	64	ctaaggccaaccgtgaaaag	accagaggcatacagggaca
Rac1	77	agatgcaggccatcaagtgt	gagcaggcaggtttaccaa
Rha	92	gaatgacgagcacacgagac	tctgtttgccatatctctgc
Fak1	45	gctacaatgagggtgtcaagc	ggtaagggtggcagtgg
Col $\alpha$ 1	15	atgttcagctttgtggacctc	gcagctgacttcagggatgt
Co1 $\alpha$ 2	83	accctgctacgacaacgtg	ccttggttagggtaaatcca
Co3 $\alpha$ 1	7	gagcaagtcttcggaaacca	tcattaatgggcctttcctct
TGF- $\beta$ 1	72	tggagcaacatgtggaactc	gtcagcagccggttacca
TGF- $\beta$ 2	73	tggagttcagacactcaacaca	aagcttcgggatttatggtgt
VEGF $\alpha$	1	aaaaacgaaagcgaagaaa	tttctccgctctgaacaagg
MMP3	7	ttgttctttgatgcagtcagc	gatttgcgcaaaaagtgc
MMP9	19	acgacatagacggcatcca	gctgtggttcagttgtggtg

**Table 2. Primer sets and UPL probes combinations for the RT-PCR.**

RT-PCR program:



## Materials and Methods

1. Initial denaturation: To activate the DNA polymerase and denature the cDNA, the temperature is increased to 95 °C at a rate of 4.4 °C/s and held for 5 min. This is only one cycle without any data collection.
2. Amplification: The temperature is first increased to 95 °C (4.4 °C/s) and held for 10 sec. The temperature is ramped down to 60 °C (2.2 °C/s) according to the annealing temperature of the primer (As the primers were selected using the online program from the manufacture, <https://www.roche-applied-science.com/sis/rtPCR/upl/center.jsp?id=030000>, the annealing temperature is standardized by 60 °C). 60 °C will be held for 10 sec to allow the binding of primer to DNA. The temperature is then increased to 72 °C (4.4 °C/s) and held for 1 sec whereby the DNA will be amplified and the fluorescence signal will be detected only once. The whole process will be normally repeated for 45 cycles and could be extended by additional cycles if the input cDNA amount was limited.
3. Cooling. After the amplification steps are completed the samples will be cooled down to 40 °C (2.5 °C/s).

After the completing of the RT-PCR, the data were relatively quantified using the Light Cycler 480 relative quantification software version 1.0. For this analysis an efficient-method was applied, which is based on the true efficiency of the each reaction of every genes. Thus the efficiency for PCR is gene specific and the reliability of the quantification is strongly enhanced as compare to the  $\Delta\Delta C_T$ -method which is based on the assumption that the efficiency of all reactions is always 2.

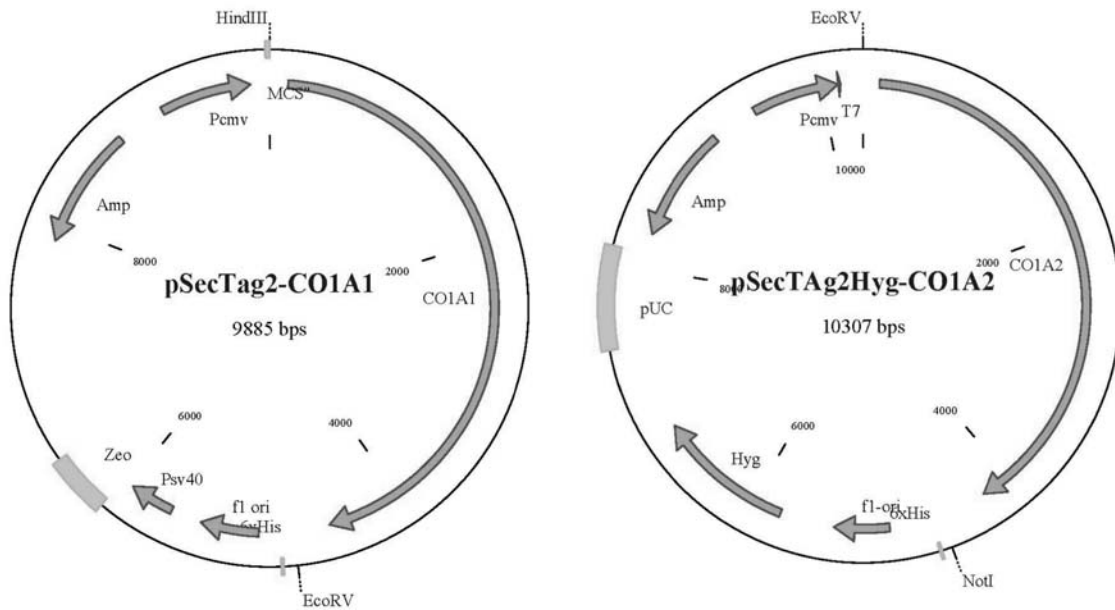
### 2.15 Type I collagen expression

#### 2.15.1 Cell culture and transfection

To study human type I collagen, the full length cDNA clones of CO1A1 (IRATp970G0431D) and CO1A2 (IRATp970A0576D) were purchased from Resource center of genome research (RZPD, Berlin, Germany). The plasmids were first isolated from the *E. coli* and purified using Miniprep Kit (Qiagen, Germany) following the instruction of manufacture. The coding sequences of the both collagen genes were then amplified by PCR using high fidelity *Pfu* DNA polymerase (Promega). The PCR products were purified using QIAquick PCR purification Kit (Qiagen, Germany) and proved by Agarose-gel electrophoresis.

## Materials and Methods

These purified PCR products were then cloned into pSecTag 2 A-Zeocin and pSec Tag 2 A-Hygromycin vector respectively. For the insertion of CO1A1, EcoR V and Hind III were applied. For the CO1A2 insertion, EcoR V and Not I cut sites were chosen.



**Figure 7. Construction of the expression vector inserted with Col 1 $\alpha$ 1 and Col1 $\alpha$ 2 full length DNA.** The plasmid with inserts were then transferred into *E. coli* DH5 $\alpha$  strain and amplified. After the isolation of the plasmid from *E. coli*, the insertion was verified by digestion with different restriction enzymes such as Hind III, EcoR V and BamH I.

To express the inserted collagen genes, the plasmids were further transfected into competent Chinese Hamster ovary cells K1 (CHO) in the means of the electroporation following construction of manufacture. The transfected cells were then cultivated in 6-well plate. After 24 h, the medium was removed and the plate was washed with fresh medium. 48 h later, the medium was changed again and fresh medium supplemented with zeocin or hygromycin were given. The cells were split if the growth reached 25% confluent into new plate with fresh selective medium. After four more days, the stable clone was picked out and cultivated in fresh medium using culture flasks for expression of the collagen  $\alpha_1$  and  $\alpha_2$  chains.

### 2.15.2 Expression and purification

The cells were cultivated for 7 days and centrifuged by 2000xg for 3 min. the supernatant was directly loaded onto the nickel-chelating resin column (ProBond<sup>TM</sup>. Metal-Binding Resin, Invitrogen) following manufacture's recommendation. Through the 6xHis tag, the expressed collagen peptides could be specifically bound on the nickel-chelating resin, thus the expressed peptides could be concentrated and purified. The elution was performed using 8 M urea solution. The purified peptides were analyzed using SDS-PAGE and Western blots.

## Results

### 3. Results

#### 3.1 Collagen extraction and SDS-PAGE analysis.

In order to study the structure and function of Type I collagen, we developed a new method for collagen extraction. Firstly, a procedure for isolation and purification of Type I collagen from rat tail tendon was established and this collagen is judged to be more suitable for application in tissue engineering.

Rat tail tendons were extracted using 9 M urea and after centrifugation at 4000xg for 15 min, the collagen samples obtained yielded a clear solution. The rate of the extraction was calculated from three independent isolations using different amount of tendons. The tendons were first washed in PBS then lyophilized in tubes. The lyophilized tendons could be easily dissolved in 9 M urea and showed a clear solution after the centrifugation. The dry weight was weighted out and pursuant urea solution was given in every tube so that the dry tendon/urea ratio was 50 mg/ml. After 20 h incubation the extracted collagen was centrifuged and the supernatant was removed and dialyzed against water. The yield of purified collagen is between 90% and 95% (see also 2.3.2).

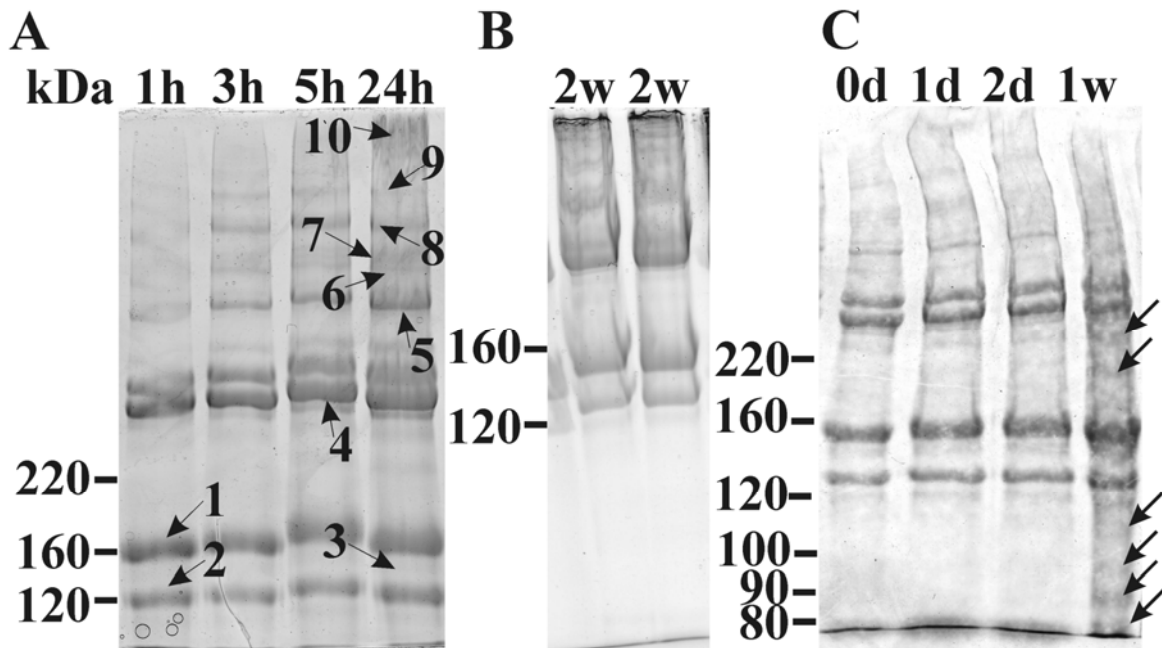
Isolation	RTT. Dry weight (g)	PBS soluble Dry weight (g)	Insoluble Dry weight (g)	yield
1	0.1955	0.0048	0.0047	95.1%
2	0.2644	0.0037	0.0146	93.1%
3	0.3202	0.0085	0.0167	92.1%

**Table 3. Yield of collagen by urea-extraction.**

The isolated collagen was analyzed by 4% SDS-PAGE gel and the characteristic bands for  $\alpha_1$  and  $\alpha_2$  were indicated with arrows numbered with 1 and 2 (Fig. 8). During the incubation in 9 M urea several samples were picked after 1 h, 3 h, 5 h, and 24 h respectively to test if the collagen is degraded. As shown in Fig. 8A, the longer the incubation the more collagen could be dissolved in the 9 M urea. The amounts of the high molecular weight fractions were increased during the incubation. A smear cross by the high molecular weight fractions was observed in all samples (Fig. 8A). Even though tendons were incubated for 2 weeks in 9 M urea solution, no degradation was observed. There was no difference in the SDS-PAGE if the extracted collagen was treated with DTT prior to loading, indicating that the high molecular weight fractions are not due to disulfide bridging (Fig. 8B). Furthermore, the urea-extracted collagen was treated in 0.5 M acetic acid for 7 days. The collagen was degraded continuously

## Results

during the incubation which is shown in Fig. 8C (arrows). The intensity of the high molecular weight fractions were significantly reduced over time ( Fig. 8C).



**Figure 8. Urea extraction of collagen and acetic degradation.** (A) Extraction of collagen from rat tail tendons. Lanes 1 to 4 were the samples taken at 1h, 3h, 5h and 24h during the incubation with 9 M urea. Arrows numbered from 1 to 10 were bands cut for ESI-MS/MS protein identification. The band 1 and 2 in lane 1 were the  $\alpha_1$  and  $\alpha_2$  chains of the collagen Type I respectively. All the arrow indicated protein bands were identified using ESI-MS/MS and these bands contains only Type I collagen. (B) Tendons were incubated in 9 M urea for 2 weeks. Lane 1 shows the sample neither reduced by DTT nor heated at 100 °C prior to loading. Lane 2 shows the reduced and heated sample. (C) Lyophilized extract from urea extraction was incubated in 0.5 M acetic acid. Lane 1 shows the input sample at the beginning, Lane 2 and 3 are the sample taken after one and two days respectively; Lane 4 is sample after 1 week incubation. The arrows indicate the bands which appeared during the incubation.

SDS-PAGE analysis performed under reducing conditions showed a complicated protein profile, with many bands above the expected molecular weight of collagen  $\alpha_1$  and  $\alpha_2$  of about 130 kDa ( $\alpha_1$ ) and 140 kDa ( $\alpha_2$ ), respectively (Figure 8A lane 1). Due to the structure of the collagen, the relative migration in polyacrylamide gels results in about a 20% increase of observed mass in comparison to that of the predicted mass [100]. No Coomassie-stained band below 100 kDa was observed in Fig. 8A and B. The urea-extracted collagen was then purified further for more detailed characterization (see below).

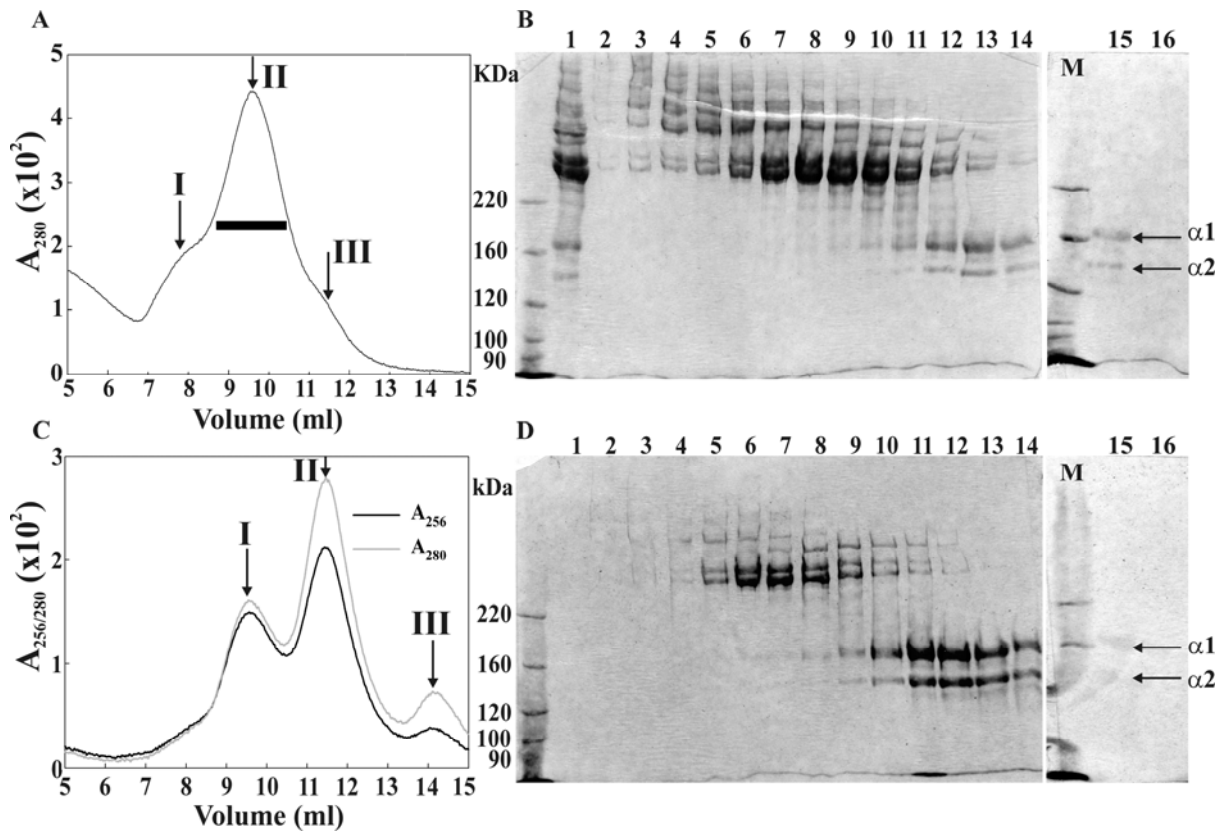
### 3.2 Superose 12 gel filtration and SDS-PAGE analysis

Subsequent purification of the urea-extracted samples by Superose 12 gel filtration in 8 M urea yielded an unexpected profile (Fig. 9A) showing three overlapping peaks at an elution

## Results

volume where high molecular weight species were expected. The SDS-PAGE lanes of the higher molecular weight fractions showed that the sub-peak I (Fig. 9A) corresponded predominantly to aggregates of an apparent high molecular weight >300 kDa but also contained lower molecular weight aggregates found predominantly in sub-peak II (Fig. 9B). However, sub-peak II also contained species with molecular weights corresponding to  $\alpha_1$  and  $\alpha_2$  which dominated the SDS-PAGE profile found in sub-peak III. We note that the aggregates observed by SDS-PAGE cannot be accounted for disulfide bridge formation as samples were routinely reduced with 5%  $\beta$ -ME prior to loading. Also, SDS-PAGE profiles obtained with thioglycolic acid in the running buffer were identical to those shown in Fig. 9 (data not shown). However, to our surprise, re-chromatography of each of the sub-peaks on the same column with 8 M urea running buffer yielded an almost identical elution profile as well as SDS-PAGE profile, although the bias to high or low molecular weight species appeared to be protein-concentration dependent. Thus, re-chromatography of the pooled fractions from sub-peak II (predominantly species with molecular weight of about 300 kDa) yielded a now larger fraction of monomeric  $\alpha_1$  and  $\alpha_2$  than the SDS-PAGE profile of the original pool suggested (Fig. 9C, D). Re-chromatography of the lower molecular weight pool III showed a higher proportion of aggregates than were present originally (data not shown). A third chromatographic step of any sub-peak also yielded the characteristic three-component profile. The distribution of protein is biased to decreasing molecular weight with each subsequent gel filtration run. This is opposite to what would be expected for aggregation of polypeptides due to covalent cross linking or non-specific hydrophobic interactions. Thus, the chromatographic behavior of the urea-extracted samples seemed to indicate a reversible association of collagen in 8 M urea. Moreover, dialysis of any of the Superose sub-peaks I, II and III against  $H_2O$  yielded a sol-gel preparation [157, 158], completely lacking any precipitates.

## Results

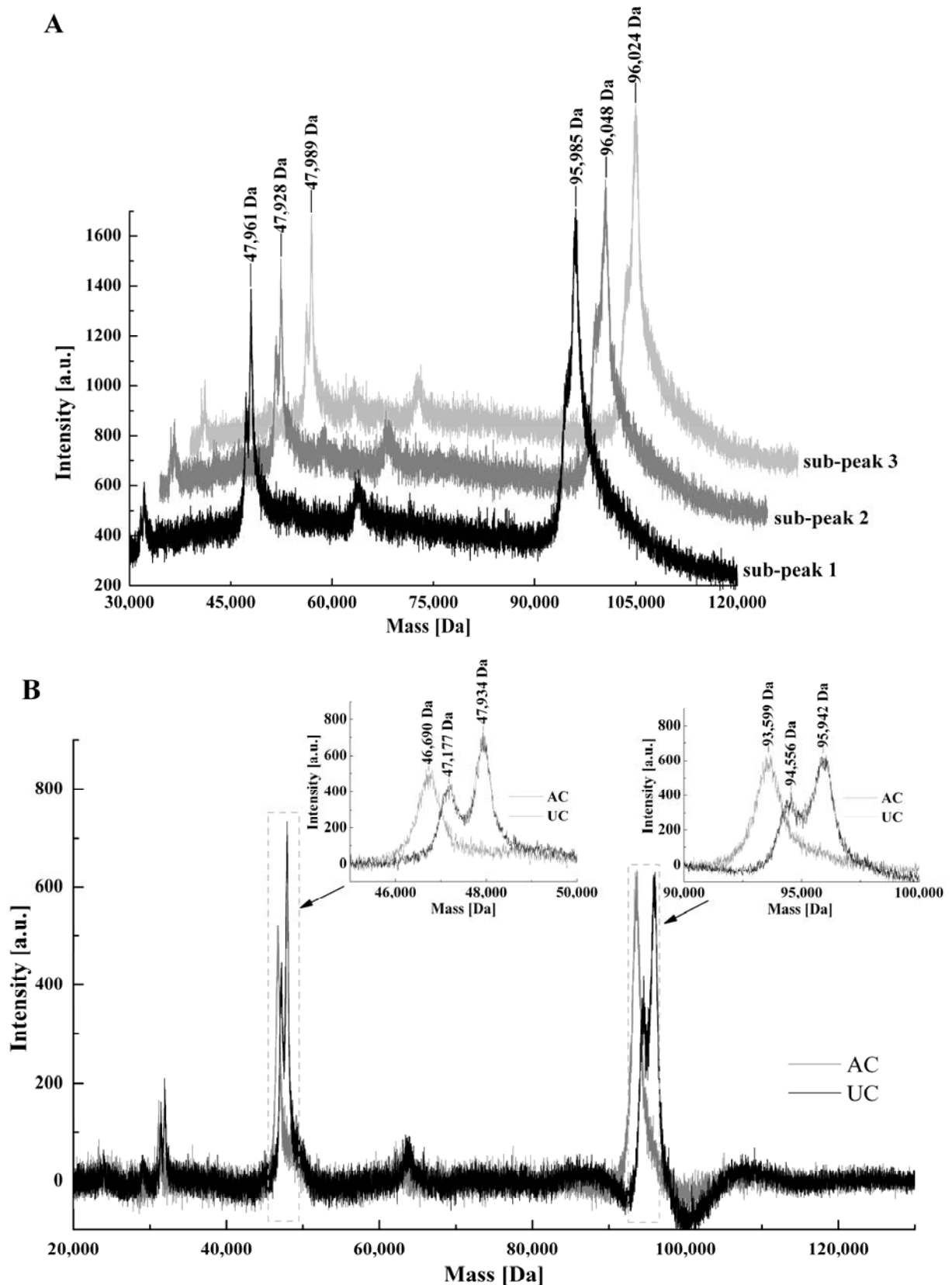


**Figure 9. Superose 12 chromatography and SDS-PAGE analysis of urea-extracted collagen.** (A) Chromatogram of the gel filtration profile following the initial urea extraction (first run). The black rectangle indicates the fractions which were pooled prior to re-chromatography on the same column (second run shown in (C,D)). (B) SDS-PAGE profile of the fractions shown in (A); (C). Re-chromatography of fractions 8-12 from the first run and the corresponding SDS-PAGE profile (D). The arrows shown in panels A and C indicated the sub-peaks (I, II, and III) mentioned in the main text.

### 3.3 Mass determination and tandem MS/MS

The mass of the isolated collagen from all three sub-peaks of the gel filtration were determined with MALDI-TOF MS using  $\beta$ -galactosidase (monomer 116.483 kDa, SwissProt Nr.:P00722) and glycogen phosphorylase B (97.289 kDa sp:P00489) as standards. The mass spectrum of the collagen about 96 kDa is shown in the Fig. 10A. The peaks to two- and three-fold charged collagen peptides were also observed in this spectrum. The fractions from the three sub-peaks were also measured by this method and no obvious difference of the mass outside of the accuracy-range (20ppm) could be observed as shown in Fig. 10A. The masses of urea-extracted and acetic acid-extracted collagen (UC; AC) were also measured (Fig. 10B) and shown to differ by up to 3 kDa. In particular, urea-extracted collagen showed 2 peaks at 95.942 kDa and 94.556 kDa while the acetic acid-extracted collagen has only one signal at 93.559 kDa. The two-fold charged peaks showed the same patterns (Fig. 10B). Scans in a high molecular weight range up to 300 kDa showed no detectable peak.

## Results



**Figure 10. (A) and (B). Mass determination and tandem MS.** (A) shows the mass determination and comparison of fractions from sub-peaks as shown in Fig. 9A. (B) shows the mass comparison of urea-extracted collagen and the acetic acid-extracted collagen. Inserts are enlarged spectra of one-fold

## Results

charged and two-fold charged collagen. The black highlighted spectrum is from UC and grey spectrum is from AC.

For protein identification, tandem MS analysis was performed as described above. The isolated collagen was digested using either trypsin or CNBr as described in methods. Nano-LC-ESI MS/MS was first applied to analyze the tryptic digested collagen samples. All the fractions in SDS-PAGE (shown in Fig. 9B) in correspondence with the three sub-peaks in chromatogram (Fig. 9 A) were analyzed separately and mass lists were then submitted to database search. All three sub-peaks showed the protein content of all three fractions to be identical.

They contained only collagen  $\alpha_1$  and  $\alpha_2$  chains and contained only peptides corresponding to rat Type I collagen with a coverage above 90% of the mature collagen  $\alpha_1$  (92.4%, shown in Fig. 11A) and  $\alpha_2$  (78.2% Fig. 11B) sequences (sp:P02454, P02466). The same results were confirmed by MALDI-TOF/TOF which submitted the mass lists to MASCOT for protein database search. Also automatic *de novo* based MS/MS protein identification using PEAKS showed the same results.



## Results

A

Sequence coverage and distribution of hydroxylations detected on Type I Collagen  $\alpha_1$ ,  
Swissprot entry P02454

						HyP <sup>a</sup> hyK <sup>b</sup>	
1	MFSFVDLRL	LLL	GATALLT	HGQEDIPEVS	CIHNGLRVPN	GETWKPDVCL	
51	ICICHNGTAV	CDGVLCKEDL	DCPNPQKREG	ECCPFCPEEY	VSPDAEVIGV		
101	EGPKGDPPGQ	GPRGPVGGPG	QDGI PGQPGL	PGPPGPPGPP	GPPGLGGNFA		
151	SQMSYGYDEK	SAGVSVPGPM	GPSGPRGLPG	PPGAPGPQGF	QGPPGEPGEP	4 (2)	
201	<b>G</b> ASGPMG <b>P</b> R <b>G</b>	<b>P</b> PGPPG <b>K</b> NGD	<b>D</b> GEAG <b>K</b> PGR <b>P</b>	<b>G</b> ERGPPGPQ <b>G</b>	ARGLPGTAGL	6 (6)	
251	PGMKGHRGFS	GLDGA	KGDTG	PAGPKGEPGS	PGENGA <b>P</b> QOM	GPRGLPGERG	4 (3) 1
301	RPFP <b>P</b> S <b>S</b> AGA	RGNDGAVGAA	GPPGPTGPTG	PPGFPGAAGA	KGEAGPQGAR	6 (5)	
351	GSEGPQGV <b>R</b> G	EPGPPG <b>P</b> AGA	AGPAGNPGAD	GQPGAKGANG	APGIAGAPGF	8 (7)	
401	PGARGPSGPQ	GPSGAPGPKG	<b>N</b> SGEPGAPGN	KGDTGAKGEP	GPAGVQGGPPG	5 (4)	
451	PAGEEGKRGA	RGEPG <b>P</b> SGLP	GPPGERGGPG	SRGFPGADGV	AGPK <b>G</b> P <b>A</b> GER	4 (3)	
501	GSPGPAGPKG	SPGEAGRPGE	AGLPGAKGLT	GSPGSPGPDG	KTGPPGPAGQ	7 (7) 1	
551	DGRPGAGPP	GARGQAGVMG	FPGPKGTAGE	PGK <b>A</b> GERGVP	GPPGAVGPAG	8 (5) 1 (0)	
601	KDGEAGA <b>Q</b> GA	PGPAGPAGER	GEQGPAGSPG	FQGLPGPAGP	PGEAGKPGEQ	5 (5)	
651	GVPDGLGAPG	PSGARGERGF	PGERGVQGGP	GPAGPRGNNG	APGNDGAKGD	3 (3) 1	
701	TGAPGAPGSQ	GAPGLQGM <b>P</b> G	ERGAAGLPGP	KGDRGDAGPK	GADGSPGKDG	6 (6)	
751	VRGLTGPIGP	PGPAGAPGDK	GETG <b>P</b> SGPAG	PTGARGAPGD	RGE <b>P</b> GGPPGA	5 (3)	
801	GFAGPPGADG	QPGAKGEPGD	TGVKGDAGPP	GPAGPAGPPG	PIGNVGAPGP	6 (5)	
851	KGSRGAAGPP	GATGFPGAAG	RVGPPGPPSGN	AGPPGPPGPV	GKEGGKGP <b>R</b> G	7 (5)	
901	ETGPAGRPGE	VGPPGPPGPA	GEKGS <b>P</b> GADG	PAGSPGTPGP	QGIAGQ <b>R</b> GVV	7 (5)	
951	GLPGQ <b>R</b> GERG	FPGLPGPSGE	PGKQGP <b>S</b> GAS	GERGPPGPMG	PPGLAGPPGE	7 (4) 1 (0)	
1001	SGREGSPGAE	GSPGRDGAPG	AKGDRGETGP	AGPPGAPGAP	GAPGPVGPAG	8 (5)	
1051	KNGDRGETGP	AGPAGPIGPA	GARGPAGPQ <b>G</b>	PRGDKGETGE	QGDRGIK <b>G</b> HR	0 1	
1101	GFSG <b>L</b> Q <b>G</b> PPG	SPGSPGEQGP	SGASGPAGPR	GPPGSAGSPG	KDGLNGLPGP	7 (5)	
1151	IGPPGPRGRT	GDSGPAGPPG	PPGPPGPPGP	PSGGYDF <b>S</b> FL	PQPPQ <b>E</b> KSQD	7 (4)	
1201	GGRY <b>R</b> ADDA	NVVRDRDLEV	DTTLK <b>S</b> LSQ <b>Q</b>	IENIR <b>S</b> PEGS	RKNPARTCRD	0	
1251	LKMCHSDWKS	GEYWIDPNQ <b>G</b>	CNLDAIKVY <b>C</b>	NMETGQ <b>T</b> CVF	PTQPSV <b>P</b> QKN		
1301	WYISPNPKEK	KHVWFGESMT	DGFQFEY <b>G</b> SE	GSDPADVA <b>I</b> Q	LTFLRLM <b>S</b> TE		
1351	ASQ <b>N</b> ITYHCK	NSVAYMDQ <b>Q</b> T	G <b>N</b> LK <b>S</b> LL <b>L</b> Q	GSNEIE <b>L</b> RGE	GNSRFTY <b>S</b> TL		
1401	VDG <b>T</b> SHTGT	WGKT <b>V</b> IEYKT	TK <b>T</b> SRL <b>P</b> IID	VAPLD <b>I</b> GAPD	Q <b>E</b> FGMD <b>I</b> GPA		
1451	CFV						
total:						120 (92) 6 (4)	

## Results

### B

Sequence coverage and distribution of hydroxylations detected on Type I Collagen  $\alpha_2$ ,  
Swissprot entry: P02466

					hyP <sup>a</sup>	hyK <sup>b</sup>
1	MLSFVDTRL	LLLAVTSCLA	TCQSLQMGSSV	RKGPTGDRGP	RGQRGPAGPR	
51	GRDGVDPVVG	PPGPPGAPGP	PGPPGPPGLT	GNFAAQYSDK	<b>GVSAGPGPMG</b>	0
101	<b>LMGPRGPPGA</b>	<b>VGAPGPQGFQ</b>	<b>GPAGEPGEPEG</b>	<b>QTGPAGSRGP</b>	<b>AGPPGKAGED</b>	5 (3)
151	<b>GHPGKPRPG</b>	<b>ERGVVGPQGA</b>	<b>RGFPPTPLP</b>	<b>GFKGIRGHNG</b>	<b>LDGLKQPGA</b>	7 2
201	<b>QGVKGEFAP</b>	<b>GENGTGQAG</b>	<b>ARGLPGERGR</b>	<b>VGAPGPAGAR</b>	<b>GSDGVSVPVG</b>	4 (3) 1
251	<b>PAGPIGSAG</b>	<b>PGFPGAPGPK</b>	<b>GELGPVGNPG</b>	<b>PAGPAGPRGE</b>	<b>AGLPGLSGPV</b>	7 (5) 1 (0)
301	<b>GPPGNPGANG</b>	<b>LTGAKGATGL</b>	<b>PGVAGAPGLP</b>	<b>GPRGIPGPVG</b>	<b>AAGATGPRGL</b>	6
351	<b>VGEPGPAGSK</b>	<b>GETGNKGEPE</b>	<b>SAGAQQPPGP</b>	<b>SGEEGKRGSP</b>	<b>GEPGSAGPAG</b>	4 (2)
401	<b>PPGLRGSPGS</b>	<b>RGLPGADGRA</b>	<b>GVMGPPGNRG</b>	<b>STGPAGVRGP</b>	<b>NGDAGRPEPE</b>	4
451	<b>GLMGRPLPG</b>	<b>SPGNVGPAGK</b>	<b>EGPVGLPGID</b>	<b>GRPGPIGPAG</b>	<b>PRGEAGNIGF</b>	4 (3)
501	<b>PGPKGPSGD</b>	<b>GKPGEKHPG</b>	<b>LAGARGAPGP</b>	<b>DGNNGAQGP</b>	<b>GPQGVQGGK</b>	5 1
551	<b>EQGPAGPPGF</b>	<b>QGLPGPSGTA</b>	<b>GEVGKPERG</b>	<b>LPGEFGLPG</b>	<b>AGPRGERGPP</b>	6 (4)
601	<b>GESGAAGPSG</b>	<b>PIGIRGPSGA</b>	<b>PGPDGNKGEA</b>	<b>GAVGAPGSAG</b>	<b>ASGPGGLPGE</b>	4 (3)
651	<b>RGAAGIPGK</b>	<b>GEKGETGLRG</b>	<b>EIGNPGRDGA</b>	<b>RGAPGAIGAP</b>	<b>GPAGASDRG</b>	4
701	<b>EAGAAGPSGP</b>	<b>AGPRGSPGER</b>	<b>GEVGPAGPNG</b>	<b>FAGPAGSAGQ</b>	<b>PGAKGEKGTK</b>	1 (0) 2
751	<b>GPKGENGIVG</b>	<b>PTGVPGAAGP</b>	<b>SGPNGPPGPA</b>	<b>GSRGDGGPPG</b>	<b>MTGFPGAAGR</b>	4 (3) 1 (0)
801	<b>TGPPGPSGIT</b>	<b>GPPGPPGAAG</b>	<b>KEGIRGPRGD</b>	<b>QGPVGRGTGEI</b>	<b>GASGPPGFAG</b>	7 (5)
851	<b>EKGPSGEPGT</b>	<b>TGPPGTAGPQ</b>	<b>GLLGAPGILG</b>	<b>LPGSRGERGQ</b>	<b>PGIAGALGEP</b>	5 (3)
901	<b>GPLGIAGPPG</b>	<b>ARGPPGAVGS</b>	<b>PGVNGAPGEA</b>	<b>GRDGNPGSDG</b>	<b>PPGRDQQPGH</b>	6
951	<b>KGERGYPGNI</b>	<b>GPTGAAGAPG</b>	<b>PHGSVAPGAK</b>	<b>HGNRGEPPGA</b>	<b>GSPVGPVAVG</b>	3 (2)
1001	<b>PRGSPGQGI</b>	<b>RGDKGEPGDK</b>	<b>GARGLPGLKG</b>	<b>HNGLQGLPGL</b>	<b>AGLHGDQGAP</b>	3 (2) 3
1051	<b>GPVGPAGPRG</b>	<b>PAGPSGPIGK</b>	<b>DGRSGHPGPV</b>	<b>GPAGVRSQG</b>	<b>SQGPAGPPGP</b>	1 (0)
1101	<b>PGPPGPPGVS</b>	<b>GGGYDFGFEG</b>	<b>GFYRADQPRS</b>	<b>QPSLRPKDYE</b>	<b>VDATLKSLNN</b>	
1151	<b>QIETLLTPEG</b>	<b>SRKNPARTCR</b>	<b>DLRLSHPEWK</b>	<b>SDYYWIDPNQ</b>	<b>GCTMDAIKVY</b>	
1201	<b>CDFSTGETCI</b>	<b>QAQPVNTPAK</b>	<b>NAYSRAQANK</b>	<b>HVWLGETING</b>	<b>GSQFEYNAEG</b>	
1251	<b>VSSKEMATQL</b>	<b>AFMRLLANRA</b>	<b>SQNITYHCKN</b>	<b>SIAYLDEETG</b>	<b>RLNKAVILQG</b>	
1301	<b>SNDVELVAEG</b>	<b>NSRFTYTVLV</b>	<b>DGCSKKTNEW</b>	<b>DKTVIEYKTN</b>	<b>KPSRLPFLDI</b>	
1351	<b>APLDIGGTNQ</b>	<b>EFRVEVGPVC</b>	<b>FK</b>			
					total: 90 (70)	11 (9)

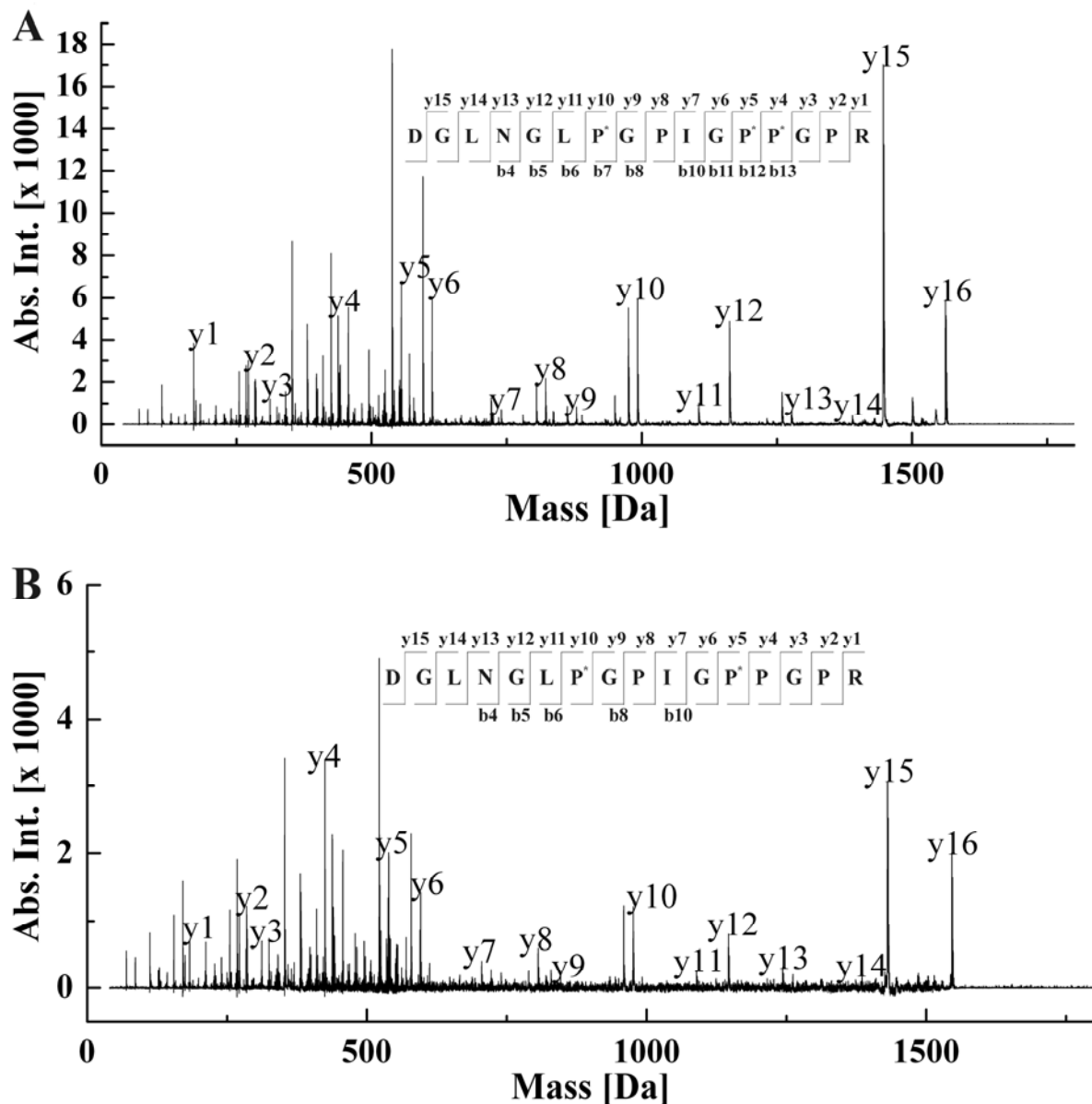
**Figure 11. (A) and (B). Tandem MS results.** The bold highlighted letters show the identified sequences and the cursive sequence were not found in the spectra. The grey parts prior and behind the core region are the propeptide sequences, which should not be present in mature collagen. The tandem MS analysis was also able to identify different PTMs in collagen Type I. About 170 proline residues in the samples were found to be hydroxylated and some peptides with the same sequence were found to contain other modifications such as phosphorylation (Table 4 and Fig. 12A, B), sulphation and glycation (data not shown).

- Maximum number of hydroxyprolines detected in this range. The minimum number detected is denoted in brackets.
- Number of hydroxylysines detected in this range.

Post-translational modifications (PTM) are thought to be crucial for formation of the tertiary and quaternary structure of collagen. Therefore, we determined some of the PTMs present in purified urea-extracted collagen (UC) using tandem MS analysis. We could detect that about 170 proline residues in the samples were found to be hydroxylated indicating that at least 25% of the proline residues in Type I collagen are post-translationally modified. Furthermore, some peptides with the same primary sequence were found to be differentially modified with

## Results

regard to Pro- and Lys-hydroxylation (Table 4 and 5; Fig. 11 A,B and Fig. 12 A, B). Differentially hydroxylated proline and lysine residues are detected in both ESI-MS/MS and MALDITOF/TOF approaches. Several peptides are picked to show this difference (Table 4 and 5).



**Figure 12. (A) and (B) Demonstration of the fragment spectra.** The peptides with same sequence but differentially modified with regard to the hydroxylation of proline residues were shown here. (A) is the fragment with mass of 1651 Da and the (B) with a mass of 1645 Da. The different hydroxylation took place on the y5. All Hyp are designated with a star. In both spectra the y-ions are indicated.

During the analyses of the MS/MS data, three different programs were applied and compared, including MASCOT, PEAKS and OMSAA.

## Results

Sequence <sup>a</sup>	Position	Mr <sup>b</sup>	Ions <sup>c</sup>	E-value <sup>d</sup>
<b>DGLNGL<u>P</u>GPIG<u>P</u>PPGPR<sup>e</sup></b>	1142-	1560.7895	14	3.1E-04
DGLNGL <u>P</u> GPIG <u>P</u> PPGPR	1157		13	8.3E-04
DGLNGL <u>P</u> GPIG <u>P</u> PPGPR			13	8.3E-04
DGLNGL <u>P</u> GPIG <u>P</u> PPGPR			12	2.6E-02
<b>DGLNGL<u>P</u>GPIG<u>P</u>PPGPR<sup>e,f</sup></b>	1142-	1544.7949	13	1.9E-04
<b>DGLNGL<u>P</u>GPIG<u>P</u>PPGPR<sup>e,f</sup></b>	1157		14	3.2E-04
<b>DGLNGL<u>P</u>GPIG<u>P</u>PPGPR<sup>e,f</sup></b>			14	3.2E-04
DGLNGL <u>P</u> GPIG <u>P</u> PPGPR			11	2.1E-02
DGLNGL <u>P</u> GPIG <u>P</u> PPGPR			12	3.7E-02
DGLNGL <u>P</u> GPIG <u>P</u> PPGPR			12	3.7E-02
TGPPGPAGQDGRPGPAGPPGAR	542- 563	2029.9571	24	2.6E-08
TGPPGPAGQDGRPGPAGPPGAR			24	3.8E-08
<b>TGPPGPAGQDGRPGPAGPPGAR<sup>e</sup></b>			25	1.9E-07
TGPPGPAGQDGRPGPAGPPGAR			25	2.7E-07
TGPPGPAGQDGRPGPAGPPGAR			25	5.2E-07
<b>TGPPGPAGQDGRPGPAGPPGAR<sup>e,f</sup></b>	542- 563	2013.9623	33	1.3E-15
<b>TGPPGPAGQDGRPGPAGPPGAR<sup>e,f</sup></b>			33	6.4E-15
TGPPGPAGQDGRPGPAGPPGAR			31	6.0E-14
TGPPGPAGQDGRPGPAGPPGAR			31	3.0E-13
TGPPGPAGQDGRPGPAGPPGAR			31	3.7E-13
TGPPGPAGQDGRPGPAGPPGAR			31	1.5E-12
<b>GETGPAGR<u>P</u>GEV<u>G</u>PP<u>G</u>PPG<u>P</u>AGEK<sup>e</sup></b>	900- 923	2215.0517	40	<1.0E-20
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			38	<1.0E-20
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			36	<1.0E-20
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			36	<1.0E-20
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			34	1.1E-15
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			38	1.4E-15
<b>GETGPAGR<u>P</u>GEV<u>G</u>PP<u>G</u>PPG<u>P</u>AGEK<sup>e,f</sup></b>	900- 923	2231.0460	38	2.7E-17
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK <sup>f</sup>			38	2.7E-17
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			38	8.6E-17
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			36	1.6E-15
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			36	2.8E-15
GETGPAGR <u>P</u> GEV <u>G</u> PP <u>G</u> PPG <u>P</u> AGEK			36	2.8E-15

**Table 4. Differential hydroxylation of proline and lysine residues in  $\alpha 1$  chain.** Three peptides with different amount and positions of Hyp are selected for illustration. The data represent a summary of data received from 4 different collagen isolates (UC). Bold peptides are found in all measured samples with varied hydroxylation patterns. No-bold peptides are found in at least one sample. Both amount and position of hydroxylation varied.

- Hydroxylated residues are underlined. Observed hydroxylation pattern(s) are highlighted. Peptides with differentially hydroxylated prolines are listed separately.
- Relative Molecular Mass calculated from measured m/z.
- Number of fragment ions matched.
- Expect value from program OMSSA.
- This pattern was supported by measurements on different peptides.
- The ranking of these patterns varied in different measurements.

## Results

Sequence <sup>a</sup>	Position	Mr <sup>b</sup>	Ions <sup>c</sup>	E-value <sup>d</sup>
<b>RGSP<u>G</u>EP<u>G</u>SAGPAG<u>P</u><u>P</u>GLR<sup>e</sup></b>	387- 405	1747.8601	31	<1.0E-20
RGSP <u>G</u> EP <u>G</u> SAGPAG <u>P</u> <u>P</u> GLR			31	<1.0E-20
RGSP <u>G</u> EP <u>G</u> SAGPAG <u>P</u> <u>P</u> GLR			28	3.0E-13
RGSP <u>G</u> EP <u>G</u> SAGPAG <u>P</u> <u>P</u> GLR			22	4.8E-07
<b>GSP<u>G</u>EP<u>G</u>SAGPAG<u>P</u><u>P</u>GLR</b>	388- 405	1607.7545	21	8.9E-12
GSP <u>G</u> EP <u>G</u> SAGPAG <u>P</u> <u>P</u> GLR			20	9.4E-11
GSP <u>G</u> EP <u>G</u> SAGPAG <u>P</u> <u>P</u> GLR			17	3.9E-06
HGNRGE <u>P</u> GPAGSVGPVGA <u>V</u> GPR	981- 1002	2040.0249	42	1.6E-19
<b>HGNRGE<u>P</u>GPAGSVGPVGA<u>V</u>GPR<sup>e</sup></b>			47	2.7E-19
<b>HGNRGE<u>P</u>GPAGSVGPVGA<u>V</u>GPR</b>	981- 1002	2024.0299	43	1.6E-19

**Table 5. Differential hydroxylation of proline and lysine residues in  $\alpha 2$  chain.** Three peptides with different amount and positions of Hyp are selected for illustration. The data represent a summary of data received from 4 different collagen isolates (UC). Bold peptides are found in all measured samples with varied hydroxylation patterns. No-bold peptides are found in at least one sample. Both amount and position of hydroxylation varied. Hydroxylated Residues are underlined. Observed hydroxylation pattern(s) are highlighted.

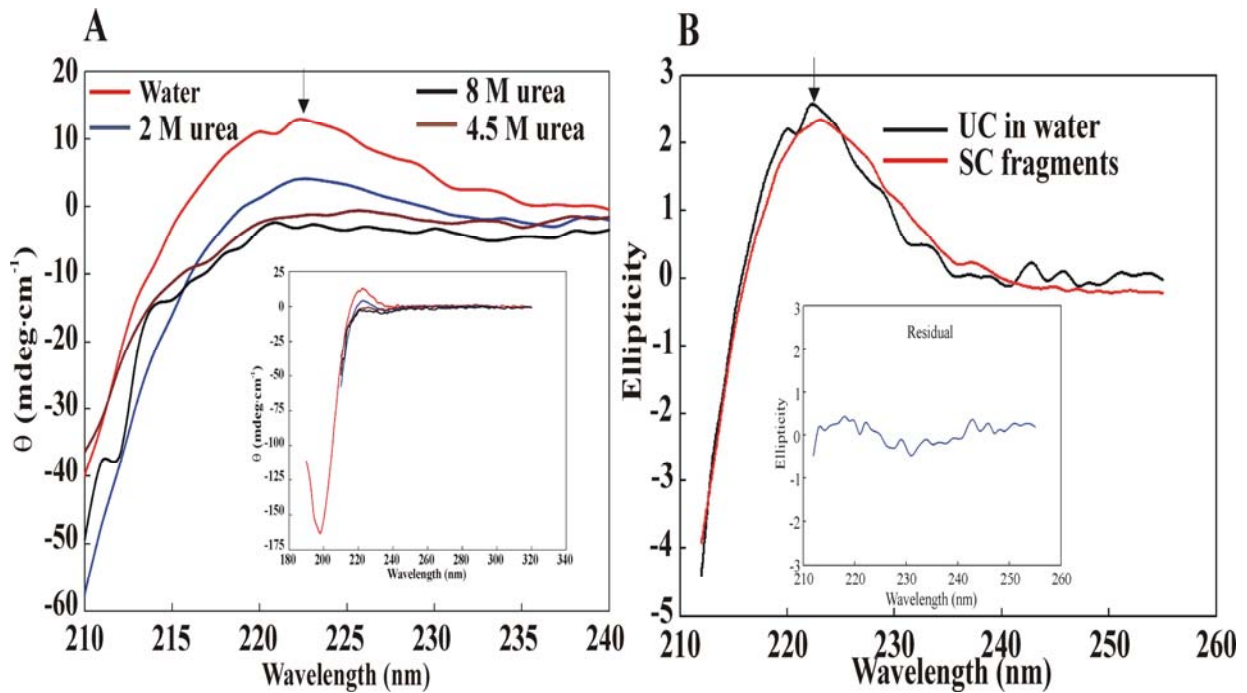
- Hydroxylated Residues are underlined. Observed hydroxylation pattern(s) are highlighted.
- Relative Molecular Mass calculated from measured m/z.
- Number of fragment ions matched.
- Expect value from program OMSSA.
- This pattern was supported by measurements on different peptides.

### 3.4 UV-CD spectroscopy

UV-CD and UV-VIS spectroscopy were employed to characterize the physical state as well as possible change of conformation of the purified UC under a variety of conditions.

UV-CD spectra (Fig. 13) of the collagen sol-gel showed the characteristic peak at 225 nm, attributable to the presence of triple helix [5].

## Results



**Figure 13.** UV-CD spectra obtained from the Superose 12-purified collagen polypeptides dissolved in different concentrations of urea. The main panel shows the spectral region accessible in the presence of different concentrations of urea. The characteristic peak at 225 nm corresponding to intact collagen triple helix is indicated by an arrow. The insert shows the complete spectrum obtained from the soluble collagen hydrogel, with the spectra from the main panel superimposed. (B) comparison of the UV-CD spectra of the UC in water obtained here and that of “pure triple helix” obtained using synthetic peptides, here designated as SC [5].

We could demonstrate a perfect correspondence between the UV-CD data shown in [5], obtained from synthetic polypeptides which form triple helix exclusively, and our data obtained from the sol-gel preparation of urea-purified collagen (Fig. 13A, red line). The fit with the published data is shown in Fig. 13B. The match of both spectra is about 98 %.

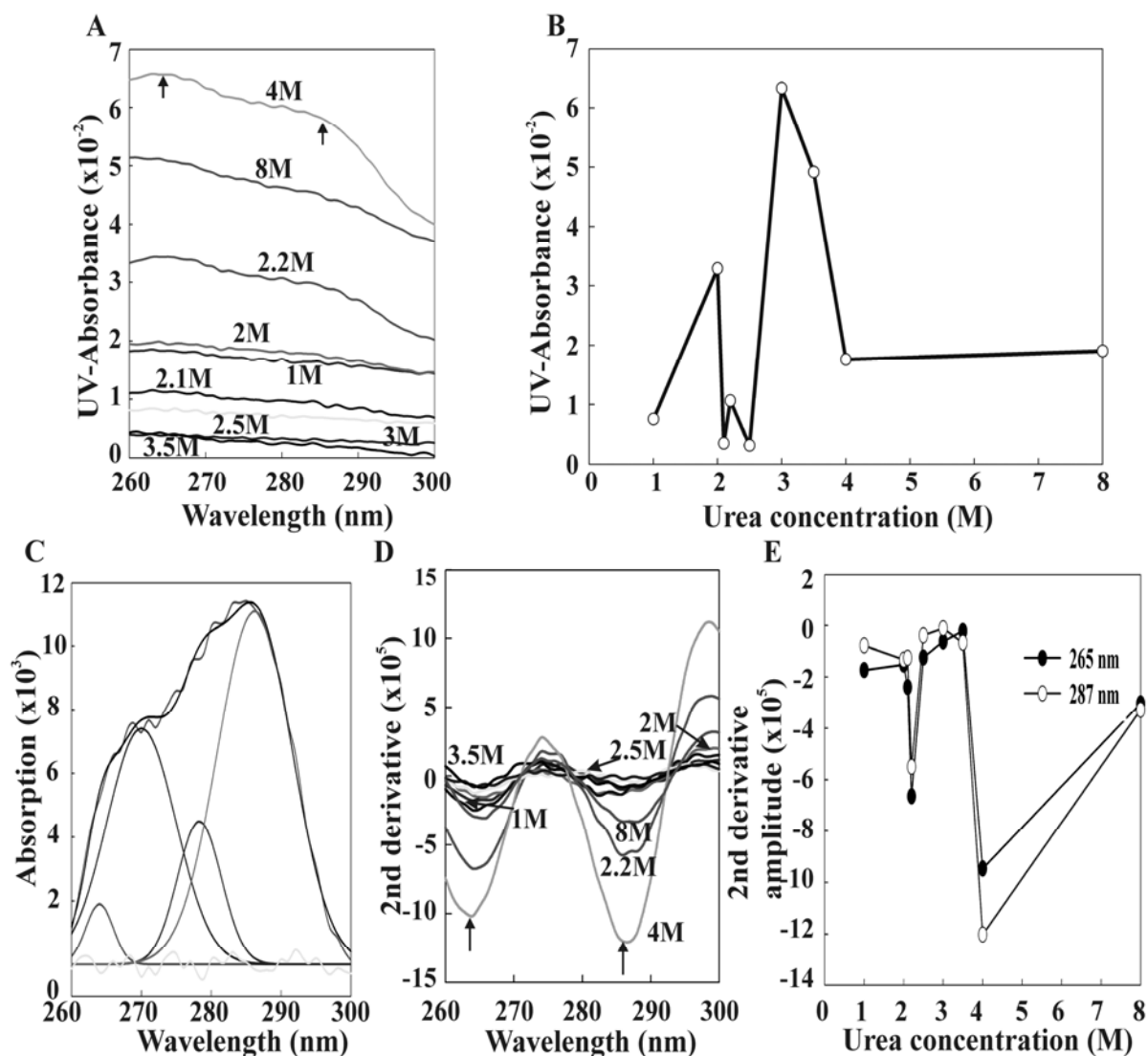
In order to examine the mechanism of collagen assembly from the urea-purified polypeptides in more detail, we measured the UV-CD spectra for collagen polypeptides dissolved in various concentrations of urea. In 8 M urea, the UV-CD spectra are characteristic of random coil as expected. Essentially, no change in the UV-CD spectrum was observed until the urea concentration was lowered to 2 M (Fig. 13A). Below 2 M urea, the characteristic triple helix peak at 225 nm as described previously [5] appeared and was maximal in the absence of urea (Fig. 13A). We did not observe a time-dependent change in the UV-CD spectra taken at concentrations 0-4 M urea over a period of about 2 h.

## Results

### 3.5 UV-VIS spectroscopy of urea-extracted collagen

To characterize the urea-mediated effects further, we obtained UV-VIS spectra of the UC in the presence of different concentrations of urea. (Fig. 14). In the range 260-300 nm two features could be observed. First, the change in turbidity as the urea concentration was lowered from 8 M to 1 M was not a monotonic function, as would be expected for the formation of increasingly larger non-specific aggregates, but followed a complex profile (Fig. 14A, B). Strikingly, between 2-4 M urea, where triple helix begins to form, the turbidity dropped to a value below that of the UC in 8 M urea, where only isolated random coil chains are present. This behavior is consistent with the appearance of ordered structures (here triple-helical aggregates), which display smaller effective hydrodynamic radii than the sum of their parts. The second feature observable is the presence of three very weak absorption maxima, with major components at 270 nm, 280 nm and 290 nm (Fig. 14C). To assess the relative intensity changes of the two major maxima we performed second-derivative analysis (which essentially removes the monotonic contribution from turbidity, while “sharpening” the peaks (Fig. 14D). The urea-profiles of the second derivative values at 265 nm and 287 nm (Fig. 14E) show a striking similarity to that of the turbidity profile (Fig. 14B). The positions of the peaks would correspond well to the peak maxima for tyrosine and its ionized form, tyrosinate. Tyrosine is present in the core region ( $5/\alpha_1$  and  $2/\alpha_2$ ). At present, we cannot explain why the intensities of the putative tyrosine peaks are so weak. Possibly, the tyrosines exist in a variety of slightly differing environments, which leads to absorption peaks with slightly different maxima, thus “smearing” the intensity over a relatively large range.

## Results



**Figure 14.** UV-spectral analysis of Superose 12-purified collagen dissolved in various concentrations of urea. (A) UV-VIS absorption spectra obtained with different concentrations of urea (indicated). The spectra shown have been obtained after subtraction of the reference which contained the appropriate urea concentration. The concentrations of urea used are indicated. (B) Turbidity plotted from the absorptions at 270 nm. (C) Gaussian multicomponent analysis of the UV-VIS spectra obtained for collagen dissolved in 2.2 M urea (shown in (A)) (D) Second-derivative analysis of the spectra shown in (A); (E) Urea concentration-dependence of the absorption maxima shown by the arrow in (A)

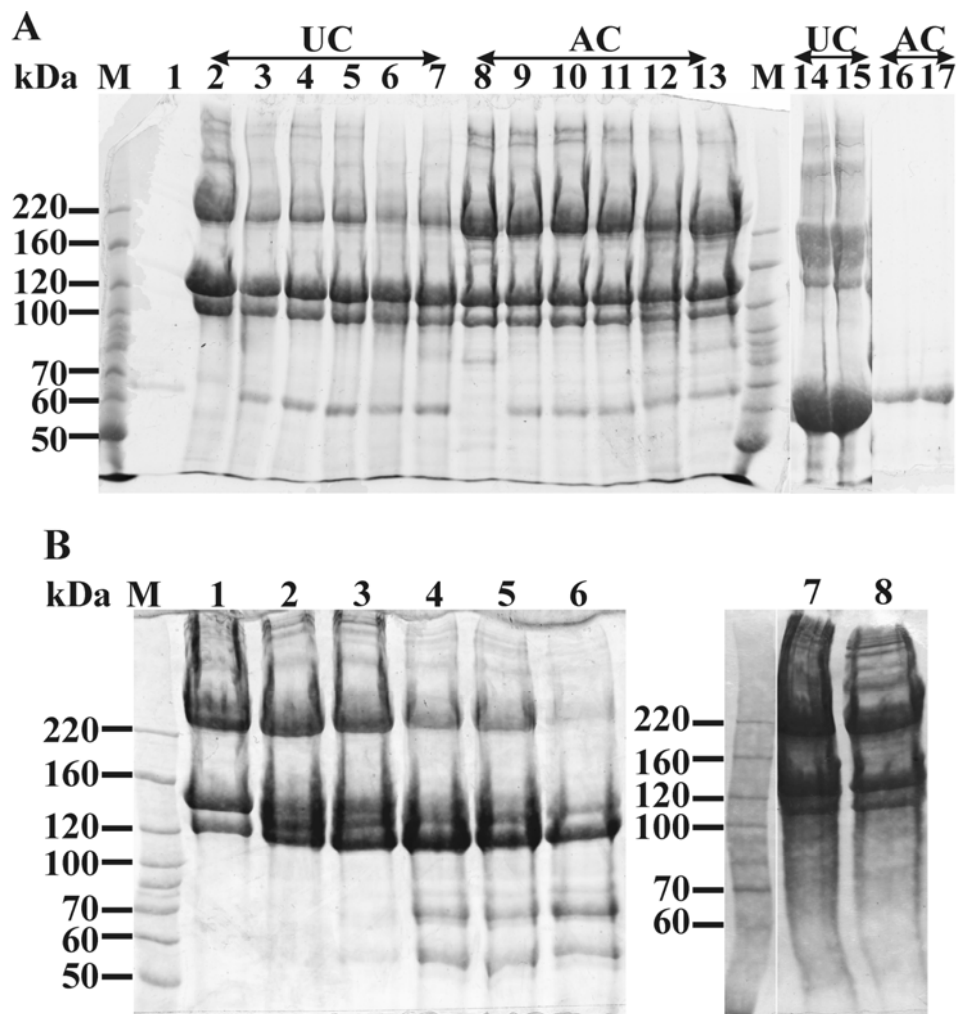
### 3.6 Carboxypeptidase and chymotrypsin digestion

To compare the structural integrity of the urea-extracted and acetic acid-extracted collagen (AC), dialyzed UC and AC samples were digested with yeast carboxypeptidase Y for 1 min to 20 h. The enzyme/collagen ratio varied from 1/200 to 1/25 (w/w). The digestions were carried out at 37 °C and stopped by addition of PMSF. The digested samples were then reduced and analyzed by 6% SDS-PAGE (Fig. 15A). As the enzyme/collagen ratio was 1/200, there was



## Results

little difference seen between the UC and AC samples. However, if the E/C ratio was increase to 1/50 and 1/25, the AC was digested much faster than the UC as shown in lane 14-17. Since the chymotrypsin digestion was shown to be useful to test the stability of the collagen and collagen  $\alpha$ -chains [159, 160], the chymotrypsin digestion was also carried out with UC samples with a E/C ratio for 1/50 at 20 °C for 1 min to 20 h. The digestion was stopped by addition of PMSF and analyzed by 6% SDS-PAGE. The proteolysis of collagen was very slow despite the high E/C ratio (Fig. 15B). These results indicated that the UC is more stable than AC due to more intact structure.



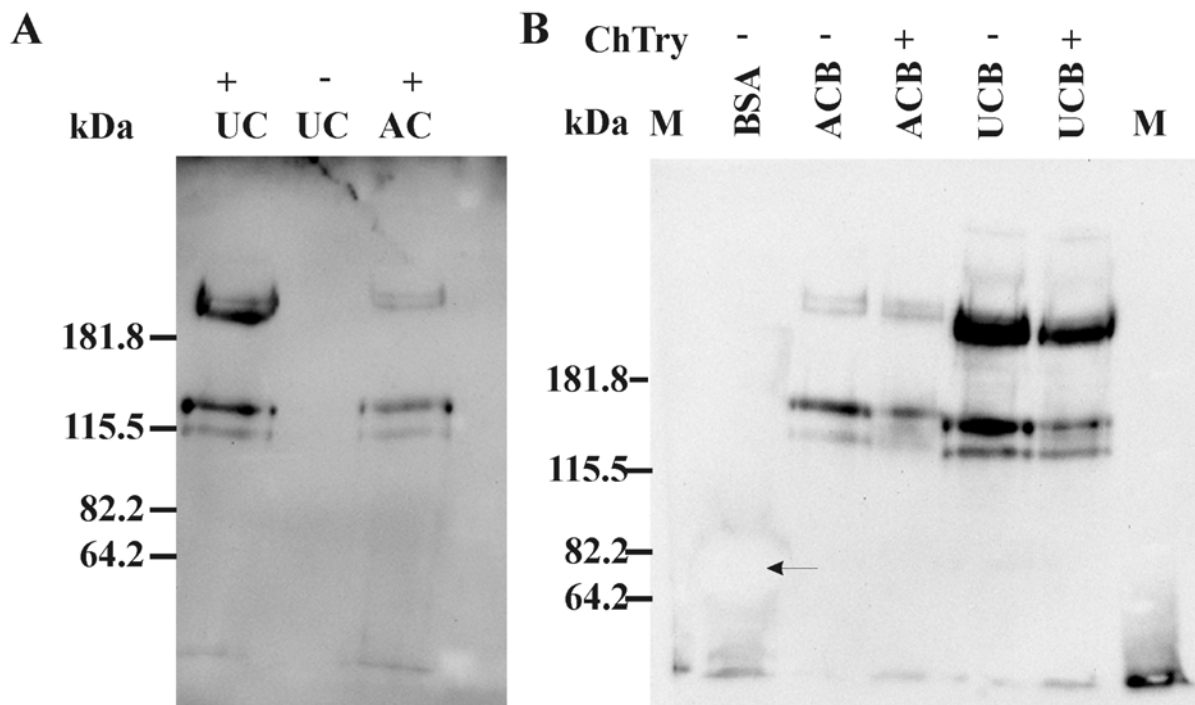
**Figure 15. SDS-PAGE (6%) profile of the carboxypeptidase digestion and chymotrypsin limited digestion.** (A) Lane 1 is yeast carboxypeptidase Y (61.0 kDa). This band can also be detected in lane 2-13. Lane 2 and 8 are the non-digested urea-extracted and acetic acid-extracted collagen respectively. The lane 3-7 are digestion for 1min, 5min, 10min, 1h and 20h; appropriate digestions were performed by AC in the same time series, as shown in lane 9-13; The enzyme to collagen ratio is 1:200 as suggested by producer. Lanes 14-15 were digestions of UC and 16-17 of AC for 20h with enzyme to collagen ratio for 1:50 (lanes 14 and 16) and 1:25 (lanes 15 and 17). (B) Chymotrypsin limited digestion of UC. Lane 1 was non-digested UC and lane 2-6 were digested for 1min, 2min, 10min, 1h

## Results

and 20h. Lanes 7 and 8 were the digestion using mixture of ChyTrp and trypsin for 2min, chymotrypsin to trypsin ratio is 10:1 (1 $\mu$ g:0.1 $\mu$ g) and 1:10 (0.1  $\mu$ g:1 $\mu$ g) respectively; 100  $\mu$ g collagen was digested using the enzyme mixture.

### 3.7 Biotinylation of collagen

To detect whether cysteine and disulfide bonds exist in UC and AC the samples were modified using EZ-link-maleimide PEO<sub>2</sub>-Biotin so that the specifically labeled collagen could be easily detected by western blot. As shown in Fig. 16A the biotin signal of the UC sample is much more intensive than the signal by AC and the non-labeled UC has no signal. The signal in Fig. 16A lane 1 correlates well to the running profile of the collagen  $\alpha_1$  as shown in Fig. 9B. Also, the higher molecular weight band corresponds to the high molecular weight fractions observed in Fig. 9B. The signal from AC sample is rather an unspecific labeling of the EZ-link-maleimide PEO<sub>2</sub>-Biotin. The blot was exposed for 20 sec. The biotinylated UC and AC were digested using chymotrypsin to see whether the triple helix structure could be affected by the biotinylation. The results after detection by Western blot indicated that the biotin-signal is weakened after the digestion. Not only at the low molecular weight range but also the high molecular weight aggregate was affected significantly by chymotrypsin.



**Figure 16. Biotinylation of UC and AC.** (A) NEM-biotin labeled UC and AC samples were shown in lane 1 and 3. Control-sample non-labeled UC was put between the both samples. (B) Chymotrypsin digested UC and AC samples. Control sample is biotinylated non-digested BSA with same amount of protein as the other samples, whereby an over-exposed signal was detected (arrow); lane 2 and 3 are

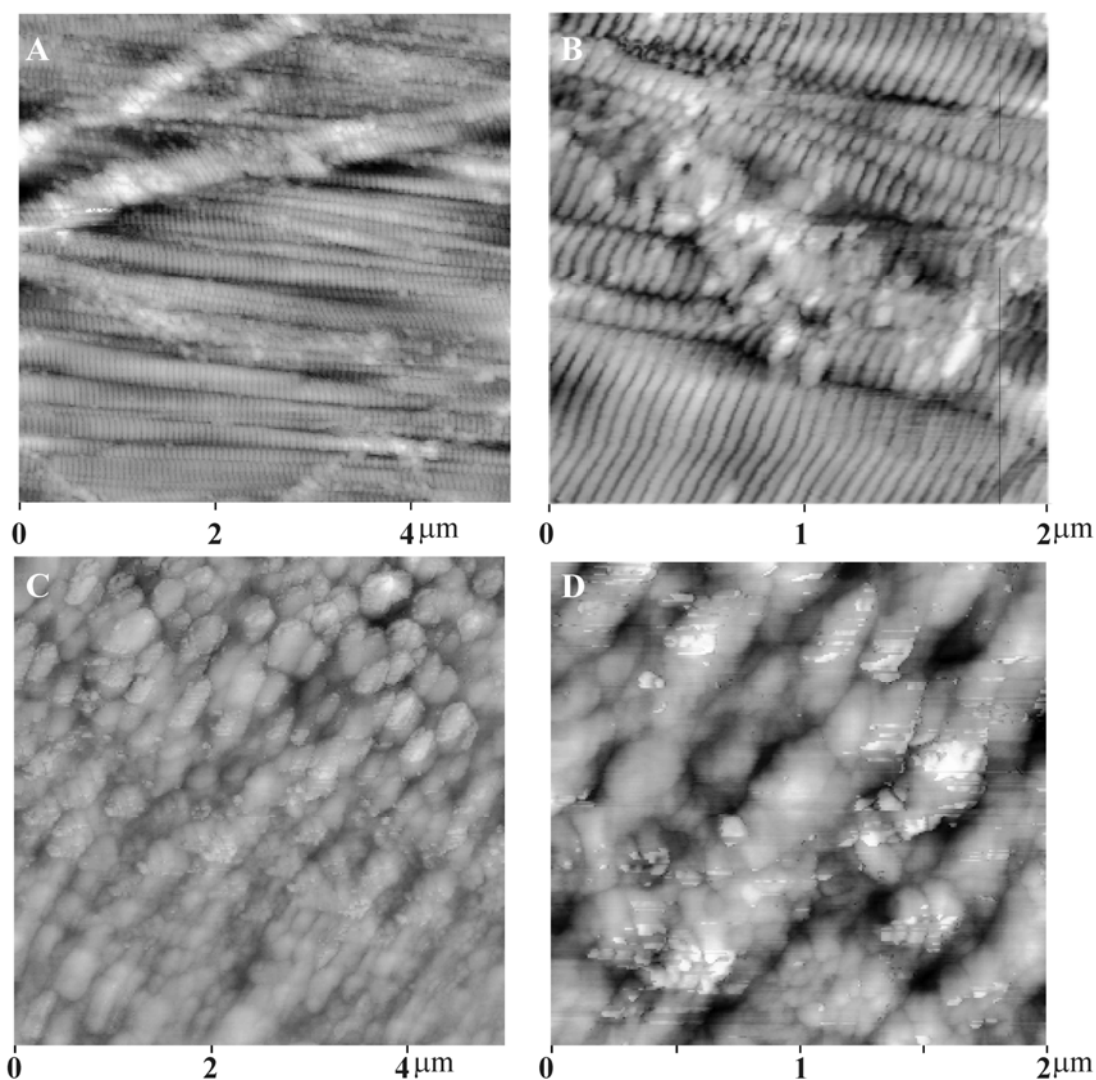
## Results

non-digested and digested biotinylated AC samples; the lane 4 and 5 are the non-digested and digested biotinylated UC samples. At the both sides of the blot are the pre-stained Bench Marker for the molecular weights standard (indicated as M).

### 3.8 AFM and SEM of collagen

Atomic force microscopy is established to study the fibril structure and fibril assembly of collagens *in vitro*. In this study dialyzed UC samples were also investigated using AFM. This was compared directly with the AFM images from PBS-washed rat tail tendons.

AFM images of RTT (Fig. 17A and B) directly after the preparation showed the characteristic fibril structure of rat tail tendon and its arrangement. The AFM images of urea-extracted collagen after gel filtration and dialysis against water also showed RTT-like fibrils (Fig. 17C and D), although the diameter of single fibrils was larger than those of natural collagen (diameter < 100 nm) [161].

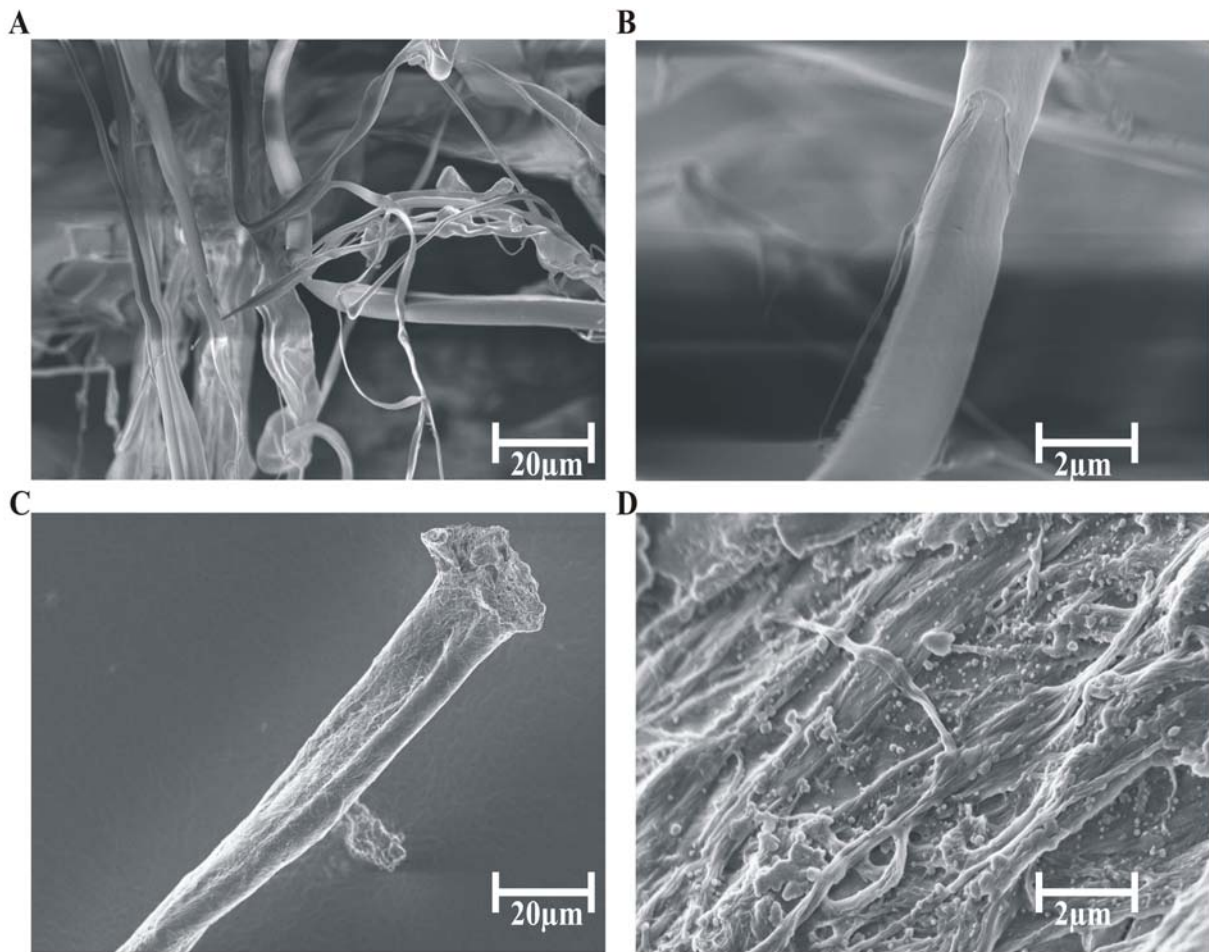


**Figure 17. Comparison of AFM images of PBS washed rat tail tendon (RTT) and dialyzed Superose 12 column fraction of UC. (A, B) the images of PBS washed rat tail tendon directly after**

## Results

the teasing out from the tail. (C, D) showed the images of urea-extracted collagen dialyzed against water after the gel filtration on Superose 12 column.

The images obtained using scanning electron microscopy (SEM) showed randomly arranged fibrils with a length more than 10  $\mu\text{m}$  and an average diameter of approximately 2.4  $\mu\text{m}$  (Fig. 18A, B), if the urea-extracted collagen was directly dialyzed against water repeatedly after the extraction without any further purification steps. After the gel filtration on a Superose 12 column and repeated dialysis against water, the SEM images also showed that fibrils were present in these samples (Fig. 18C). By zooming on the fibril, sub-fibrils what are aligned into the large fibril could also be observed (Fig. 18D).



**Figure 18. SEM Images different protein samples.** (A) and (B) show SEM images of dialyzed urea extracts which were gelled by dialysis against water. (C) and (D) were SEM images of lyophilized protein fractions after the gel filtration step as shown in Fig. 9 A and B.

### 3.9 DTNB reaction and histidine measurement

To verify if only mature or procollagen species were present in the UC, we attempted to determine the amount of Cys in UC, Cys exists only in procollagen. For determination of the

## Results

cysteine residues and procollagen concentration, DTNB reaction was applied [153, 154]. First, an equation was determined directly by calibration curve obtained with cysteine solutions as standard. Because the cysteine residues exist exclusively in both propeptides of procollagen the followed calculation under assumption that procollagen exists in UC sample.

$$A_{412} = 0.0134 \times (\text{nmol} \cdot \text{procollagen}) + 0.309(1)$$

Using this equation, calculated that a weighed 2.5 mg UC sample has 1.1 mg (pro)-collagen after calculation. This could mean that from a possible 45 Cys residues, present exclusively in the N- and C-propeptides, that only 20 residues are present in the sample. This is in contradiction to our MS data, where no propeptides could be identified. Therefore, we have to assume that other components or most probably unknown post-translational modifications might be present in the UC leading to unspecific reaction with DTNB.

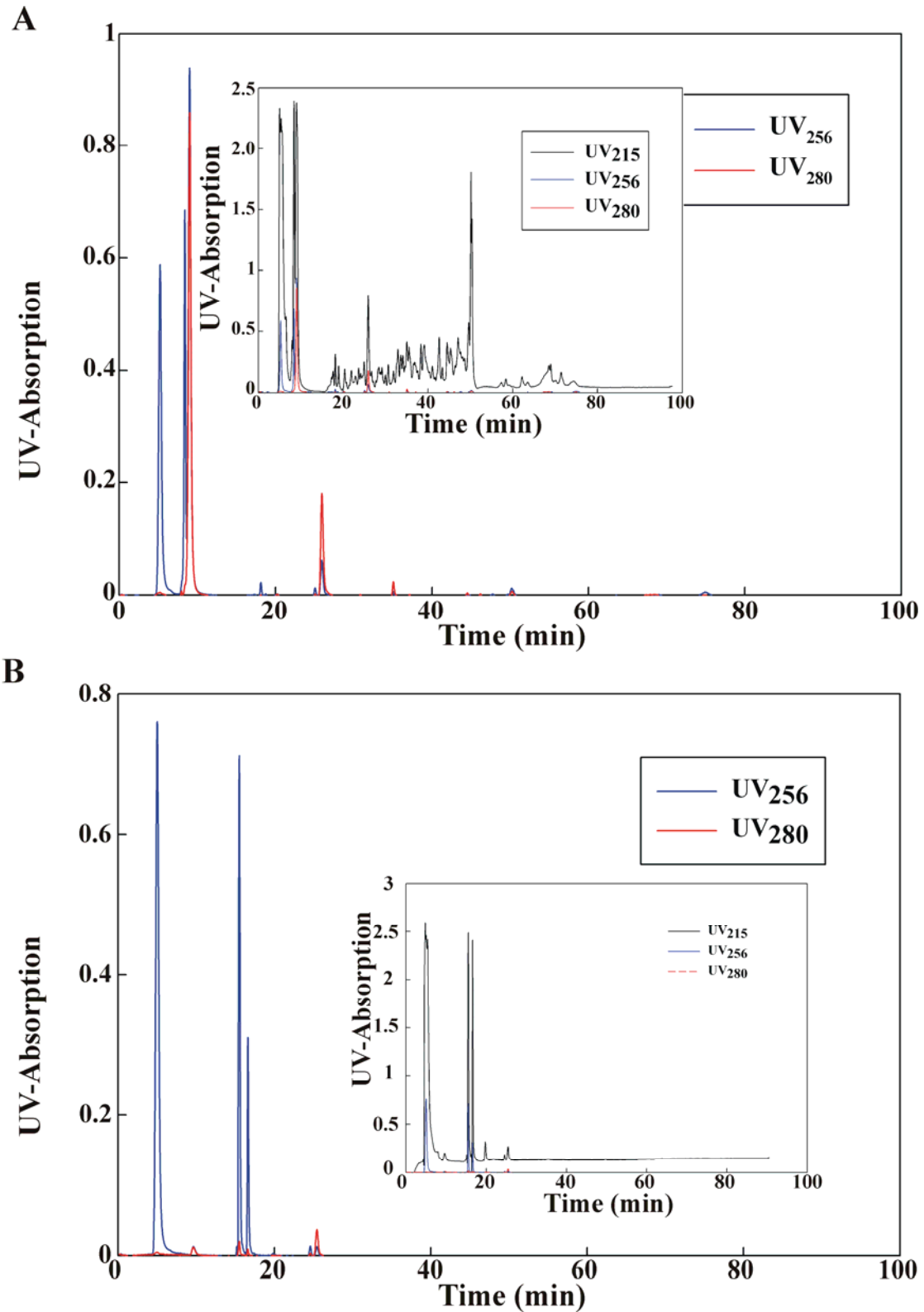
To confirm this assumption, we determined the amount of His using UV-spectroscopy. In  $\alpha_1$  chain two His are present and in  $\alpha_2$  9 His, that means in whole collagen molecule 13 His should be found. UV-spectra at 220 nm ( $\Delta A_{220}$ ) determined that a weighed 1.3 mg sample (4.6 nmol collagen) has a value of 1.6 mg (5.6 nmol collagen). This UV-spectroscopic determination confirmed the result that only mature collagen is present in UC.

### 3.10 Chromatographic analysis of cleaved collagen

To simplify the MS/MS analysis and study the UC structure in detail, UC samples were digested using trypsin and the tryptic peptides were separated by RP-HPLC using a C18 column (30 nm pore size). The fractionation and peptides were monitored with three wavelengths (215 nm, 256 nm and 280 nm) so that the absorption of peptide bonds and aromatic amino acids could be monitored simultaneously (Fig. 19). As a blank, the same amount of trypsin was also applied to the column and chromatographed using same running program (Fig. 19B).

Tryptic-digested UC samples have a very complex chromatogram as compared to the trypsin blank. Interestingly, a very strong aromatic absorption was observed only in digested UC samples as indicated in Fig. 19 A,  $A_{256}$  was shown in blue and  $A_{280}$  in red. The small inserts are the entire chromatogram which additionally shows the  $A_{215}$  in black.

## Results



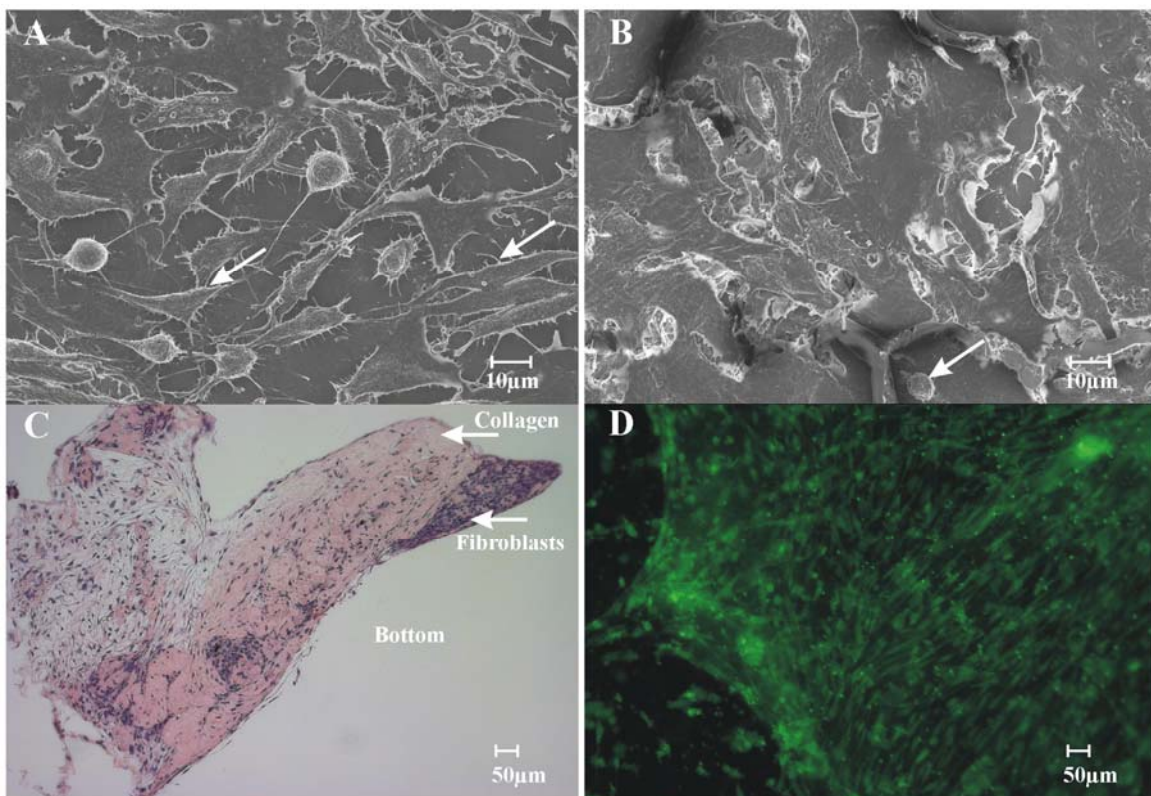
**Figure 19. RP-HPLC of tryptic digested UC samples.** (A) Tryptic digested UC sample. (B) Trypsin was used as blank sample. Black line is  $A_{215}$ ; blue line is  $A_{256}$  and red line is  $A_{280}$ . Insert are the complete chromatogram with all three wavelengths shown in one diagram.

## Results

### 3.11 Phenotypic comparison of AC and UC upon cell cultures

In order to test the suitability of UC for tissue engineering we compared the behavior of NIH 3T3 fibroblast cell cultures when grown on urea-extracted collagen and acetic acid-extracted collagen. SEM images showed significant differences between cells seeded on the different collagen extracts. The cells seeded on AC (Fig. 20B) were less abundant than the cells seeded on UC (Fig. 20A) and spread irregularly. In contrast to the AC-grown culture, the UC-grown culture was evenly distributed over the UC surface (Fig. 20A). In addition, the cells grown on UC show clearly-defined filapodia (Fig. 20A, arrows).

Also, primary fibroblasts show a distinct phenotype in UC (Fig. 20C and D.) Interestingly, although cells are most concentrated at the bottom of the insert but the cells showed a good free distribution after 14 days cultivation (Fig. 20C) and maybe began to build a new matrix so that the upper side is looser than the under side. The vitality test showed a relative homogeny arrangement of cells along an axial, which is very similar as cells *in vivo*.

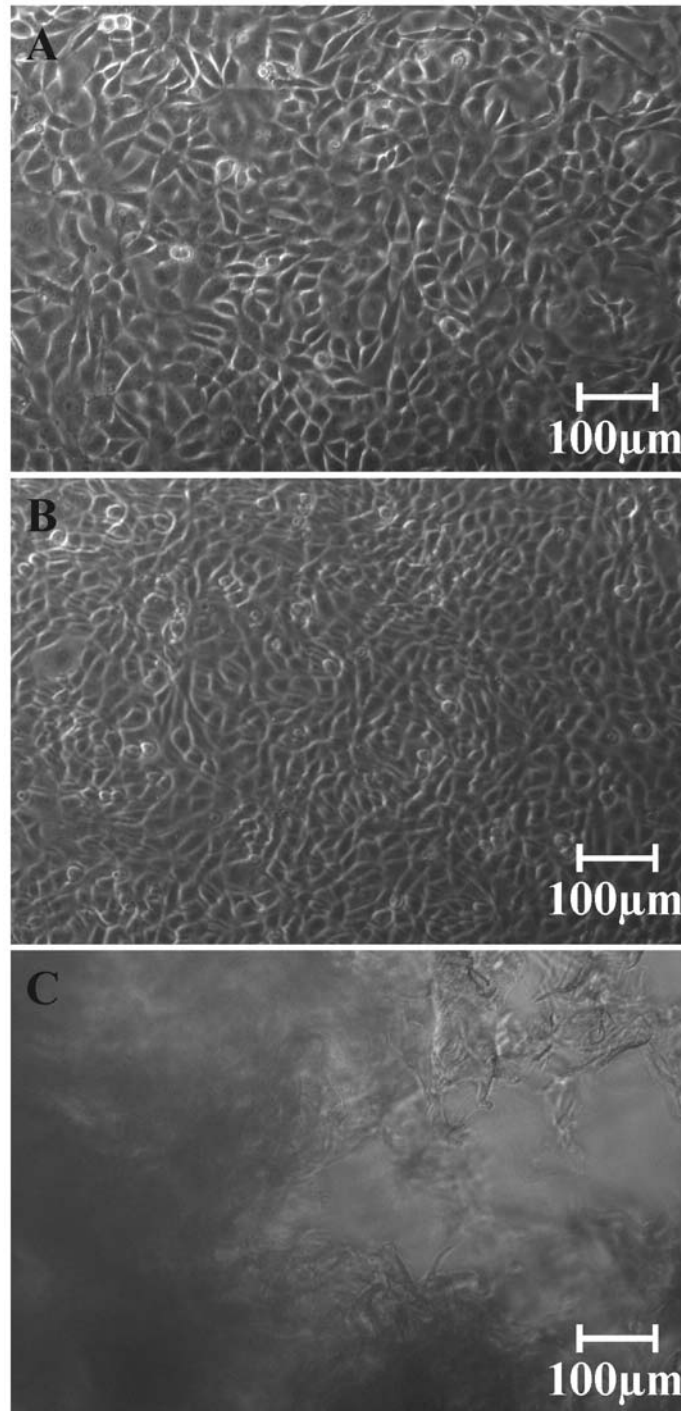


**Figure 20. Microscopic images of NIH 3T3 fibroblasts cultivated on the UC and AC coated glass slides.** (A) is the UC-grown culture and (B) AC-grown culture. The well-defined filapodia of the cells in the UC-seeded culture are indicated by arrows. (C) Primary fibroblasts (prepared from foreskin) mixed with UC and cultivated for 14 days. After H&E staining, the cells have an even distribution through the gel. (D) Vitality test showed high vitality of cells cultivated in UC.

The phenotype was also compared using the light microscopy (Fig. 21), even at this resolution the difference was significant. While the cell morphology was very similar between the

## Results

control (Fig. 21A) and cells on UC (Fig. 21B), they both show the typical fibroblast form and arrangement under both culture conditions, the cells grown on AC appeared irregular in shape even deformed (Fig.21C).

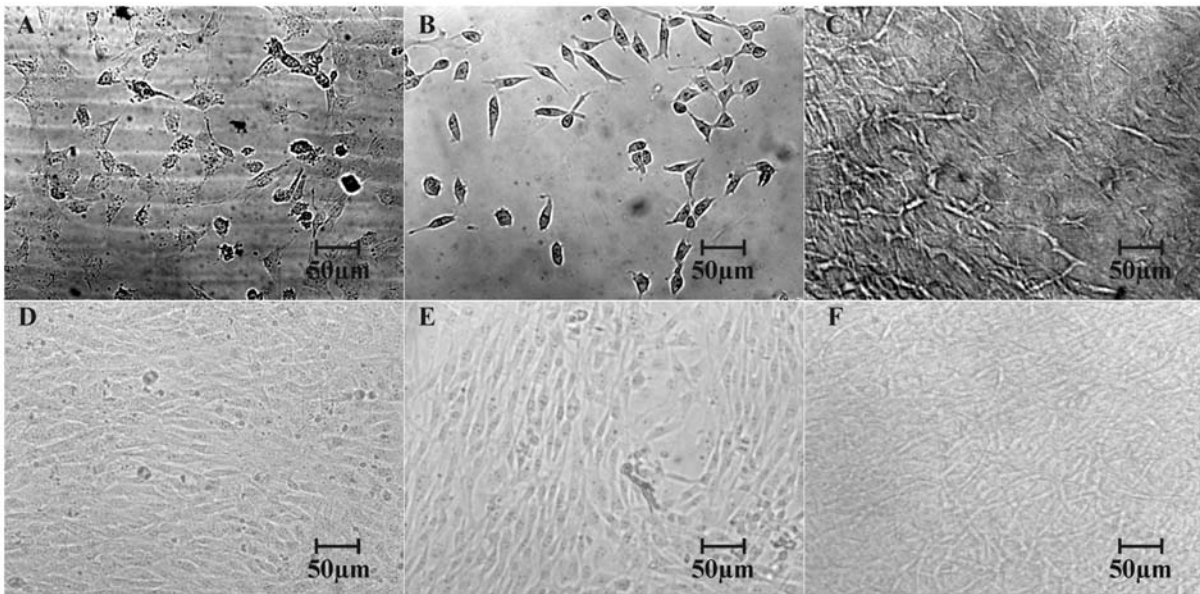


**Figure 21. Light microscopy of fibroblasts cultivated in different matrices.** (A) 3T3 fibroblast cultivated directly in D-MEM medium as control. (B) 3T3 fibroblasts mixed in UC and grown for 7 days. (C) 3T3 fibroblasts mixed in AC and grown for 7 days. All light microscopic images are shown with a magnification of 100-fold.



## Results

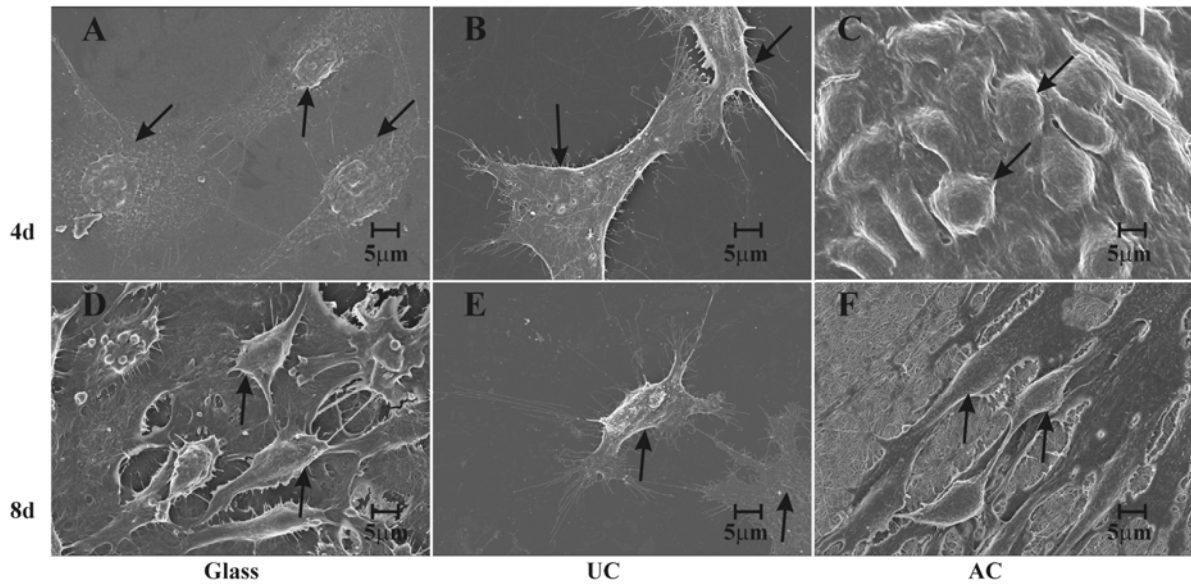
The fibroblasts cultivated on the cover slips coated with UC or AC showed significant difference according to the cell motility, growth and stress. The cells grown on cover slips without collagen as a standard are shown in Fig. 22 (A, D). Cells grown on UC-coated cover slips (Fig. 22B and E) have similar growth, orientation, and cellular morphology as in the standard. In contrast to those in the control and also the UC culture, the cells which grew on AC-coated cover slips show normal growth but a totally random alignment of the fibroblasts in culture. Fibroblasts grown either on glass cover slips or UC-surface, however, have an alignment which is typical for fibroblasts in a collagen Type I fibril environment [162].



**Figure 22. 3T3 Fibroblasts cultivated on the cover slips coated with different matrices.** (A), (D) control cultures 5000 3T3 fibroblasts seeded directly on cover slips and cultivated in D-MEM medium. (B), (E) cover slips coated with UC and seeded with 5000 3T3 Fibroblasts. (C), (F) AC coated cover slips seeded with 5000 3T3 fibroblasts. (A)-(C) Cultivated for 4 days and (D)-(F) for 8 days. All light microscopic images are shown with a magnification of 200-fold.

We could observe significant differences between NIH 3T3 fibroblasts seeded on the different collagen extracts during a time course of up to 8 days (Fig. 23). The NIH 3T3 fibroblasts seeded on AC show abnormal morphology after 4 days as compared to glass grown fibroblasts. They were packed closely together with globular form. After 8 days, NIH 3T3 fibroblasts showed mostly deformed morphology without significant amount of lamellipodia (Fig. 23C and F). In contrast to the AC grown cells, the UC grown NIH 3T3 fibroblasts were evenly distributed over the UC surface (Fig. 23B and E) with similar morphology as the control samples directly cultivated on the cover-slips (Fig. 23A and D). The cells grown on UC showed clearly defined lamellipodia throughout the 8 days and have a spindle like form which is characteristic for the fibroblasts in a 3D-matrix.

## Results



**Figure 23. SEM images of the 3T3-fibroblasts grown on different matrices during 8 days.** (A) and (D) are the control culture directly grown on the cover slips in DMEM for 4 and 8 days respectively. (B) and (E) are the cells grown on urea-extracted collagen for 4 and 8 days respectively.

### 3.12 Real time PCR and regulation at transcriptional level

To test if the difference in morphology observed in the differently prepared collagen matrices could also be manifested on a molecular level, we performed transcriptional analysis of genes implicated in the regulation of cell adhesion and cell motion, matrix remodelling as well as mechanical stress responses using qRT-PCR. The house-keeping gene glyceraldehyde-3-phosphate dehydrogenase was used as standard. The ratios of expression levels of the genes shown in Table 6 were calculated as ratio between UC/AC grown fibroblasts and cover-slip grown fibroblasts. Upregulation and downregulation were indicated with plus and minus respectively.

As the SEM images in Fig. 23 indicated that gene expression levels of typical marker genes indicating proliferation and motility as well as mechanical stress level are significantly different in UC or AC grown fibroblasts. The qRT-PCR results furthermore confirmed the observation that the cells in AC were more stressed than in UC, especially mechanical stress which is indicated by the 23 fold upregulation of  $TGF\beta_1$  in cells grown on AC. The MMP9 is strongly suppressed, up to 100-fold, in cells grown on AC as compared to UC and control cultures. In contrast to the AC grown cells, the collagen type I synthesis is 2.5-fold up regulated in UC grown cell indicating a higher level of matrix remodeling if compared to AC grown cells. In our study, up regulated Fak1 and consequently down regulated Rac1

## Results

confirmed the reduced matrix rigidity and mechanical stress level of the fibroblasts in UC than in AC.

	UC			UC <sub>mean±</sub> deviation	AC			AC <sub>mean±</sub> deviation
Fak1	10.84	7.1	10.64	9.53±1.6	-3	1	-3.45	-2.15±1.88
Rac1	-14	-3.84	-2	-6.61±4.9	-7.7	-3.45	-3	-4.72±1.99
Rho1	1.29	-2.63	1	-0.11±1.7	-3.33	-8.33	-1.6	-4.42±2.61
MMP9	37.64	42.47	-2	26.04±18.7	-263.16	-4.76	-100	-122.64±93.68
TGFb1	-5	5.09	-2	-0.64±3.8	1.51	23.42	1	8.64±9.85
CO1A1	1.18	1.18	5.27	2.54±1.8	-25	-1.2	1.44	-8.25±11.16
Actb	-5.26	-11.11	-10	-8.79±2.4	11.74	2.57	2.42	5.58±4.11

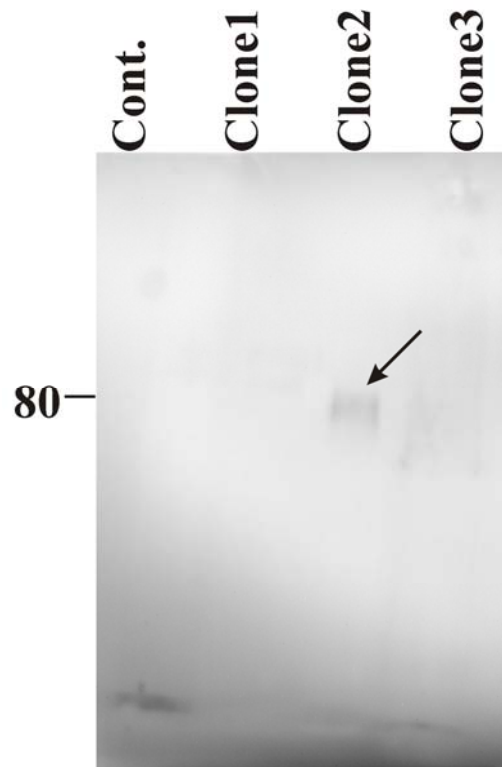
**Table 6. Comparison of fibroblasts grown on UC and AC using RT-PCR.** Values represent mean values from three different experiments (7 days cultivation) using different batches of collagen. Values are compared between UC/AC and glass surface.

### 3.13 Type I Collagen expression

For tissue engineering, animal derived collagen has still some shortages with regard to transmission of animal diseases. Therefore, recombinant human collagen may be a good choice in the future. For this purpose, cDNA clones described as full length cDNA clones of CO1A1 (IRATp970G0431D) and CO1A2 (IRATp970A0576D) were obtained from RZPD (present Ima Genes, Germany) and re-clones into mammalian expression vector to ensure the secretion with the specific signal peptide and ease the purification with the C-terminal His-Tag. Then the clones were transfected into CHO-K12 cells. Positive integrated clones were selected using zeocin and hygromycin respectively. The positive clones were cultivated for 14 days to produce the  $\alpha_1$  and  $\alpha_2$  chains. The clones were flanked by typical secretion-signal peptide (Ig K: METDTLLLWVLLLWVPGSTGD) so that the synthesized collagen chains could be directly secreted into the medium. The medium was separated from the cells by centrifugation (4000g for 15 min at 4 °C). The cell-free medium was then loaded onto the nickel-chelating resin column. The fully synthesized collagen peptides could bind to the nickel molecules through the C-terminal flanked His-tag. The peptides were then eluted with 250 mM imidazole solution. The isolated peptides were then detected using SDS-PAGE and western-blot. Unfortunately, only one clone showed a positive signal with strongly decreased mass of about 80 kDa, as indicated in Fig. 24. In the selected clones for collagen  $\alpha_2$  chains, no

## Results

signal was observed after the elution. Also the cell lysates tested using anti-His-tag antibodies showed no positive signals.



**Figure 24. Western-blot detected using anti-His-Tag antibodies for CO1A1-expression.** Collagen  $\alpha_1$  chain was expressed in CHO-K12 cell line. After purification using nickel-chelating column the isolated peptides containing a His-tag on the C-terminus were detected using western-blot and anti-His-Tag antibodies.

## Discussion

### 4. Discussion

The molecular mechanism for the biosynthetic assembly of collagen is still of great interest. Collagen is the major component in the extracellular matrix and Type I, II and III are the most abundant collagens that form fibrils responsible for tensile strength [3, 163]. Due to its high accessibility and compatibility, Type I collagen is one of the most well-studied collagens and is already widely used as a bioscaffold in medicine and cell biology [164-166]. Type I collagen is trimeric  $[(\alpha_1)_2\alpha_2]$  and exists as triple helix. The structure and assembly of the triple helix have been extensively studied for more than 40 years. The spontaneous formation of triple helices from isolated  $\alpha_1$  and  $\alpha_2$  polypeptides *in vitro* has been studied extensively [21, 51, 167, 168] and recent studies using synthetic model peptides [5, 169] show the self-association or self-assembly of these peptides to a native-like triple helix structure characteristically found in collagen fibrils. However, the studies using naturally derived collagens have been strongly hindered due to the high molecular weight of native collagen and limited understanding of the function of the collagen peptides, in particular the N- and C-terminal regions.

This study was originated by the necessity of the standardization of the acetic acid-extraction method for collagen used for tissue engineering. It turned out that it is comparably difficult to be reproduced due to extraction conditions such as pH, concentrations and significant extraction time, resulted in differently hydrolyzed products. Thus, a non-hydrolyzing method has been looked for since decades. Although the urea-extraction of collagen Type I was established more than 30 years ago [28, 92, 170] using 8-10 M urea, this method so far has been limited in its application, probably due to reports that only low quality collagen can be obtained with this procedure.

Although Adelman [92] observed that urea extraction led to an irreversible denaturation of collagen, this collagen extract has been used nevertheless for immunological studies. Becker [28, 171] and coworkers described an urea-extraction of collagen where the animals were made lathyrctic by feeding with  $\beta$ -aminopropionitrile three weeks before extraction. For further analysis the extracted collagen was then dialyzed and purified by ion-exchange chromatography. These studies also focused on the antigenicity of the collagen peptides without any consideration of the structure of urea-extracted collagen or its assembly. Similarly, Reuterberg [170] has used the same method to isolate collagen to study the cross-links in non-helical regions. All these groups have used animal skin to isolate collagen.

## Discussion

Other researchers have tried to extract procollagen Type I, II and III from skin with urea [93, 172, 173]. Extractions under these conditions were reported to result in large proportions of high molecular weight aggregates [93], corresponding to dimers or trimers of which only 5% could be converted to low molecular weight molecules. In all of these studies no reversible aggregation was reported. In contrast to these findings, Becker et al. [93] have shown that renaturation depends upon the formation of disulfide bridges.

In our study, we chose urea-extraction, since this method seems to imply to be less-hydrolyzing. We isolated Type I collagen from rat tail tendon using 9 M urea solution followed by a Superose 12 gel filtration and dialysis against water. In contrast to the above described studies, we did not observe any degradation and hydrolysis as observed with acetic acid-extraction or those reported by the other researchers who have used urea-extraction. Moreover, using urea-extraction, no proteinase inhibitor “cocktail” is needed and the whole process can be performed at room temperature without any cooling step as in other methods [91]. Additionally, in contrast to classical acetic or neutral solution extractions, the urea-extraction is considerably more effective as shown by its yield (Table 3) and in spite of the short incubation time of about 20 h, the purity and integrity of the collagen is much better than that obtained from acetic acid-extraction for an incubation time of around 14 days or more as shown in Fig. 8. Therefore, in this study, urea-extraction enhanced the reproducibility and quality of the collagen preparation.

As shown in Figs. 8 and 9, after the gel filtration, we obtained clear Coomassie-stained bands in SDS-gels without any smear, indicating a removal of substances during the gel filtration which probably enhance the gelling of collagen, such as glycosylated proteins [174-176]. The integrity and homogeneity of UC were further characterized using MALDI-TOF as discussed below.

In the last two decades, most analysis of Type I collagen have relied upon chromatography using CM-cellulose as column material following acetic acid-extraction and different cleavages using either CNBr or trypsin [100, 177, 178]. In our studies, we tried also to reproduce the well-accepted CM-cellulose profiles using our UC. However, we could not obtain this “classic” profile using UC. This suggests that UC is more stable against trypsin than AC.

In prior experiment using a PAS-staining, which is specific for sugars and simple to detect, we obtained a significant staining for urea-extracted collagen prior to gel filtration indicating presence of sugar probably glycosylation or proteoglycan. Whether these modifications are resistant against the urea-extraction procedure established in our study is still unclear.

## Discussion

Another technical advantage of UC is that the UC could be re-dissolved in aqueous solutions, thus making the storage and transport much easier than with AC, which is commonly kept in acetic solution and stored cold.

In our study, we found that the extracted collagen has unexpectedly reversible aggregational properties observable by SDS-PAGE (see Fig. 9A and C). This observation is in contrast to a previous report [92], which might be due to the purer urea used in this study [30]. Also re-chromatography of UC showed that the observed aggregation is reversible, with peaks observable at both higher and lower molecular weight fractions than present in the fractions selected from the first column (Fig. 9A and C). The appearance of lower molecular weights is not consistent with the increasing appearance of chemical crosslinks. In addition, the running pattern in gel was very complex and the aggregational profiles were not affected by re-chromatography as well as presence of reducing reagents. The latter observation suggests that formation of SDS-PAGE observable aggregation is not due to cross-linking or disulfide bridges. Furthermore, the MALDI-TOF data showed only a single mass by 96 kDa, indicating that every fraction from the gel filtration contains only collagen Type I and that the collagen is not cross-linked. A classical observation is that collagen migrates anomalously in polyacrylamide gels, resulting in an apparent mass increase over the predicted mass of about 20% [100]. This is supported by our SDS-PAGE analyses, where the  $\alpha_1$  and  $\alpha_2$  chains show molecular mass of 140 kDa and 130 kDa respectively. A more accurate determination of molecular mass is achieved by mass spectrometry using suitable calibrating proteins. Actually, the collagen mass from UC matched the theoretical database mass perfectly (Fig. 8A). The MALDI spectra of the fractions obtained after Superose 12 gel filtration confirmed the homogeneity of UC indicating that every fraction from the gel filtration contains only Type I collagen. Unfortunately, we could not determine whether the aggregates are dimers, heterotrimer/homotrimer or higher aggregates by either SDS-PAGE or MALDI-TOF spectra. As indicated in the SwissProt data, the  $\alpha_1$  and  $\alpha_2$  chains should have masses about 94.8 kDa and 91.7 kDa respectively. If we calculate the hydroxylation of proline residues detected in the ESI-MS/MS data, we obtain a 2.1 kDa- increase of the mass, which indicates that the  $\alpha_1$  and  $\alpha_2$  should now correspond to 96.9 kDa and 93.8 kDa respectively. These two masses now account well for those observed for UC experimentally. Furthermore, two peaks could be detected in UC samples which probably represent the  $\alpha_1$  and  $\alpha_2$  chains. In contrast to UC, in all measured AC samples only a single peak could be observed and always with a significantly lower mass (about 3 kDa) than the theoretical mass (Fig. 10B). Previous studies reported the degradation of collagen during acidic extraction [28, 30, 92], which could be an

## Discussion

explanation for the lowered masses for AC. Nevertheless, the measured masses contain some uncertainty because we could not determine whether these two peaks are really from  $\alpha_1$  and  $\alpha_2$  chains respectively, or whether the two chains overlapped each other and the second peak is from partial degraded  $\alpha$  chains. One means of answering this question is to use specific monoclonal antibodies against the NC-domains so that the two types of  $\alpha$  chain can be distinguished. However, so far no commercial antibodies designed to bind the NC-domains are available, and the antibodies against triple-helical region show high cross-reactivity, as the triple-helical regions share a high homology (see Appendix III). Regardless of this minimal dissatisfaction, the tandem MS analysis resulted in very high sequence coverage from the start of the N-terminus to end of the C-terminus of the mature collagen, so that the degradation, if it happens, could be neglected.

The results also indicate that the Superose 12 column effectively had removed other components which probably may enhance collagen gelling, such as glycosylated proteins [174-176]. The removal of these unknown components may lead to the high solubility of urea-extracted collagen in water. In contrast, acid-extracted rat tail tendon yields core collagen containing a variety of impurities [29]. Hence UC can be useful to study cell-collagen contacts and the effect of collagen in cell culture but without gelling effect and contraction characteristic of AC.

Interestingly, the UV-chromatogram at 280 nm and 256 nm showed significant absorption which normally is a sign for aromatic amino acids such as tryptophan and tyrosine. This in turn raised the question if any contamination or propeptides of collagen are present in the extraction. If the extract contains procollagen, especially the C-terminal propeptide which contains most of Trp, it may explain the aromatic absorption. Moreover, this suggests that we isolated or at least co-isolated procollagen during the urea-extraction, although it is generally accepted that mature tendon from rat older than 100 days contain no procollagen quantities. Furthermore, it has been reported that procollagen Type I forms irregular and irreversible aggregates in urea solution [92], which should precipitate..

Until now, the molecular assembly of the triple helix and its transport from cytosol into the ECM-region are still not fully understood. A well accepted theory is that the assembly of the triple helix begins in the ER by the assembly of C-terminal propeptides immediately after the full synthesis of the pro- $\alpha_1$  and pro- $\alpha_2$  chains [48]. Veis and coworkers have shown that the folding of propeptides to initiate the triple helix formation involve intrachain disulfide bonds e.g. in C-pro- $\alpha_1$ , residues C81-C244, and C152-C197, and in C-pro- $\alpha_2$ , C84-C245 and C153-C-198 form intrachain disulfide bonds which are important in proper folding of the individual



## Discussion

chain propeptide domains. Evidence for a transporter which is capable of transferring the 300 kDa collagen molecules to the ECM is still unknown. Although we added reductive reagents to the UC, we did not observe any change in the SDS-PAGE profiles, which suggests that also *in vivo* the assembly of mature collagen chains could be possible. Although we have not detected propeptides in the isolates. The reversible assembly of  $\alpha$ -chains into well-ordered triple-helical structure (Fig. 13A) indicates that the processed mature collagen  $\alpha_1$  and  $\alpha_2$  chains can assemble and fold to the characteristic triple helix in urea. This led to our hypothesis that the pro- $\alpha$  chains are not essential for folding, at least during dialysis of the UC in 8 M urea against water.

Another fundamental step which is not well understood is how the three chains recognize each other and initiate the folding. Many studies have shown that the C-NC domains in the pro- $\alpha_2$  play a key role through the disulfide bridges. However, the self-aggregation of UC to triple helices in water cannot be explained by these results. The studies from model peptides consisting only of the Gly-Pro-Hyp repeats also show self-assembly to triple helical structure. Generally accepted theory is that the hydroxylation can enhance triple-helical folding and stabilize the triple-helices. Thermodynamic studies by different groups have shown that this is fastest at a temperature just below the  $T_m$  (37 °C) [18]. If we assume that the processed chains are first transported into ECM before folding begins, these should be able to fold easily under body temperature and form triple-helices without specific order [12]. However, as the random helices encounter other molecules such as glycoproteins, which could lend negative charge to the C-terminal and positive charge to the N-terminal, thereby affecting the hydrophobic interaction, random triple helices may still generate well-ordered fibril forms as *in vivo* [12].

This self-aggregational phenomenon contributes not only to an understanding of collagen assembly but also yields new perspectives for protein design or protein engineering, in particular in generating materials with an improved stability and reversibility.

The ESI-MS approaches using fractions from the 3 sub-peaks, successfully covered 93% and 78% of the sequences of collagen Type I  $\alpha_1$  and  $\alpha_2$  respectively (Fig. 11A, B). No other proteins could be found in these samples, which again confirmed that UC contains exclusively collagen Type I  $\alpha_1$  and  $\alpha_2$  chains.

The high molecular weight bands (Fig. 9B) were also collagen Type I as shown by MS, containing no other high molecular weight contamination as reported previously by another group [30] or any other impurities. In contrast to the high purity of UC, we also have found Type II collagen, rat albumin and rat keratin with significant amount in AC which was shown by MS/MS with high score and significance (data not shown). From commercially purchased

## Discussion

AC-product we were not only able to observe degraded fragments but also a variety of contaminations using SDS-PAGE. For this reason, it is not surprising that cell culture experiments performed using these collagen products were not consistently reproducible. This may also be the reason that many researchers are troubled about the usage of animal derived collagen in human medicine approaches [96-98, 101].

A difficulty is the biochemical quantitation of the amounts of Pro and Hyp using amino acid composition analysis, whereby the collagen is hydrolyzed in 6N HCl for 24h at 106 °C as described previously [179]. A specific color indicator which binds to Hyp and is measured at 240 nm has been also proposed alternatively. In the present study, the MS/MS data showed that 92-120 proline residues from 236 Pro in  $\alpha_1$  chain and 70-90 Pro from 201 Pro in  $\alpha_2$  chain were detected as Hyp in our UC samples, which is far different from the sequence data in the database. Results from independent groups in the last decades also varied strongly.

It is known that Hyp is important for the stability of collagen molecules and is thus critical for the biological functions of collagen. Recent work has also proposed that Hyp is required for the triple helix self assembly [23]. Our MS/MS data allow us not only identify the total amount of hydroxylation but also indicate the possible sequence position of the Hyp, hence facilitating further detailed studies. To our knowledge, no one has applied MS/MS to study the hydroxylation of Pro residues.

Until now, the studies about the stability and folding of the collagen triple helices were mostly carried out using model peptides containing GPP repeats, which are strongly length limited. Moreover, these synthetic peptides consist of either non-interrupted GPP repeats or substituted at several positions by other amino acids. Although these model peptides have given insight into the collagen assembly and stabilization, they differ from real natural collagen molecules which consists at least three chains with more than 3000 amino acids in total. The natural interruptions in triple-helical regions have been long studied and it is accepted that these interruptions serve as binding sites for cells or collagen-cleaving enzymes like MMPs. Therefore, our MS/MS approach could be useful to study the cell-collagen contacts.

In general most collagen sequences especially Type XII- Type XXVIII have been generated and annotated from cDNA-clones. Even though collagen sequencing using tandem MS approach is still a challenge [148, 149] we were able to achieve a high sequence coverage, which could be probably explained relying on the improved collagen quality using urea-extraction. Mutations in collagen are widely correlated to different collagen related diseases. In our MS/MS data, many amino acids differed from the database-sequences suggesting new

## Discussion

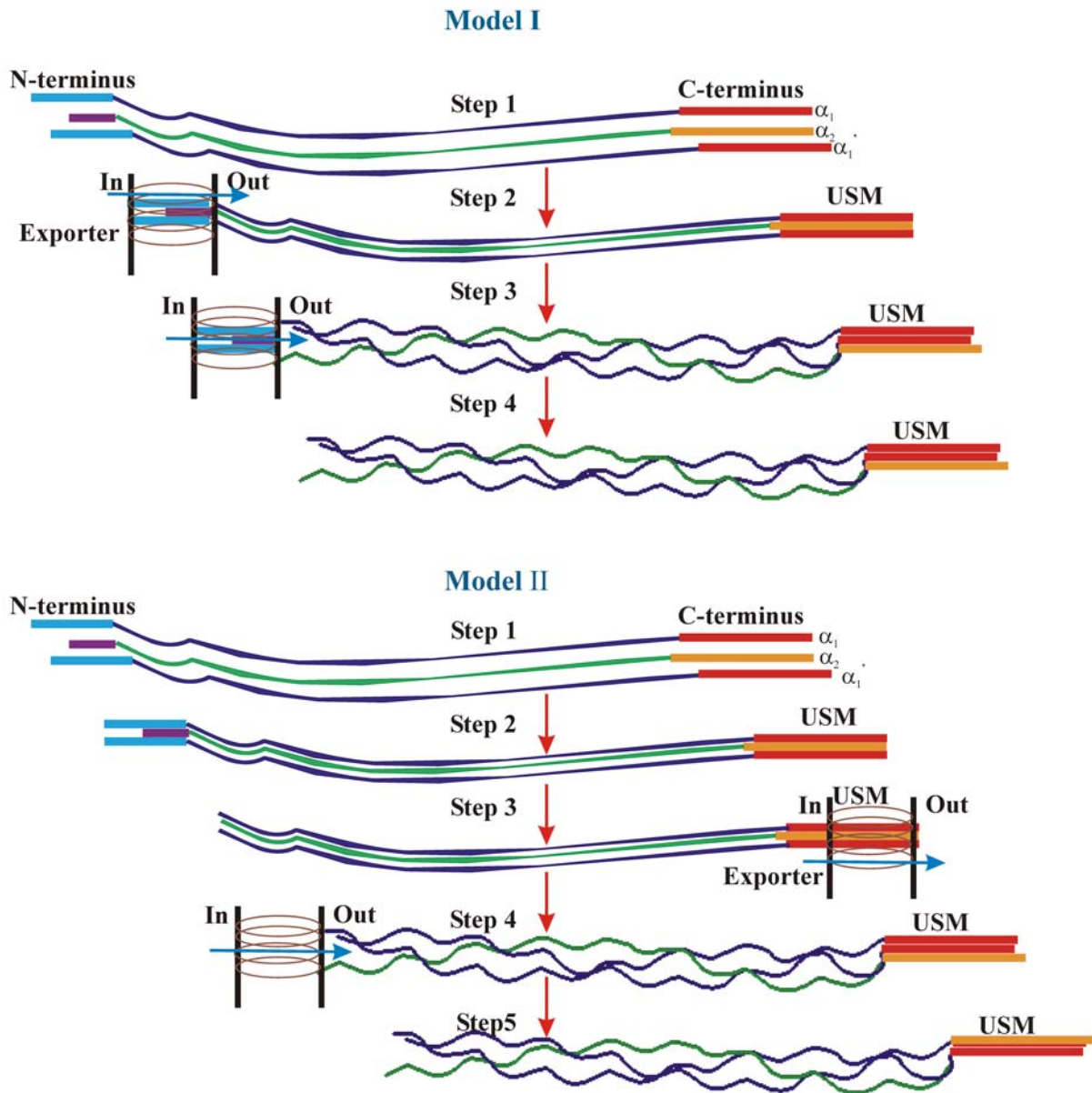
insight to study the polymorphisms in collagen and thus advancing the studies of collagen-related diseases. This promising approach is expedited by combining urea-extraction and MS/MS techniques.

Another new insight from these MS/MS data is the differential hydroxylation of Pro as shown in Table 4 and 5. For this reason we would like to hypothesize that the Type I collagen contains not only  $[\alpha_1]_2\alpha_2$  but differentially modified  $\alpha_1$  and  $\alpha_2$  chains. They might be differentially post-translationally modified so that they may play different roles during assembly, folding and cellular function. In addition, the two  $\alpha_1$  chains with different interruptions in the triple-helical regions may yield differentiated possibilities for cell binding and enzymatic cleavages.

To explain this strong structural stability and reversibility, I would like to hypothesize that in C-terminal region a highly stable peptide domain is present, which holds the three chains together strongly after the peptide docking. Hence this region might induce the following folding process. Urea might dissociate this specific region by altering its hydrophobic interaction between these three peptides. As soon as urea is removed by dialysis against water, these three peptides are able to reassociate.

Based upon our measurements and observations, we have developed two models to illustrate our hypothesis about the folding and export of Type I collagen (Fig. 25). In the first model, we assume that the single  $\alpha$ -chains are exported to the extracellular spaces separately and this process should be ensured by the recognition of the C-terminal sequences by the exporter on the cell membrane or bound directly through a ligand in the membrane. The second model retains the hypothesis derived from studies of model peptides and molecular dynamics [48] which indicate that export is controlled by C-terminal propeptides and their interchain disulfide bridges. Correct folding of three  $\alpha$ -chains into triple helices, yielding a mass of more than 300 kDa has been suggested to occur prior to export. After cleavage of the propeptides the mature collagen is exported by exocytosis. However no direct evidence has shown this transmembrane step. In our second model, only the C-terminal peptides from processed  $\alpha$ -chains dock to form tight triple-helical C-terminal structures (designated here as urea stable motive USM), ensuring the chain recognition. Then the C-terminally docked non structured chains are exported by an unknown mechanism through the membrane. After export, the chains are induced by the docked termini to fold into triple-helical structures. In our models, differential hydroxylation of Pro is thought to play an important role.

## Discussion



**Figure 25. Hypothesized models for folding and export of mature Type I collagen.** Urea-stable motive is indicated as USM.

After dialysis against water and lyophilization the UC could be redissolved in water to yield a clear sol-gel type solution. As indicated by UV-CD spectroscopy, the urea-extracted collagen dissolved in water, exhibited a triple helical structure. This measurement confirms the reversible aggregation of UC. The spectra are also comparable to those performed using synthetic model peptides [5] indicating the triple helices in our measurements are not miss-aligned peptides showing randomly triple helical similar structure.

In contrast to our procedure, collagen extracted using either neutral salt or acidic solution must be either lyophilized or stored cold in acetic acid [99] to avoid degradation.

This reversible aggregation of collagen is first observed by us in this study and the complex behavior of the absorption maxima during reversible aggregation suggests strongly that the

## Discussion

assembly process itself is complex, probably proceeding through several conformational/aggregational states. In fact, the UV-VIS data would be quite consistent with the general consensus in the literature [48] that the C-terminal regions of the mature collagen are responsible for catalyzing triple helix formation.

The most surprising result was that the reaction of UC samples with DTNB indicated that cysteine could be present. This was contradicted by the MS/MS analysis whereby no peptide containing cysteine was found, indicating that no propeptide sequences were detected in all MS/MS approaches. In all samples only mature Type I collagen was found and identified so that protein-contamination is excluded. Procollagen is known to be extracted by neutral salt solution so that if procollagen Type I is in the RTT, it should be removed at the first step wash using PBS solution. However, (Table 3) no proteins were found in the wash solution.

If we assume that the UC contains short C-terminal telopeptide sequences which can already associate and fold at a urea concentration of less than 4 M, then changes in turbidity might be explained by the following speculative model. The increase in turbidity between 4-8 M urea might correspond to the initial aggregation of C-termini without triple helix formation. Triple helix formation is possibly initiated between 2-4 M urea, thereby leading to more compact structures with a concomitant reduction in turbidity, and below 2 M urea, the triple helical structures associate to form larger structures, thereby causing an increase in turbidity to the final value observed in water. The corresponding changes in the putative absorption peaks due to tyrosine may reflect the different environments present during this aggregation process. However, by MS/MS data, we have no evidence that a C-terminal telopeptide is present.

A very interesting observation by the UV-spectra is weak aromatic absorption. In core collagen Trp is absent. However, a puzzling aspect of the aromatic absorption spectrum is that the characteristic peaks at about 275-280 nm and 260 nm, corresponding to Trp and Tyr appear to be very weak (see Fig. 14A). Closer examination of the spectra obtained in the presence of different concentrations of urea revealed that an additional optical density with strong wavelength dependence is superimposed upon that arising from the aromatic absorption. Interestingly, the presence of detectable absorption peaks varied with the urea concentration, with most distinct peaks being present under conditions where the “baseline” due to the wavelength-dependent optical density was highest (corresponding to 4M urea). This concentration corresponds to the random coil-triple helix transition as detected by UV-CD spectroscopy (Fig. 13A).

We surmise that the wavelength-dependent additional optical density arises primarily due fibril formation. To focus on the aromatic environment we performed a second-derivative

## Discussion

analysis of the spectra (Fig. 14D), which essentially removes the contributions due to turbidity. The plot of the UV maxima at 265 nm and 287 nm revealed that the environment(s) of both Tyr and Trp residues change in a complex manner with the varying urea concentration (Fig. 14B) and not systematically as one would expect from a simple two-state transition, which is typical for denaturation/renaturation. However, these could only be partially confirmed by MS/MS sequencing, that these aromatic peaks are from tyrosine and its ionized form, tyrosinate ( $5/\alpha_1$  and  $2/\alpha_2$ ).

Collagen concentration determination is still a problem; some groups determine the concentration by lyophilization and sample weighing, other groups use the chemical or immunological reactions, which are based upon either anti-collagen antibodies or chemical reaction followed with photometric detection [180]. In principle, the dry weight methods are the most accurate methods but time consuming; whereas photometric methods are complicated due to the chemical reaction, which need an optimum condition for detection; and antibody reaction is only accurate when purified monoclonal antibody are employed. Unfortunately, most commercial antibodies against collagen type I show high cross-reactivity (up to 10%) to other fibril collagens such as Type III and the use in immunoblotting is not accurate and also time consuming. In comparison to those methods, the His-determination may be a new fast assay system whereby the determination requires only a UV-spectrum for His, however, this methods is mostly suitable for pure collagen or collagen containing matrices.

Both AFM and SEM images showed that the formation of fibrils in UC is similar in the structure and the arrangement as for natural collagen, however the diameters of single fibrils from water-soluble urea-extracted collagen were larger than those of natural collagen (diameter < 100 nm) [161, 181]. The difference in fibril diameter is possibly attributable to the lack of additional components (bound carbohydrate, glycoproteins or lipid) present in the starting material and subsequently removed by gel filtration [21, 182] which may improve the packing of collagen triple helical units.

As described above, during the fractionation of the UC absorption due to aromatic groups was observed in chromatogram. In order to separate peptides containing aromatic amino acids, we cleaved the collagen using CNBr and trypsin respectively and separated these peptides by reverse-phase HPLC using  $C_{18}$  materials. Interestingly these procedures also yielded fractions showing aromatic absorption profiles (Fig. 19). The fraction showing aromatic absorption around 256-280 nm and peptide absorption at 215 nm were collected and analyzed by MALDI-TOF/TOF. However, none of these isolated peptides contained aromatic amino acids

## Discussion

or Tyr. To our surprise, every time the gradient was increased to 100% ACN and maintained for 30 min, we obtained a striking peak with very strong aromatic absorption and peptide absorption. Unfortunately, these fractions were not sequencable by neither ESI-MS/MS nor MALDI-TOF/TOF. Here I hypothesize that peptides containing Trp or Tyr are present in collagen preparation but with a very unusual structural effect so that these peptides were bound strongly to the C<sub>18</sub>-column. Even though, they were eluted from the column they would be again attached to the Nano-LC column in MS/MS step, leading to missed sequencing results. Another explanation would be the presence of non-protein component which have also aromatic absorption.

Structural integrity of UC and AC was compared by enzymatic digestion using carboxypeptidase Y and chymotrypsin. UC shows a high resistance against yeast carboxypeptidase Y (Fig. 15A) and chymotrypsin (Fig. 15B), which indicates the very stable triple helix is formed after the dialysis, so that the termini of UC is not susceptible to these proteases [159, 183]. Thus, whereas the dialyzed AC samples were almost completely digested by carboxypeptidase (Fig. 15A lane 16 and 17), the UC exhibited significant resistance against this protease. These results are also consistent with the UV-CD-spectra which revealed that UC in water exists as triple helix. Prior to this study carboxypeptidase was never applied to judge the structure of collagen triple helix.

Here, we surmise that the C-terminal region fold tightly and initiate the triple helix packing. This would be also explainable that this domain is very stable.

Another curious result is that the biotinylation of UC using a modified NEM-biotin molecule which binds tightly and irreversible to free or reduced Cys residues shows a clear signal as judged by Western-blot (Fig. 16). A weak signal could also be observed from AC sample. That means in the UC sample may really contain some Cys residues but due to an unknown structural effect, these were not able to be detected using MS. Alternatively, unknown PTMs might cross react with the NEM-biotin.

For biophysical and biochemical studies a rapid method for the determination of protein amount in collagen samples is very meaningful. In our study we have shown that the UV-spectroscopic determination of His-amount in collagen is very simple and fast as compared with other popular methods and this method shows a high accuracy as described above. So far, this method, first tested here, may be a very good choice for a cheap, fast and accurate alternative determination for collagen.

As reported previously, the cell-matrix interaction can control cellular phenotype. In our studies, fibroblasts cultivated on different collagen matrix showed significant differences both

## Discussion

at phenotypic and transcriptional levels. Images using both light microscope and SEM showed significant difference between the UC and AC. While the cells grown on UC has similar form and alignment as in control culture directly on plastic surface the cells on AC were totally deformed may be due to the shrinkage of the AC matrix during the course of cultivation for 14 days. The contraction may have an important impact on the cell growth as it may change the cellular signaling and responses [65]. As reported previously, in 3-D collagen lattices, migrating fibroblasts and tumor cells frequently develop a length axis of up to 100  $\mu\text{m}$  and the mean maximal diameter of the cell body may surpass 20  $\mu\text{m}$  [89]. This observation is in well correspondence to our observation using SEM during a time course for 8 days. This measurement was only found by cells grown on UC but not in AC. This bipolar or stellate morphology is characteristic of fibroblasts in 3D environments [83]. It is also known that normally bipolar or stellate shaped fibroblasts adhere to stiff surfaces and adopt dramatic spread morphology [184]. In our AC-culture, the cells have all deformed to a round morphology and packed closely together as shown in Fig. 23 C and F, most highly due to the mechanical stress originated from the contracted acetic collagen gel during the cultivation. Interestingly, in some epithelial cells this round morphology means the beginning of the transition from proliferation to apoptosis [185].

The cellular behavior was also compared using qRT-PCR. Several genes involved in regulation of cell motility, adhesion, matrix-remodeling and mechanical stress response were investigated.

MMP-9 is a gelatinase responsible for breaking down collagen degradation products [186]. The breakdown of collagen chains and the subsequent degradation of collagen breakdown products could affect many cellular behaviors, especially by matrix remodeling [187].

In correspondence to the matrix-remodeling the Type I collagen synthesis was significantly up regulated by UC but is strongly suppressed in AC. The regulation of matrix remodeling revealed that in the contracted AC-gel the cellular activity was significantly reduced with regards to matrix remodeling and collagen synthesis.

It is well known that ECM can regulate the cell-cell signaling by binding of cellular receptors like RANTES and growth factors such as TGF- $\beta_1$  [188]. Moreover the soluble elements in the ECM and liberated substances during matrix-remodeling can also signal the cells. TGF- $\beta$  has three isoforms in mammals: TGF- $\beta_1$  and  $\beta_2$  are described as being pro-fibrotic while the TGF- $\beta_3$  has been described being anti-fibrotic [189]. TGF- $\beta_1$  is generally inducer to ECM production however increased expression of TGF- $\beta_1$  is correlated to numerous diseases with increased fibrosis. TGF- $\beta_1$  is reported to stimulate fibroblasts contraction of ECM but



## Discussion

fibroblasts in a contracting matrix have been observed to undergo apoptosis. TGF- $\beta_1$  is a potent activator of fibroblasts stimulating fibroblast proliferation, production of extracellular matrix and differentiation into myofibroblasts. Because of these actions, TGF- $\beta_1$  driven fibroblast activation is believed to play a major role in wound repair, scar formation and tissue fibrosis. Interestingly in contrast to the monolayer culture, in 3D-culture TGF- $\beta_1$  induced apoptosis [189]. In our study, TGF- $\beta_1$  is strongly down-regulated by UC while it is up regulated in AC. This could indicate that the fibroblast rounded closely in the AC matrix may undergo apoptosis. If these cells are undergoing apoptosis should be determined more carefully through specific methods.

Another aspect in this study is to compare several genes involved in the regulation of cell motility, including Fak, Rho and Rac.

Cell motility is regulated by the chemical and physical properties of ECM. The surface density and distribution of ECM proteins can determine the speed and direction of cell migration. The topography and rigidity of ECM regulate cell migration through topographic guidance and durotaxis respectively. The mechanical forces acting on ECM affect the polarity of cells. The light-microscopic images showed a similar alignment of cells on cover-slip and UC pre-coated cover-slip which is not observed in any tested culture using AC. This observation could be explained by the regulation through the topographic property of the matrix. As assumed the UC in aqueous solutions exists as natural identical triple helix so that the structure is more similar to natural collagen than the AC so that the cells sense this structural alignment thus grow along the collagen fibrils. Moreover the cells have decreased motility if they are cultivated in a rigid ECM where the rigidity of ECM strengthens the linkage of integrins to cell cytoskeleton. In another word, the flexible matrix makes the cells easier to migrate and cells tend to move toward stiffer matrices due to the stronger traction force generated by this rigidity [57]. Fak is important to sense this rigidity and activates the downstream regulation pathways. In our experiments, Fak1 was strongly up regulated by UC, whereas in AC it was significantly down regulated. This fits the suggested theory very well. Furthermore in 3D-ECM the Rac pathway was required for invasive growth of the cells into ECM through the PI3K pathway. In our RT-PCR results, the Rac1 is stronger down regulated by AC than UC, which reveals a lower motility of cells in AC than in UC. Similarly a down regulation of Rho $\alpha$  also reveals the lowered motility.

All of these data indicate that urea-extracted collagen may have advantages to be applied as scaffold for tissue engineering and cell cultures. We expect that the urea-extracted collagen

## Discussion

preparation described here will be very useful for cell culture and tissue engineering techniques in the future.

In order to establish a simplified and more productive as well as easier to be standardized procedure for collagen preparation, we re-examined the urea-extraction with regard to elimination of hydrolysis during the extraction, which is usual by acetic acid-extraction. In the course of this study, we were fascinated by many new insights in aspects of both collagen structure and biological function. To our knowledge, several attempts in this study were first reported here and may contribute to understanding of the collagen self-assembly and export as well as its function in a fully new way. For example, we could show the PTMs in collagen and correlated the position of PTMs to the collagen assembly. Additionally, we developed two speculative models based upon our data, to explain how the collagen Type I could be easier exported and re-fold after the export.

Furthermore, we tried to approve the suitability of UC as a matrix for tissue engineering not only at phenotypical level but also at molecular biological level, so that the physiological impact on cell metabolism, morphology, apoptosis and differentiation/re-differentiation could be judged in detail. As shown in our data, we obtained preliminary data indicating that the UC may have advantages to be applied as a scaffold or a major component in a 3D-matrix for tissue engineering.

As mentioned above, despite of intensive studies of collagen, our understanding is still limited. Thus, we hope, this study may be helpful as an initiative contribution to this field.

## Reference

### REFERENCE

1. Exposito, J.Y., et al., *Evolution of collagens*. Anat Rec, 2002. **268**, 302-16.
2. Gelse, K., E. Poschl, and T. Aigner, *Collagens--structure, function, and biosynthesis*. Adv Drug Deliv Rev, 2003. **55**, 1531-46.
3. Kadler, K.E., et al., *Collagens at a glance*. J Cell Sci, 2007. **120**, 1955-8.
4. Tsuchiya, S., et al., *Collagen type I matrix affects molecular and cellular behavior of purified porcine dental follicle cells*. Cell Tissue Res, 2008. **331**, 447-59.
5. Kotch, F.W. and Raines, R.T., *Self-assembly of synthetic collagen triple helices*. Proc Natl Acad Sci U S A, 2006. **103**, 3028-3033.
6. Steffen, C., et al., *Demonstration of collagen in synovial fluid cells of rheumatoid arthritis by immunofluorescence*. Z Immunitätsforsch Exp Klin Immunol, 1972. **143**, 252-63.
7. Wick, G., et al., *The diagnostic application of specific antiprocollagen sera. II. Analysis of liver biopsies*. Int Arch Allergy Appl Immunol, 1978. **56**, 316-24.
8. Brunner, H., Dichtl, M. and Steffen, C., *Immunogenicity and specificity of chromium-stabilized calf-collage*. Z Immunitätsforsch Exp Klin Immunol, 1971. **141**, 460-70.
9. Steffen, C., et al., *Immunogenicity and specificity of collagen. XII. Demonstration by immunofluorescence and haemagglutination of antibodies with different specificity to human collagen*. Immunology, 1971. **21**, 649-57.
10. Myllyharju, J. and Kivirikko, K.I., *Collagens and collagen-related diseases*. Ann Med, 2001. **33**, 7-21.
11. Myllyharju, J. and Kivirikko, K.I., *Collagens, modifying enzymes and their mutations in humans, flies and worms*. Trends Genet, 2004. **20**, 33-43.
12. Brodsky, B., et al., *Triple-helical peptides: An approach to collagen conformation, stability, and self-association*. Biopolymers, 2008. **89**, 345-53.
13. Bruckner-Tuderman, L. and Bruckner, P., *Genetic diseases of the extracellular matrix: more than just connective tissue disorders*. J Mol Med, 1998. **76**, 226-37.
14. Prockop, D.J. and Kivirikko, K.I., *Collagens: molecular biology, diseases, and potentials for therapy*. Annu Rev Biochem, 1995. **64**, 403-34.
15. Prockop, D.J., et al., *The biosynthesis of collagen and its disorders (second of two parts)*. N Engl J Med, 1979. **301**(2): p. 77-85.
16. Ramachandran, G.N., *Structure of Collagen at the Molecular Level*. Treatise on Collagen. 1967, New York: Academic Press. p. 103-183.
17. Kadler, K.E., Hojima, Y. and Prockop, D.J., *Assembly of type I collagen fibrils de novo. Between 37 and 41 degrees C the process is limited by micro-unfolding of monomers*. J Biol Chem, 1988. **263**, 10517-10523.
18. Leikina, E., et al., *Type I collagen is thermally unstable at body temperature*. Proc Natl Acad Sci U S A, 2002. **99**, 1314-8.
19. Bretscher, L.E., et al., *Conformational stability of collagen relies on a stereoelectronic effect*. J Am Chem Soc, 2001. **123**, 777-8.
20. Brodsky, B., et al., *Collagen fibril structure in lamprey*. J Mol Biol, 1994. **243**, 38-47.
21. Brodsky, B. and Persikov, A.V., *Molecular structure of the collagen triple helix*. Adv Protein Chem, 2005. **70**, 301-39.
22. Malone, J.P., Alvares, K. and Veis, A., *Structure and assembly of the heterotrimeric and homotrimeric C-propeptides of Type I collagen: Significance of the 2(I) Chain*. Biochemistry, 2005. **44**, 15269-15279.
23. Mohs, A., et al., *Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline*. J Biol Chem, 2007. **282**, 29757-29765.

## Reference

24. Paramonov, S.E., Gauba, V. and Hartgerink, J.D., *Synthesis of collagen-like peptide polymers by native chemical ligation*. *Macromolecules*, 2005. **38**, 7555-7561.
25. Okuyama, K., et al., *Helical twists of collagen model peptides*. *Biopolymers*, 2006. **84**, 421-32.
26. Okuyama, K., et al., *Revision of collagen molecular structure*. *Biopolymers*, 2006. **84**, 181-91.
27. Orgel, J.P., et al., *The in situ supermolecular structure of type I collagen*. *Structure (Camb)*, 2001. **9**, 1061-9.
28. Becker, U. and Timpl, R., *Cyanogen bromide peptides of the rabbit collagen a1-chain*. *FEBS Lett*, 1972. **27**, 85-88.
29. Miller, E.J., Gay, S. & Leon, W.C., (1987) [1] The collagens: An overview and update. In *Methods in enzymology*, Academic Press, Paris 1987, p. 3-41.
30. Chandrakasan, G., Torchia, D.A. and Piez, K.A., *Preparation of intact monomeric collagen from rat tail tendon and skin and the structure of the nonhelical ends in solution*. *J Biol Chem*, 1976. **251**, 6062-7.
31. Rajan, N., et al., *Preparation of ready-to-use, storable and reconstituted type I collagen from rat tail tendon for tissue engineering applications*. *Nat. Protocols*, 2007. **1**, 2753-2758.
32. Brodsky, B., et al., *Variations in collagen fibril structure in tendons*. *Biopolymers*, 1982. **21**, 935-51.
33. Goodsell, D.S. *Collagen*. 2000 [cited; Available from: [http://www.pdb.org/pdb/static.do?p=education\\_discussion/molecule\\_of\\_the\\_month/pdb4\\_1.html](http://www.pdb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb4_1.html)].
34. Raspanti, M., Congiu, T. and Guizzardi, S., *Structural aspects of the extracellular matrix of the tendon: an atomic force and scanning electron microscopy study*. *Arch Histol Cytol*, 2002. **65**, 37-43.
35. Jenkins, C.L., et al., *Effect of 3-hydroxyproline residues on collagen stability*. *J Am Chem Soc*, 2003. **125**, 6422-7.
36. Jamall, I.S., Finelli V.N. and Que Hee, S.S., *A simple method to determine nanogram levels of 4-hydroxyproline in biological tissues*. *Anal Biochem*, 1981. **112**, 70-5.
37. Berg, R.A. and Prockop, D.J., *The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen*. *Biochem Biophys Res Commun*, 1973. **52**, 115-20.
38. Light, N.D., and Bailey, A.J., [16] *Covalent cross-links in collagen*. In *Methods in Enzymology*. 1982, Academic Press. p. 360-372.
39. Eyre, D., *New biomarkers of bone resorption*. *J Clin Endocrinol Metab*, 1992. **74**, 470A-470C.
40. Martin-de las Heras, S., Valenzuela, A. and Villanueva, E., *Deoxyypyridinoline crosslinks in human dentin and estimation of age*. *Int J Legal Med*, 1999. **112**, 222-6.
41. Acil, Y., et al., *Concentration of collagen cross-links in human dentin bears no relation to the individual age*. *Int J Legal Med*, 2002. **116**, 340-3.
42. Acil, Y., et al., *Detection of mature collagen in human dental enamel*. *Calcif Tissue Int*, 2005. **76**, 121-6.
43. Açil, Y. and Müller, P.K., *Rapid method for the isolation of the mature collagen cross-links, hydroxylysylpyridinoline and lysylpyridinoline*. *J Chromatogr A*, 1994. **664**, 183-188.
44. Eyre, D.R., Paz, M.A. and Gallop, P.M., *Cross-linking in collagen and elastin*. *Annu Rev Biochem*, 1984. **53**, 717-48.
45. Yamauchi, M., et al., *Collagen Structural Microheterogeneity and a Possible Role for Glycosylated Hydroxylysine in Type I Collagen*. *Proc Natl Acad Sci U S A*, 1982. **79**, 7684-7688.

## Reference

46. Kadler, K.E., et al., *Collagen fibril formation*. Biochem J, 1996. **316 ( Pt 1)**, 1-11.
47. Iozzo, R.V., *Matrix proteoglycans: from molecular design to cellular function*. Annu Rev Biochem, 1998. **67**, 609-52.
48. Khoshnoodi, J., et al., *Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers*. J Biol Chem, 2006. **281**, 38117-21.
49. Kuznetsova, N.V., McBride, D.J. and Leikin, S., *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, 2003. **331**, 191-200.
50. Koivu, J., *Identification of disulfide bonds in carboxy-terminal propeptides of human type I procollagen*. FEBS Lett, 1987. **212**, 229-32.
51. Bornstein, P., *The NH(2)-terminal propeptides of fibrillar collagens: highly conserved domains with poorly understood functions*. Matrix Biol, 2002. **21**, 217-26.
52. Egeblad, M. and Werb, Z., *New functions for the matrix metalloproteinases in cancer progression*. Nat Rev Cancer, 2002. **2**, 161-74.
53. Adams, J.C., *Cell-matrix contact structures*. Cell Mol Life Sci, 2001. **58**, 371-92.
54. Adams, J.C. and Watt, F.M., *Regulation of development and differentiation by the extracellular matrix*. Development, 1993. **117**, 1183-1198.
55. Di Lullo, G.A., et al., *Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human Type I collagen*. J Biol Chem, 2002. **277**, 4223-4231.
56. Cowin, S.C., *Tissue growth and remodeling*. Annu Rev Biomed Eng, 2004. **6**, 77-107.
57. Li, S., Guan, J.L. and Chien, S., *Biochemistry and biomechanics of cell motility*. Annu Rev Biomed Eng, 2005. **7**, 105-50.
58. Byers, P.H., *Brittle bones--fragile molecules: disorders of collagen gene structure and expression*. Trends Genet, 1990. **6**, 293-300.
59. Schnieke, A., Harbers, K. and Jaenisch, R., *Embryonic lethal mutation in mice induced by retrovirus insertion into the alpha 1(I) collagen gene*. Nature, 1983. **304**, 315-20.
60. Saga, Y., et al., *Mice develop normally without tenascin*. Genes Dev, 1992. **6**, 1821-31.
61. Griffith, L.G. and Naughton, G., *Tissue engineering--current challenges and expanding opportunities*. Science, 2002. **295**, 1009-14.
62. Eaglstein, W.H. and Falanga, V., *Tissue engineering and the development of Apligraf a human skin equivalent*. Adv Wound Care, 1998. **11**, 1-8.
63. Eaglstein, W.H. and Falanga, V., *Tissue engineering for skin: an update* J Am Acad Dermatol, 1998. **39**, 1007-10.
64. Parenteau, N.L. and Hardin-Young, J., *The use of cells in reparative medicine*. Ann N Y Acad Sci, 2002. **961**, 27-39.
65. Vogel, V. and Baneyx, G., *The tissue engineering puzzle: a molecular perspective*. Annu Rev Biomed Eng, 2003. **5**, 441-63.
66. Elias, K.L., Price, R.L. and Webster, T.J., *Enhanced functions of osteoblasts on nanometer diameter carbon fibers*. Biomaterials, 2002. **23**, 3279-87.
67. Webster, T.J., et al., *Enhanced functions of osteoblasts on nanophase ceramics*. Biomaterials, 2000. **21**, 1803-10.
68. Webster, T.J., et al., *Mechanisms of enhanced osteoblast adhesion on nanophase alumina involve vitronectin*. Tissue engineering, 2001. **7**, 291-301.
69. Koo, L.Y., et al., *Co-regulation of cell adhesion by nanoscale RGD organization and mechanical stimulus*. J Cell Sci, 2002. **115**, 1423-33.
70. Maheshwari, G., et al., *Cell adhesion and motility depend on nanoscale RGD clustering*. J Cell Sci, 2000. **113**, 1677-86.

## Reference

71. Hirschi, K.K., et al., *Vascular assembly in natural and engineered tissues*. Ann N Y Acad Sci, 2002. **961**, 223-42.
72. Kellner, K., et al., *Determination of oxygen gradients in engineered tissue using a fluorescent sensor*. Biotechnol Bioeng, 2002. **80**, 73-83.
73. Roy, P., et al., *Effect of flow on the detoxification function of rat hepatocytes in a bioartificial liver reactor*. Cell Transplant, 2001. **10**, 609-14.
74. Cukierman, E., et al., *Taking cell-matrix adhesions to the third dimension*. Science, 2001. **294**, 1708-12.
75. Corbett, S.A., Wilson, C.L. and Schwarzbauer, J.E., *Changes in cell spreading and cytoskeletal organization are induced by adhesion to a fibronectin-fibrin matrix*. Blood, 1996. **88**, 158-66.
76. Sechler, J.L., et al., *Modulation of cell-extracellular matrix interactions*. Ann N Y Acad Sci, 1998. **857**, 143-54.
77. Sosnik, A. and Sefton, M.V., *Semi-synthetic collagen/poloxamine matrices for tissue engineering*. Biomaterials, 2005, **26**, 7425-35.
78. Seliktar, D., et al., *Dynamic mechanical conditioning of collagen-gel blood vessel constructs induces remodeling in vitro*. Ann Biomed Eng, 2000. **28**, 351-62.
79. Feng, Z., et al., *Investigation on the mechanical properties of contracted collagen gels as a scaffold for tissue engineering*. Artif Organs, 2003. **27**, 84-91.
80. van der Flier, A. and Sonnenberg, A., *Function and interactions of integrins*. Cell and tissue research, 2001. **305**, 285-98.
81. Wozniak, M.A., et al., *Focal adhesion regulation of cell behavior*. Biochim Biophys Acta, 2004. **1692**, 103-19.
82. Cukierman, E., Pankov, R. and Yamada, K.M., *Cell interactions with three-dimensional matrices*. Curr Opin Cell Biol, 2002. **14**, 633-9.
83. Larsen, M., et al., *The matrix reorganized: extracellular matrix remodeling and integrin signaling*. Curr Opin Cell Biol, 2006. **18**, 463-71.
84. Friedl, P., Zanker, K.S. and Brocker, E.B., *Cell migration strategies in 3-D extracellular matrix: differences in morphology, cell matrix interactions, and integrin function*. Microsc Res Tech, 1998. **43**, 369-78.
85. Heino, J., *Biology of tumor cell invasion: interplay of cell adhesion and matrix degradation*. Int J Cancer, 1996. **65**, 717-22.
86. Aggeler, J., Frisch, S.M. and Werb, Z., *Changes in cell shape correlate with collagenase gene expression in rabbit synovial fibroblasts*. J Cell Biol, 1984. **98**, 1662-71.
87. Akiyama, S.K., Olden, K. and Yamada, K.M., *Fibronectin and integrins in invasion and metastasis*. Cancer Metastasis Rev, 1995. **14**, 173-89.
88. Langholz, O., et al., *Collagen and collagenase gene expression in three-dimensional collagen lattices are differentially regulated by alpha 1 beta 1 and alpha 2 beta 1 integrins*. J Cell Biol, 1995. **131**, 1903-15.
89. Heath, J.P. and Peachey, L.D., *Morphology of fibroblasts in collagen gels: a study using 400 keV electron microscopy and computer graphics*. Cell Motil Cytoskeleton, 1989. **14**, 382-92.
90. Venturoni, M., et al., *Investigations into the polymorphism of rat tail tendon fibrils using atomic force microscopy*. Biochem Biophys Res Commun, 2003. **303**, 508-13.
91. Miller, E.J., and Kent Rhodes, R., [2] *Preparation and characterization of the different types of collagen*, in *Methods in Enzymology*. 1982, Academic Press. p. 33-64.
92. Adelmann, B.C., *The structural basis of cell-mediated immunological reactions of collagen. Reactivity of separated -chains of calf and rat collagen in cutaneous delayed hypersensitivity reactions*. Immunology, 1972. **23**, 739-48.

## Reference

93. Becker, U. and Timpl, R., *NH<sub>2</sub>-terminal extensions on skin collagen from sheep with a genetic defect in conversion of procollagen into collagen*. *Biochemistry*, 1976. **15**, 2853-62.
94. Silver, F.H., Freeman, J.W. and Sehra, G.P., *Collagen self-assembly and the development of tendon mechanical properties*. *J Biomech*, 2003. **36**, 1529-53.
95. Lee, C.H., Singla, A. and Lee, Y., *Biomedical applications of collagen*. *Int J Pharm*, 2001. **221**, 1-22.
96. Olsen, D., et al., *Recombinant collagen and gelatin for drug delivery*. *Adv Drug Deliv Rev*, 2003. **55**, 1547-67.
97. Baez, J., Olsen, D. and Polarek, J.W., *Recombinant microbial systems for the production of human collagen and gelatin*. *Appl Microbiol Biotechnol*, 2005. **69**, 245-52.
98. Olsen, D., et al., *Expression and characterization of a low molecular weight recombinant human gelatin: development of a substitute for animal-derived gelatin with superior features*. *Protein Expr Purif*, 2005. **40**, 346-57.
99. Kessler, A., Rosen, H. and Levenson, S.M., *Chromatographic fractionation of acetic acid-solubilized rat tail tendon collagen*. *J Biol Chem*, 1960. **235**, 989-994.
100. Deyl, Z., Miksik, I. and Eckhardt, A., *Preparative procedures and purity assessment of collagen proteins*. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2003. **790**, 245-75.
101. Heath, C.A., et al., *Dura mater-associated Creutzfeldt-Jakob disease: experience from surveillance in the UK*. *J Neurol Neurosurg Psychiatry*, 2006. **77**, 880-2.
102. Buechter, D.D., et al., *Co-translational incorporation of trans-4-hydroxyproline into recombinant proteins in bacteria*. *J Biol Chem*, 2003. **278**, 645-50.
103. Toman, P.D., et al., *Production of recombinant human type I procollagen trimers using a four-gene expression system in the yeast *Saccharomyces cerevisiae**. *J Biol Chem*, 2000. **275**, 23303-9.
104. Tomita, M., et al., *Transgenic silkworms produce recombinant human type III procollagen in cocoons*. *Nat Biotechnol*, 2003. **21**, 52-6.
105. Lamberg, A., et al., *Characterization of human type III collagen expressed in a baculovirus system. Production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit*. *J Biol Chem*, 1996. **271**, 11988-95.
106. John, D.C., et al., *Expression of an engineered form of recombinant procollagen in mouse milk*. *Nat Biotechnol*, 1999. **17**, 385-9.
107. Fertala, A., et al., *Synthesis of recombinant human procollagen II in a stably transfected tumour cell line (HT1080)*. *Biochem J*, 1994. **298 ( Pt 1)**, 31-7.
108. Fichard, A., et al., *Human recombinant alpha1(V) collagen chain. Homotrimeric assembly and subsequent processing*. *J Biol Chem*, 1997. **272**, 30083-7.
109. Ruszczak, Z. and Friess, W., *Collagen as a carrier for on-site delivery of antibacterial drugs*. *Adv Drug Deliv Rev*, 2003. **55**, 1679-98.
110. Weinberg, C.B. and Bell, E., *A blood vessel model constructed from collagen and cultured vascular cells*. *Science*, 1986. **231**, 397-400.
111. Ratner, B.D. and Bryant, S.J., *Biomaterials: where we have been and where we are going*. *Annu Rev Biomed Eng*, 2004. **6**, 41-75.
112. Goissis, G., et al., *Biocompatibility studies of anionic collagen membranes with different degree of glutaraldehyde cross-linking*. *Biomaterials*, 1999. **20**, 27-34.
113. Chvapil, M., Kronenthal, L. and Van Winkle, W. Jr., *Medical and surgical applications of collagen*. *Int Rev Connect Tissue Res*, 1973. **6**, 1-61.

## Reference

114. Dieterich, C., et al., *In vitro reconstructed human epithelia reveal contributions of Candida albicans EFG1 and CPH1 to adhesion and invasion*. Microbiology, 2002. **148**, 497-506.
115. Mukhopadhyay, A., Madhusudhan, T. and Kumar, R., *Hematopoietic stem cells: clinical requirements and developments in ex-vivo culture*. Adv Biochem Eng Biotechnol, 2004. **86**, 215-53.
116. Matthiesen, R. and Mutenda, K.E., *Introduction to proteomics*. Methods Mol Biol, 2007. **367**, 1-35.
117. Marcotte, E.M., *How do shotgun proteomics algorithms identify proteins?* Nat Biotechnol, 2007. **25**, 755-7.
118. Nesvizhskii, A.I., Vitek, O. and Aebersold, R., *Analysis and validation of proteomic data generated by tandem mass spectrometry*. Nat Methods, 2007. **4**, 787-97.
119. Cohen, S.L. and Chait, B.T., *Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins*. Anal Chem, 1996. **68**, 31-7.
120. Kussmann, M., et al., *Matrix-assisted laser desorption/ionization mass spectrometric peptide mapping of the neural cell adhesion protein neurolin purified by sodium dodecyl sulfate polyacrylamide gel electrophoresis or acidic precipitation*. J Mass Spectrom, 1997. **32**, 483-93.
121. Patterson, S.D. and Aebersold, R., *Mass spectrometric approaches for the identification of gel-separated proteins*. Electrophoresis, 1995. **16**, 1791-814.
122. Mirza, U.A. and Chait, B.T., *Effects of anions on the positive ion electrospray ionization mass spectra of peptides and proteins*. Anal Chem, 1994. **66**, 2898-904.
123. Kocher, T., Allmaier, G. and Wilm, M., *Nanoelectrospray-based detection and sequencing of substoichiometric amounts of phosphopeptides in complex mixtures*. J Mass Spectrom, 2003. **38**, 131-7.
124. Choudhary, G., et al., *Use of on-line mass spectrometric detection in capillary electrochromatography*. J Chromatogr, 2000. **887**, 85-101.
125. Wilm, M. and Mann, M., *Analytical properties of the nanoelectrospray ion source*. Anal Chem, 1996. **68**, 1-8.
126. Roepstorff, P. and Fohlman, J., *Proposal for a common nomenclature for sequence ions in mass spectra of peptides*. Biomed Mass Spectrom, 1984. **11**, 601.
127. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry*. Rapid Commun Mass Spectrom, 2003. **17**, 2337-42.
128. Taylor, J.A. and Johnson, R.S., *Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry*. Anal Chem, 2001. **73**, 2594-604.
129. Frank, A. and Pevzner, P., *PepNovo: de novo peptide sequencing via probabilistic network modeling*. Anal Chem, 2005. **77**, 964-73.
130. Sadygov, R.G. and Yates, J.R., *A hypergeometric probability model for protein identification and validation using tandem mass spectral data and protein Sequence databases*. Anal Chem, 2003. **75**, 3792-3798.
131. Field, H.I., Fenyö, D. and Beavis, R.C., *RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database*. Proteomics, 2002. **2**, 36-47.
132. Clauser, K.R., Baker, P. and Burlingame, A.L., *Role of accurate mass measurement (+/-10 ppm) in protein Identification strategies employing MS or MS/MS and database searching*. Anal Chem, 1999. **71**, 2871-2882.
133. Eng, J.K., McCormack, A.L. and Yates, J.R. III, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database*. J Am Soc Mass Spectrom, 1994. **5**, 976-989.
134. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**, 3551-67.



## Reference

135. Geer, L.Y., et al., *Open mass spectrometry search algorithm*. J Proteome Res, 2004. **3**, 958-64.
136. Zhang, N., Aebersold, R. and Schwikowski, B., *ProBiD: A probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data*. Proteomics, 2002. **2**, 1406-1412.
137. Allet, N., et al., *In vitro and in silico processes to identify differentially expressed proteins*. Proteomics, 2004. **4**, 2333-2351.
138. Craig, R. and Beavis, R.C., *TANDEM: matching proteins with tandem mass spectra*. Bioinformatics, 2004. **20**, 1466-1467.
139. Salek, M., et al., *Sequence tag scanning: A new explorative strategy for recognition of unexpected protein alterations by nanoelectrospray ionization-tandem mass spectrometry*. Proteomics, 2005. **5**, 667-674.
140. Mann, M. and Wilm, M., *Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags*. Analytical chemistry, 1994. **66**, 4390 - 4399.
141. Tabb, D.L., Saraf, A. and Yates, J.R., *GutenTag: High-throughput sequence tagging via an empirically derived fragmentation model*. Anal Chem, 2003. **75**, 6415-6421.
142. Frank, A., et al., *Peptide Sequence Tags for Fast Database Search in Mass-Spectrometry*. J. Proteome Res., 2005. **4**, 1287-1295.
143. Xu, C. and Ma, B., *Software for computational peptide identification from MS-MS data*. Drug Discov Today, 2006. **11**, 595-600.
144. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry*. Anal Chem, 2003. **75**, 4646-4658.
145. Miksik, I., et al., *Capillary electromigration methods for the study of collagen*. J Chromatogr B Analyt Technol Biomed Life Sci, 2006. **841**, 3-13.
146. Chung, E., Keele, E.M. and Miller, E.J., *Isolation and characterization of the cyanogen bromide peptides from the alpha 1(3) chain of human collagen*. Biochemistry, 1974. **13**, 3459-64.
147. Chung, E. and Miller, E.J., *Collagen polymorphism: characterization of molecules with the chain composition [ $\alpha_1$ (III)]<sub>3</sub> in human tissues*. Science, 1974. **183**, 1200-1.
148. Gevaert, K., et al., *Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry*. Electrophoresis, 2001. **22**, 1645-1651.
149. Dreisewerd, K., et al., *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, 2004. **76**, 3482-91.
150. McCullagh, J.S.O., Juchelka, D. and Hedges, R.E.M., *Analysis of amino acid <sup>13</sup>C abundance from human and faunal bone collagen using liquid chromatography/isotope ratio mass spectrometry*. Rapid Commun Mass Spectrom, 2006. **20**, 2761-2768.
151. Ocaña, F.M., et al., *C-terminal sequencing by mass spectrometry: Application to gelatine-derived proline-rich peptides*. Proteomics, 2005. **5**, 1209-1216.
152. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4*. Nature, 1970. **227**, 680-5.
153. Ellman, G.L., *Tissue sulfhydryl groups*. Arch Biochem Biophys, 1959. **82**, 70-7.
154. Riddles, P.W., Blakeley, R.L. and Zerner, B., *Reassessment of Ellman's reagent*. In Methods in enzymology, 1983. **91**, 49-60.
155. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels*. Anal Chem, 1996. **68**, 850-8.
156. Habermehl, J., et al., *Preparation of ready-to-use, stockable and reconstituted collagen*. Macromol Biosci, 2005. **5**, 821-828.

## Reference

157. Atkins P., de Paula, J., *Atkins' Physical Chemistry*. 6 ed. 2002: Oxford University Press.: p. 721-5.
158. Brinker C.J. & W. SG (1990) *SOL-GEL Science: The physics and chemistry of Sol-Gel Processing*, Academic press, Boston 1990, p.1-13.
159. Bruckner, P. and Prockop, D.J., *Proteolytic enzymes as probes for the triple-helical conformation of procollagen* *Anal Biochem*, 1981. **110**, 360-368.
160. Varani, J., et al., *Inhibition of type I procollagen synthesis by damaged collagen in photoaged skin and by collagenase-degraded collagen in vitro*. *Am J Pathol*, 2001. **158**, 931-42.
161. Raspanti, M., et al., *Collagen fibril surface: TMAFM, FEG-SEM and freeze-etching observations*. *Microsc Res Tech*, 1996. **35**, 87-93.
162. Canty, E.G., et al., *Actin filaments are required for fibropositor-mediated collagen fibril alignment in tendon*. *J Biol Chem*, 2006. **281**, 38592-8.
163. Rosso, F., et al., *From cell-ECM interactions to tissue engineering*. *J Cell Physiol*, 2004. **199**, 174-80.
164. Berglund, J.D., et al., *A biological hybrid model for collagen-based tissue engineered vascular constructs*. *Biomaterials*, 2003. **24**, 1241-1254.
165. Purna, S.K. and Mary, B., *Collagen based dressings — a review*. *Burns*, 2000. **26**, 54-62.
166. Wallace, D.G. and Rosenblatt, J., *Collagen gel systems for sustained delivery and tissue engineering*. *Adv Drug Deliv Rev*, 2003. **55**, 1631-49.
167. Olsen, B.R., et al., *Purification and characterization of a peptide from the carboxy-terminal region of chick tendon procollagen type I*. *Biochemistry*, 1977. **16**, 3030-3036.
168. Ottani, V., Raspanti, M. and Ruggeri, A., *Collagen structure and functional implications*. *Micron*, 2001. **32**, 251-60.
169. Kar, K., et al., *Self-association of collagen triple helix peptides into higher order structures*. *J Biol Chem*, 2006. **281**, 33283-33290.
170. Rauterberg, J., et al., *The amino acid sequence of the carboxyterminal nonhelical cross link region of the alpha 1 chain of calf skin collagen*. *FEBS Lett*, 1972. **21**, 75-79.
171. Becker, U., Timpl, R. and Kuhn, K., *Carboxyterminal antigenic determinants of collagen from calf skin. Localization within discrete regions of the nonhelical sequence*. *Eur J Biochem*, 1972. **28**, 221-31.
172. Fessler, L.I. and Fessler, J.H., *Characterization of type III procollagen from chick embryo blood vessels*. *J Biol Chem*, 1979. **254**, 233-239.
173. Uitto, J., Allan, R.E. and Polak, K.L., *Conversion of type II procollagen to collagen. Extracellular removal of the amino-terminal and carboxy-terminal extensions without a preferential sequence*. *Eur J Biochem*, 1979. **99**, 97-103.
174. Bann, J.G., Bachinger, H.P. and Peyton, D.H., *Role of carbohydrate in stabilizing the triple-helix in a model for a deep-sea hydrothermal vent worm collagen*. *Biochemistry*, 2003. **42**, 4042-8.
175. Mizuno, K., et al., *Hydroxylation-induced stabilization of the collagen triple helix: Acetyl-(glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyprolyl)<sub>10</sub>-NH<sub>2</sub> forms a highly stable triple helix*. *J Biol Chem*, 2004. **279**, 38072-38078.
176. Parodi, A.J., *Protein glucosylation and its role in protein folding*. *Annu Rev Biochem*, 2000. **69**, 69-93.
177. Deyl, Z. and Miksik, I., *Advanced separation methods for collagen parent alpha-chains, their polymers and fragments*. *J Chromatogr B Biomed Sci Appl*, 2000. **739**, 3-31.

## Reference

178. Deyl, Z. and Miksik, I., *Comparison of different electrokinetic separation modes applicable to a model peptide mixture (collagen type I and III CNBr fragments)*. J Chromatogr B Biomed Sci Appl, 2000. **745**, 251-60.
179. Piez, K.A., *The amino acid chemistry of some calcified tissues*. Ann N Y Acad Sci, 1963. **109**, 256-68.
180. Shormanov, V.K. and Bulatnikov, G.G., *Photometric determination of collagen*. J Anal Chem, 2006. **61**, 320-324.
181. Raspanti, M., Congiu, T. and Guizzardi, S., *Tapping-mode atomic force microscopy in fluid of hydrated extracellular matrix*. Matrix Biol, 2001. **20**, 601-4.
182. Brodsky, B. and Eikenberry, E.F., *Characterization of fibrous forms of collagen*. in Methods in enzymology, 1982. **82 Pt A**, p. 127-74.
183. Layman, D.L., McGoodwin, E.B. and Martin, G.R., *The nature of the collagen synthesized by cultured human fibroblasts*. Proc Natl Acad Sci U S A, 1971. **68**, 454-458.
184. Beningo, K.A., Dembo, M. and Wang, Y.L., *Responses of fibroblasts to anchorage of dorsal extracellular matrix receptors*. Proc Natl Acad Sci U S A, 2004. **101**, 18024-9.
185. Fujisaki, H. and Hattori, S., *Keratinocyte apoptosis on type I collagen gel caused by lack of laminin 5/10/11 deposition and Akt signaling*. Exp Cell Res, 2002. **280**, 255-69.
186. Galardy, R.E., et al., *Low molecular weight inhibitors in corneal ulceration*. Ann N Y Acad Sci, 1994. **732**, 315-23.
187. Phillips, J.A., Vacanti, C.A. and Bonassar, L.J., *Fibroblasts regulate contractile force independent of MMP activity in 3D-collagen*. Biochem Biophys Res Commun, 2003. **312**, 725-32.
188. Choe, M.M., Sporn, P.H. and Swartz, M.A., *Extracellular matrix remodeling by dynamic strain in a three-dimensional tissue-engineered human airway wall model*. Am J Respir Cell Mol Biol, 2006. **35**, 306-13.
189. Wang, Z., et al., *Increased transcriptional response to mechanical strain in keloid fibroblasts due to increased focal adhesion complex formation*. J Cell Physiol, 2006. **206**, 510-7.

## Appendix I: Figure index

### Figure index

Figure 1. Structure model of collagen Type I and SEM of RTT collagen Type I fibrils. ....	6
Figure 2. Cross-linking of hydroxylysine residues and formation of collagen fibril [44]. ....	7
Figure 3. Protomer assembly of collagen I.....	9
Figure 4. SEM image of a motile fibroblast in three-dimensional matrix.....	11
Figure 5. Schematic diagram showing the cell-ECM adhesion-mediated signaling in cell migration [57]. ....	13
Figure 6. Different cell morphology whether cultivated on 2D- or 3D matrices. ....	16
Figure 7. Construction of the expression vector inserted with Col 1 $\alpha$ 1 and Col1 $\alpha$ 2 full length DNA .....	39
Figure 8. Urea extraction of collagen and acetic degradation.....	41
Figure 9. Superose 12 chromatography and SDS-PAGE analysis of urea-extracted collagen....	43
Figure 10. (A) and (B). Mass determination and tandem MS. ....	44
Figure 11. (A) and (B). Tandem MS results. ....	47
Figure 12. (A) and (B) Demonstration of the fragment spectra. ....	48
Figure 13. UV-CD spectra obtained from the Superose 12-purified collagen polypeptides dissolved in different concentrations of urea. ....	51
Figure 14. UV-spectral analysis of Superose 12-purified collagen dissolved in various concentrations of urea.....	53
Figure 15. SDS-PAGE (6%) profile of the carboxypeptidase digestion and chymotrypsin limited digestion.....	54
Figure 16. Biotinylation of UC and AC. ....	55
Figure 17. Comparison of AFM images of PBS washed rat tail tendon and dialyzed Superose 12 column fraction of UC.....	56
Figure 18. SEM Images different protein samples. ....	57
Figure 19. RP-HPLC of tryptic digested UC samples.....	59
Figure 20. Microscopic images of NIH 3T3 fibroblasts cultivated on the UC and AC coated glass slides.....	60
Figure 21. Light microscopy of fibroblasts cultivated in different matrix.....	61
Figure 22. 3T3 Fibroblasts cultivated on the cover slips coated with different matrices. ....	62
Figure 23. SEM images of the 3T3-fibroblasts grown on different matrices during 8 days.....	63
Figure 24. Western-blot detected using anti-His-Tag antibodies for CO1A1-expression.....	65
Figure 25. Hypothesized models for folding and export of mature Type I collagen. ....	73

## Appendix II: Table index

### Table index

Table 1. SDS-PAGE preparation. ....	27
Table 2. Primer sets and UPL probes combinations for the RT-PCR.....	37
Table 3. Yield of collagen by urea-extraction. ....	40
Table 4. Differential hydroxylation of proline and lysine residues in $\alpha 1$ chain. ....	49
Table 5. Differential hydroxylation of proline and lysine residues in $\alpha 2$ chain. ....	50
Table 6. Comparison of fibroblasts grown on UC and AC using RT-PCR. ....	64

### Appendix III: Type I collagen sequence alignments

## Protein sequence alignment of Type I collagen from bovine, human and rat

CO1A1\_alignment

		Section 1					
	(1)	1	10	20	30	40	52
CO1A1_Bovin	(1)	QLSYGYDEKSTG	-I	SVVPGPMGPGSPRGL	PGPPGAPGPGQGFQGGPPGEPGEPGA		
CO1A1_human	(1)	QLSYGYDEKSTGGI	SVVPGPMGPGSPRGL	PGPPGAPGPGQGFQGGPPGEPGEPGA			
CO1A1_rat	(1)	QMSYGYDEKSAQ	-V	SVVPGPMGPGSPRGL	PGPPGAPGPGQGFQGGPPGEPGEPGA		
Consensus	(1)	QLSYGYDEKSTG	I	SVVPGPMGPGSPRGL	PGPPGAPGPGQGFQGGPPGEPGEPGA		
		Section 2					
	(53)	53	60	70	80	90	104
CO1A1_Bovin	(52)	SGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPGQARGLPGTAGLPGMK					
CO1A1_human	(53)	SGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPGQARGLPGTAGLPGMK					
CO1A1_rat	(52)	SGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPGQARGLPGTAGLPGMK					
Consensus	(53)	SGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPGQARGLPGTAGLPGMK					
		Section 3					
	(105)	105	110	120	130	140	156
CO1A1_Bovin	(104)	GHRGFSGLDGAKGDA	GPAGPKGEPGSPGENGA	PGQMGPRLPGERGRPGA	PG		
CO1A1_human	(105)	GHRGFSGLDGAKGDA	GPAGPKGEPGSPGENGA	PGQMGPRLPGERGRPGA	PG		
CO1A1_rat	(104)	GHRGFSGLDGAKGDT	GPAGPKGEPGSPGENGT	PGQMGPRLPGERGRPGA	PG		
Consensus	(105)	GHRGFSGLDGAKGDAG	GPAGPKGEPGSPGENGAP	PGQMGPRLPGERGRPGA	PG		
		Section 4					
	(157)	157	170	180	190	208	
CO1A1_Bovin	(156)	PAGARGNDGAT	GAAGPPGPTGPA	GPPGFPGA	VAKGEG	GPQGRGSEGPQGV	
CO1A1_human	(157)	PAGARGNDGAT	GAAGPPGPTGPA	GPPGFPGA	VAKGEG	GPQGRGSEGPQGV	
CO1A1_rat	(156)	TAGARGNDGAT	VGAAGPPGPTGPT	GPPGFPGA	VAKGEG	GPQGRGSEGPQGV	
Consensus	(157)	PAGARGNDGAT	GAAGPPGPTGPA	GPPGFPGA	VAKGEG	GPQGRGSEGPQGV	
		Section 5					
	(209)	209	220	230	240	250	260
CO1A1_Bovin	(208)	RGEPGPPGPPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPFGARGPSSGPQ					
CO1A1_human	(209)	RGEPGPPGPPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPFGARGPSSGPQ					
CO1A1_rat	(208)	RGEPGPPGPPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPFGARGPSSGPQ					
Consensus	(209)	RGEPGPPGPPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPFGARGPSSGPQ					
		Section 6					
	(261)	261	270	280	290	300	312
CO1A1_Bovin	(260)	GPSGPPGPKGN	SGEPGAPGSKGDTGAKGEPGPTG	VQGGPPPAGEEGKRGARG			
CO1A1_human	(261)	GPSGPPGPKGN	SGEPGAPGSKGDTGAKGEPGPTG	VQGGPPPAGEEGKRGARG			
CO1A1_rat	(260)	GPSGAPGPKGT	SGEPGAPGNKGDTGAKGEPGPTG	VQGGPPPAGEEGKRGARG			
Consensus	(261)	GPSGPPGPKGN	SGEPGAPGSKGDTGAKGEPGPTG	VQGGPPPAGEEGKRGARG			
		Section 7					
	(313)	313	320	330	340	350	364
CO1A1_Bovin	(312)	EPGPAAGLP	PPGERGGPGRGFP	PGADGVAGPKGPA	AGERG	SPGPAGPKGSPGE	
CO1A1_human	(313)	EPGPAAGLP	PPGERGGPGRGFP	PGADGVAGPKGPA	AGERG	SPGPAGPKGSPGE	
CO1A1_rat	(312)	EPGPAAGLP	PPGERGGPGRGFP	PGADGVAGPKGPA	AGERG	SPGPAGPKGSPGE	
Consensus	(313)	EPGPAAGLP	PPGERGGPGRGFP	PGADGVAGPKGPA	AGERG	SPGPAGPKGSPGE	
		Section 8					
	(365)	365	370	380	390	400	416
CO1A1_Bovin	(364)	AGRPGEAGLP	PAKGLTGS	PGSPGPDGKT	GPPG	PAGQDGRPPG	GGPPGARGQA
CO1A1_human	(365)	AGRPGEAGLP	PAKGLTGS	PGSPGPDGKT	GPPG	PAGQDGRPPG	GGPPGARGQA
CO1A1_rat	(364)	AGRPGEAGLP	PAKGLTGS	PGSPGPDGKT	GPPG	PAGQDGRPPG	GGPPGARGQA
Consensus	(365)	AGRPGEAGLP	PAKGLTGS	PGSPGPDGKT	GPPG	PAGQDGRPPG	GGPPGARGQA

### Appendix III: Type I collagen sequence alignments

#### CO1A1\_alignment

		Section 9					
		(417) 417	430	440	450	468	
CO1A1_Bovin	(416)	GVMGFPPGPKGAGEPPGKAGERGVPPGAVGVPAGKDGEEAGAQGGPPGPAGPAG					
CO1A1_human	(417)	GVMGFPPGPKGAGEPPGKAGERGVPPGAVGVPAGKDGEEAGAQGGPPGPAGPAG					
CO1A1_rat	(416)	GVMGFPPGPKGTAGEPPGKAGERGVPPGAVGVPAGKDGEEAGAQGGPPGPAGPAG					
Consensus	(417)	GVMGFPPGPKGAGEPPGKAGERGVPPGAVGVPAGKDGEEAGAQGGPPGPAGPAG					
		Section 10					
		(469) 469	480	490	500	510	520
CO1A1_Bovin	(468)	ERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPPSGARGGERGF					
CO1A1_human	(469)	ERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPPSGARGGERGF					
CO1A1_rat	(468)	ERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPPSGARGGERGF					
Consensus	(469)	ERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPPSGARGGERGF					
		Section 11					
		(521) 521	530	540	550	560	572
CO1A1_Bovin	(520)	PGERGVQGPVPPGAPPRGANGAPGNDGAKGDTGAPGAPGSQGAPGLQGMPPGER					
CO1A1_human	(521)	PGERGVQGPVPPGAPPRGANGAPGNDGAKGDTGAPGAPGSQGAPGLQGMPPGER					
CO1A1_rat	(520)	PGERGVQGPVPPGAPPRGANGAPGNDGAKGDTGAPGAPGSQGAPGLQGMPPGER					
Consensus	(521)	PGERGVQGPVPPGAPPRGANGAPGNDGAKGDTGAPGAPGSQGAPGLQGMPPGER					
		Section 12					
		(573) 573	580	590	600	610	624
CO1A1_Bovin	(572)	GAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPPPAGAPGDKGEAG					
CO1A1_human	(573)	GAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPPPAGAPGDKGEASG					
CO1A1_rat	(572)	GAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPPPAGAPGDKGEAG					
Consensus	(573)	GAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPPPAGAPGDKGEAG					
		Section 13					
		(625) 625	630	640	650	660	676
CO1A1_Bovin	(624)	PSGPAGPTGARGAPGDRGEPGPPGPAFAGPPGADGQPPGAKGEPGDAGAKGD					
CO1A1_human	(625)	PSGPAGPTGARGAPGDRGEPGPPGPAFAGPPGADGQPPGAKGEPGDAGAKGD					
CO1A1_rat	(624)	PSGPAGPTGARGAPGDRGEPGPPGPAFAGPPGADGQPPGAKGEPGDAGAKGD					
Consensus	(625)	PSGPAGPTGARGAPGDRGEPGPPGPAFAGPPGADGQPPGAKGEPGDAGAKGD					
		Section 14					
		(677) 677	690	700	710	728	
CO1A1_Bovin	(676)	AGPPGPAFAGPPGPIGNVGAPEKGRGASAGPPGATGFPGAAGRVPVPPGPPS					
CO1A1_human	(677)	AGPPGPAFAGPPGPIGNVGAPEKGRGASAGPPGATGFPGAAGRVPVPPGPPS					
CO1A1_rat	(676)	AGPPGPAFAGPPGPIGNVGAPEKGRGASAGPPGATGFPGAAGRVPVPPGPPS					
Consensus	(677)	AGPPGPAFAGPPGPIGNVGAPEKGRGASAGPPGATGFPGAAGRVPVPPGPPS					
		Section 15					
		(729) 729	740	750	760	770	780
CO1A1_Bovin	(728)	GNAGPPGPPGPAKKEGSKGPRGETGPAGRPGEVGGPPPPGPPAGEKGPAGADG					
CO1A1_human	(729)	GNAGPPGPPGPAKKEGSKGPRGETGPAGRPGEVGGPPPPGPPAGEKGPAGADG					
CO1A1_rat	(728)	GNAGPPGPPGPAKKEGSKGPRGETGPAGRPGEVGGPPPPGPPAGEKGPAGADG					
Consensus	(729)	GNAGPPGPPGPAKKEGSKGPRGETGPAGRPGEVGGPPPPGPPAGEKGPAGADG					
		Section 16					
		(781) 781	790	800	810	820	832
CO1A1_Bovin	(780)	PAGAPGTPGPPQGIAGQRGVVGLPGQRGERGFPPGLPGPSGEPGKQGPSGASGE					
CO1A1_human	(781)	PAGAPGTPGPPQGIAGQRGVVGLPGQRGERGFPPGLPGPSGEPGKQGPSGASGE					
CO1A1_rat	(780)	PAGAPGTPGPPQGIAGQRGVVGLPGQRGERGFPPGLPGPSGEPGKQGPSGASGE					
Consensus	(781)	PAGAPGTPGPPQGIAGQRGVVGLPGQRGERGFPPGLPGPSGEPGKQGPSGASGE					





## Appendix III: Type I collagen sequence alignments

CO1A2\_alignment

					Section 9		
	(417)	417	430	440	450	468	
CO1A2_Bovin	(415)	GPKGPSGDPSKAGFKGHAGLAGARGAPGPDGNNGAQQGPPGLQGVQGGKGEQG					
CO1A2_human	(417)	GPKGPTGDPGKNGLKGHAGLAGARGAPGPDGNNGAQQGPPGLQGVQGGKGEQG					
CO1A2_rat	(417)	GPKGPSGDPSKAGFKGHAGLAGARGAPGPDGNNGAQQGPPGLQGVQGGKGEQG					
Consensus	(417)	GPKGPSGDPSKAGFKGHAGLAGARGAPGPDGNNGAQQGPPGLQGVQGGKGEQG					
						Section 10	
	(469)	469	480	490	500	510	520
CO1A2_Bovin	(467)	PAGPPGFQGLPGFAGTAGEAGKPGERGLPGEFGLPGPAGARGERGPPGESGA					
CO1A2_human	(469)	PAGPPGFQGLPGFSGPAGEVVKPGERGLHGEFGLPGPAGPRGERGPPGESGA					
CO1A2_rat	(469)	PAGPPGFQGLPGFSGTAGEVVKPGERGLPGEFGLPGPAGPRGERGPPGESGA					
Consensus	(469)	PAGPPGFQGLPGFSGTAGEVVKPGERGLPGEFGLPGPAGPRGERGPPGESGA					
							Section 11
	(521)	521	530	540	550	560	572
CO1A2_Bovin	(519)	AGPTGPIGSRGSPGPPGPDGNKGEFGVVGAVGTAGPSGSPSGLPGERGAAGIIP					
CO1A2_human	(521)	AGPTGPIGSRGSPGPPGPDGNKGEFGVVGAVGTAGPSGSPSGLPGERGAAGIIP					
CO1A2_rat	(521)	AGPSGPIGIRGSPGAPGPDGNKGEAGAVGAVTSSAGASGPGGLPGERGAAGIIP					
Consensus	(521)	AGPTGPIGSRGSPGPPGPDGNKGEFGVVGAVGTAGPSGSPSGLPGERGAAGIIP					
							Section 12
	(573)	573	580	590	600	610	624
CO1A2_Bovin	(571)	GGKGEKGETGLRGETIGSPGRDARGAPGATGAPGPAGANGDRGEAGPAGPAG					
CO1A2_human	(573)	GGKGEKGETGLRGETIGNPGRDARGAPGAVGAPGPAGATGDRGEAGPAGPAG					
CO1A2_rat	(573)	GGKGEKGETGLRGETIGNPGRDARGAPGATGAPGPAGASGDRGEAGPAGPAG					
Consensus	(573)	GGKGEKGETGLRGETIGNPGRDARGAPGATGAPGPAGASGDRGEAGPAGPAG					
							Section 13
	(625)	625	630	640	650	660	676
CO1A2_Bovin	(623)	PAGPRGSPGERGEVGPAGPNGFAGPAGAGQPGAKKERGTGKPKGENGVVGP					
CO1A2_human	(625)	PAGPRGSPGERGEVGPAGPNGFAGPAGAGQPGAKKERGAKGPKGENGVVGP					
CO1A2_rat	(625)	PAGPRGSPGERGEVGPAGPNGFAGPAGAGQPGAKKERGTGKPKGENGVVGP					
Consensus	(625)	PAGPRGSPGERGEVGPAGPNGFAGPAGAGQPGAKKERGTGKPKGENGVVGP					
							Section 14
	(677)	677	690	700	710	728	
CO1A2_Bovin	(675)	TGPVGAAGPSGPNGPPGPAAGSRGDGGPPGATGFPGAAGRTGPPGPGSGISGPP					
CO1A2_human	(677)	TGPVGAAGPSGPNGPPGPAAGSRGDGGPPGATGFPGAAGRTGPPGPGSGISGPP					
CO1A2_rat	(677)	TGPVGAAGPSGPNGPPGPAAGSRGDGGPPGATGFPGAAGRTGPPGPGSGISGPP					
Consensus	(677)	TGPVGAAGPSGPNGPPGPAAGSRGDGGPPGATGFPGAAGRTGPPGPGSGISGPP					
							Section 15
	(729)	729	740	750	760	770	780
CO1A2_Bovin	(727)	GPPGFAGKEGLRGRGDQGPVGRSGETGASGPPGFVGEKGPSGEAGTAGPPG					
CO1A2_human	(729)	GPPGFAGKEGLRGRGDQGPVGRSGEIVGAVGPPGFVGEKGPSGEAGTAGPPG					
CO1A2_rat	(729)	GPPGAAGKEGLRGRGDQGPVGRSGEIVGAVGPPGFVGEKGPSGEAGTAGPPG					
Consensus	(729)	GPPGFAGKEGLRGRGDQGPVGRSGEIVGAVGPPGFVGEKGPSGEAGTAGPPG					
							Section 16
	(781)	781	790	800	810	820	832
CO1A2_Bovin	(779)	TGGPQGLLGAAGFLGLPGRGERGLPGVAGSVGEFGLGIAGPPGARGPPGN					
CO1A2_human	(781)	TGGPQGLLGAAGFLGLPGRGERGLPGVAGAVGEPGFLGIAGPPGARGPPGA					
CO1A2_rat	(781)	TAGPQGLLGAAGFLGLPGRGERGLPGVAGAVGEPGFLGIAGPPGARGPPGA					
Consensus	(781)	TGGPQGLLGAAGFLGLPGRGERGLPGVAGAVGEPGFLGIAGPPGARGPPGA					



## Acknowledgements

### Acknowledgements

Accomplishing of this PhD-project and writing of this dissertation have been the first significant challenge in my scientific carrier. To following people I owe my deepest gratitude, because of their patience, support, and guidance during my work.

- **Prof. Dr. Herwig Brunner**, who was the supervisor of my work despite his other academic and professional activities. His experience, knowledge, commitment and patience encouraged and motivated me through the whole work.
- **Prof. Dr. Robin Ghosh**, who gave me the practical supervise and brought me into the experimental field of protein structure biology; He was always patient to teach me his experience and techniques, additionally to improve my English.
- **PD-Dr. Steffen Rupp**, he gave me the possibility to complete this work in his department and was always disposed for my questions and discussions, also intensive supervision in molecular biological and cell biological experiments.
- **Prof. Dr. Arthur Veis**, from Northwestern University USA. Who very critically reviewed my manuscript for publication and gave me many useful practical, literature suggestions.
- **Dr. Kai Sohn**, who gave me many suggestions both in scientific work and real life. Especially, his humorous talking has always inspired me despite the high pressure in the experiments.
- **Mr. Ekkehard Hiller**, who taught me the mass spectrometric techniques and permanent being “free” for a discussion and helpful by any difficulties. Specially, he never gave up correcting my German language and tried to better my German continuously!
- **Dr. Friedel Drepper**, from University Freiburg. Who helped me to obtain the collagen sequence with high coverage and accuracy.
- **Dr. Hans Weber, Mr. Marija Dukalska, Mr. Marc Röhm**, who gave me their lab-experiences and encouraged me through all the years.
- **All the friendly colleagues in the Department of Molecular Biotechnology**, who supported and helped me during my work and residence in Germany.
- **Cuifeng Zhao**, my wife, without whom, Completion of this dissertation would be almost impossible; She decelerated her own study to support me and taught me to sacrifice, being patient and to compromise.

## Acknowledgement

- *My parent*, who have always supported, encouraged and believed in me, in all my endeavors.
- *Fengni Xiong*, my daughter, who was just born several months before this dissertation was completed. She detracted me from the stressful experiments and writing in her own special way.

## Acknowledgements

### Curriculum vitae

Xin Xiong

Born on December 06, 1977 in Liaoning, P. R. China

### Education

- Jan. 2005-                      PhD-thesis work at Institute for Interfacial Engineering, University  
Stuttgart/ Fraunhofer Institute for Interfacial Engineering and  
Biotechnology (IGB). Supervised by Prof. Dr. H. Brunner.  
Title: New insights into Structure and Function of Type I collagen
03. 2004 - 12. 2004              Diploma thesis at Fraunhofer IGB supervised by Prof. Dr. H.  
Brunner.  
Title: Functional Characterization of Tsa1p in *Candida albicans*
10. 2002 - 12. 2004              Study of Technical Biology at University Stuttgart  
Degree: Diplom-Biologe (technisch orientiert)
04. 2002 - 12.2004              Research assistant at Fraunhofer IGB, Dept. Molecular  
Biotechnology
08. 2002 - 11. 2002              Practical training at MANN+HUMMEL GmbH in Ludwigsburg  
Focus: Characterization of ceramic-hollow fiber membrane  
Pre-development department by Dr. S. Tudyka
10. 2001 - 03. 2002              Intensive course for DSH at the University of Stuttgart
01. 2000 - 06. 2000              Bachelor thesis at Dalian Institute of chemical physics,  
Chinese Academy of Science, supervised by Prof. X. J. MA  
Title: Effect of chitosan properties on the membrane strength of  
chitosan-alginate microcapsules
09. 1996 - 07. 2000              Study of biochemical engineering at Huaqiao University, China  
Degree: Bachelor-Biochemical Engineering

### Awards

10. 2005                              First Hugo-Geiger Price
01. 2005-07.2008                  Scholarship of Peter und Traudl Engelhorn foundation

## Publications

### Publications

1. Xiong, X., Ghosh, R., et al., Molecular characterization of urea-extracted rat tail collagen Type I shows unusual assembly- and hydroxyproline-patterns. FEBS J. Submitted
2. Xiong, X., Ghosh, R., et al., Comparison and characterization of Type I collagen as biomaterial for tissue engineering using different isolation methods. Accepted by J. Biotechnol., in Proceeding of IBS2008.
3. Xiong, X., Brunner, H., et al., Verfahren zur Isolierung von Kollagen aus kollagenhaltigem Gewebe patented 30. 04. 2008.
4. Urban, C., Xiong, X., et al., The moonlighting protein Tsa1p is implicated in oxidative stress response and in cell wall biogenesis in *Candida albicans*. Mol Microbiol, 2005. **57**(5): p. 1318-41.

## Acknowledgements

### Declaration

I, hereby declare that this dissertation entitled “**New Insights into Structure and Function of Type I Collagen**” is my own work and all the sources I have used or quoted have been clearly indicated or acknowledged by means of completed references.