# Characterisation of Extracellular Matrix Cross-Linking in Systemic Sclerosis

**Mariya Evgenieva Semkova**

A thesis submitted for the degree of Doctor of Philosophy

Supervised by:

Professor Justin Hsuan

Dr Voon Ong

University College London

2019

*I, Mariya Semkova confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.*

Mariya Semkova,
October 2019

## Abstract

Systemic sclerosis (SSc) is a complex connective tissue disease associated with a high mortality. As development of skin fibrosis in SSc patients leads to major organ-based complications, finding a treatment that resolves this condition can help prevent disease progression. A major obstacle in the discovery of new effective anti-fibrotic therapies is that the exact mechanisms that lead to fibrotic progression and irreversibility have not been fully elucidated.

The fibrotic phenotype is associated with increased accumulation of extracellular matrix (ECM) in an organ that severely impedes its function. Excessive cross-linking of the ECM is a major factor causing its resistance to remodelling by rendering sites of proteolysis less accessible to matrix metalloproteinases. The discovery of new cross-linking sites during disease progression may therefore improve our molecular understanding of the resistance of the ECM to proteolytic breakdown, act as a clinical marker for resolvable vs non-resolvable fibrosis, and contribute to the development of new therapeutic approaches.

Transglutaminases are a group of cross-linking enzymes that play a pivotal role in fibrotic progression. A mass spectrometry-based method for the identification of transglutaminase cross-linking sites from complex biological samples has been developed in this project. A model transglutaminase cross-linked peptide was synthesized and its fragmentation behaviour was analysed. After having found the optimal fragmentation conditions the analysis of cross-linking sites was automated using MassMatrix software. The resulting score values were utilised to identify true-positive matches. This technique was successfully applied to an analysis of the transglutaminase cross-linking site of purified fibrinogen dimers. In the final stages of the method development comparative

studies of the ECM cross-linking pattern between control and pro-fibrotic dermal fibroblast cell culture and mouse models were performed. The limited number of new cross-linking sites found in the fibrotic models suggest that more significant changes occur at later fibrotic stages.

## Impact statement

Fibrotic diseases are a major health burden in the UK and worldwide. Consequent economic and social costs are high as disease progression commonly spans decades. Tissue fibrosis is a condition that can affect any type of organ. It can severely impair organ function and finally lead to organ failure. At present, there is no available treatment to directly inhibit or reverse fibrosis and the only way to treat irresolvable fibrosis is an organ transplant. There is therefore an overwhelming clinical need for new effective strategies for the treatment of fibrosis.

Fibrotic development is a very complex process in which multiple molecular mechanisms are involved. A large proportion of the mechanisms that are crucial for disease progression are not yet fully characterised. Thus, in order to find new effective anti-fibrotic drugs, the corresponding core pathological pathways need to be fully understood.

Systemic sclerosis (SSc) is a fibrotic malignancy that affects a number of organs. Skin fibrosis is an almost universal symptom among SSc patients. It deteriorates their quality of live and increases the risk of other organ-based complications. Therefore, it is believed that an effective treatment for this condition would require a therapy that is able to reverse skin fibrosis.

In this project, we aimed to contribute to the study of a pathway that plays a major role in fibrotic progression and irreversibility. More precisely, the focus of our research was to investigate alterations in the sites of protein cross-linking in the extracellular matrix that lead to its excessive accumulation. To address this question, a key advance that would be applied throughout our project was to develop a method that enabled the analysis of cross-linked sites in biological

samples. In the course of our project, we developed a mass spectrometry based technique that can be employed to identify cross-linked sites induced by transglutaminases – a group of enzymes whose hyperactivation has been observed in several types of fibrosis. The method can be exploited to reveal key pathogenic mechanisms not only in skin fibrosis, but also in other affected organs. This method enables the discovery of new targets for anti-fibrotic therapies that may also serve as diagnostic biomarkers. Thus, a long term goal of our research was to contribute to new strategies for the diagnosis and treatment of different types of fibrotic disease. In addition, the method we developed can serve as a new analytical tool to identify transglutaminase cross-linked sites relevant to other biological processes. In other words, it can also help address a range of scientific questions in the fields ranging from structural biology to human pathology.

## Acknowledgements

First of all, I would like to express my deepest gratitude to my primary supervisor Prof Justin Hsuan for all his support and guidance throughout our PhD project. His scientific knowledge and his encouragement has been invaluable not only to my thesis but also to my development as a researcher.

I would also like to thank our secondary supervisor Dr Voon Ong for the discussions and advices especially during our experiments with cell cultures and mouse models.

Furthermore, I would like to thank Dr Konstantinos Thalassinos for sharing his expertise in the field of mass spectrometry. Special thanks to Mark Crawford for all his invaluable support during my mass spectrometry measurements.

I am grateful to the very kind staff from the Rheumatology department – especially to Dr Xu Shiwen for providing training and assistance during my cell culture experiments and Prof David Abraham for his readiness to help. I was also happy to join the ILDH department that created a friendly and supportive atmosphere.

I would like to thank my colleagues Dr Adam Dyer, Dr Xiufang Lyu and Xiaoping Huang MD whose help has been extremely useful.

Special thanks to Dr Andrew Hall for his training and advice during my histology experiments, Dr Korsa Khan and Prof Elisabeth Blankenhorn for providing mouse skin samples and Prof Alethea Tabor for offering her advice during my synthesis experiments.

Finally, I would like to thank my family for their moral support and encouragement.

## Contents

## List of figures

## List of tables

## List of abbreviations

| | |
|---|---|
| ACE | angiotensin-converting enzyme |
| ACN | acetonitrile |
| AEBSF | 4-(2-aminoethyl)benzenesulfonyl fluoride |
| AECA | anti-endothelial cell antibody |
| AH | aldolhistidine |
| AIF-1 | allograft inflammatory protein-1 |
| BCA | bicinchoninic acid |
| bFGF | basic fibroblast growth factor |
| Boc | *tert*-butyloxycarbonyl protecting group |
| BSA | bovine serum albumin |
| CCL | chemokine (C–C motif) ligand |
| CCR | chemokine (C-C motif) receptor |
| CDI | 1,1'-Carbonyldiimidazole |
| CID | collision induced dissociation |
| CNBr | cyanogen bromide |
| CTGF | connective-tissue growth factor |
| dcSSc | diffuse cutaneous SSc |
| DHLN | dihydroxylysinonorleucine |
| DIEA | N,N-Diisopropylethylamine |

DMF dimethylformamide

DMSO dimethyl sulfoxide

DTT dithiothreitol

ECM extracellular matrix

ED ethylene diamine

EDC 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

EDTA ethylenediaminetetraacetic acid

ESI electrospray ionisation

ET1 endothelin-1

FBS fetal bovine serum

FDR false discovery rate

FGFR fibroblast growth factor receptor

Fmoc fluorenylmethyloxycarbonyl protecting group

GAPDH glyceraldehyde 3-phosphate dehydrogenase

HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HCCA alpha-cyano-4-hydroxycinnamic acid

HCD higher-energy C-trap dissociation

HCl hydrochloric acid

HHMD histidinohydroxymerodesmosine

HLD hydroxylysinonorleucine

HMD hydroxymerodesmosine

HOCl hypochlorous acid

HP hydroxylysylpyridinoline

HPLC high performance liquid chromatography

HRP horseradish peroxidase

HSC hepatic stellate cells

IAA iodoacetamide

IEF isoelectric focusing

IFN interferon

IL interleukin

IPF idiopathic pulmonary fibrosis

LC-MS liquid chromatography-mass spectrometry

LH2b lysyl hydroxylase 2b

LLC large latent complex

LN lysinonorleucine

LOX lysl oxidase

LOXL lysyl oxidase like

LP lysylpyridinoline

LTBP latent transforming growth factor beta binding protein

MALDI-TOF matrix-assisted laser desorption/ionization - time-of-flight

| | |
|---|---|
| MCP-1 | monocyte chemoattractant protein-1 |
| MES | 2-morpholin-4-ylethanesulfonic acid |
| MDC | monodansylcadaverine |
| MMP | matrix metalloproteinase |
| MS | mass spectrometry |
| NAPDH | nicotinamide adenine dinucleotide phosphate |
| NHS | N-hydroxysulfosuccinimide |
| PAH | pulmonary arterial hypertension |
| PBS | phosphate-buffered saline |
| PDGF | platelet-derived growth factor |
| PDGFR | platelet-derived growth factor receptor |
| PH | pulmonary hypertension |
| PLOD2 | procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 |
| PNGase F | Peptide:N-glycosidase F |
| PVDF | polyvinylidene fluoride |
| RLU | relative luminescent units |
| ROS | reactive oxygen species |
| RTX | rituximab |
| SDS-PAGE | sodium dodecyl sulfate–polyacrylamide gel electrophoresis |
| SCX | strong cation exchange |

| | |
|---|---|
| TβRIIΔk | kinase-deficient type II TGFβ receptor |
| TBST | Tris-buffered saline Tween 20 |
| TFA | trifluoroacetic acid |
| TG | transglutaminase |
| TGFβ | transforming growth factor β |
| TGFβR | transforming growth factor β receptor |
| TH | T helper cell |
| THF | tetrahydrofuran |
| TIMP | tissue inhibitor of metalloproteinase |
| TNF | tumor necrosis factor |
| TREG | regulatory T cells |
| TSK | tight skin |
| TXLP | tryptic cross-linked peptide |
| VEGF | vascular endothelial growth factor |
| VEGFR | vascular endothelial growth factor receptor |

# Chapter 1

# Introduction

## 1.1 Fibrotic diseases

Fibrotic diseases are a major health issue that is estimated to be responsible for 45% of the mortality in the Western world [1]. The fibrotic process is marked by replacement of normal tissue with non-functional fibrotic tissue that disrupts the normal organ architecture and finally leads to organ failure [2]. Any organ of the human body can be affected by fibrosis. The high mortality rate linked to this disease type is a result of the large number of cases, the lack of validated biomarkers for early diagnosis, the unrevealed underlying molecular mechanisms during disease manifestation, the variations in the fibrotic etiology depending on the affected organ and the lack of effective disease-modifying therapies [3]. Fibrotic disorders occur in a wide spectrum of diseases including systemic diseases such as systemic sclerosis (SSc), sclerodermatous graft vs. host disease and nephrogenic systemic fibrosis; and fibrotic disorders affecting individual organs, such as pulmonary, liver, cardiac and kidney fibrosis [4]. This condition is caused by abnormalities in normal tissue repair mechanisms that are controlled by regulatory mechanisms under homeostatic conditions. Under pathologic conditions however, these regulatory mechanisms are suppressed, causing the uncontrolled accumulation of fibrotic tissue [3]. The molecular pathways causing fibrosis can be divided into common pathways that can be found in every type of fibrosis [5], and disease-specific pathways that are unique for the affected organ [6]. Current strategies for anti-fibrotic therapies focus on

fundamental common pathways in order to find universal anti-fibrotic drugs, that could be used to treat any affected organ [6].

### 1.2 Systemic sclerosis

Systemic sclerosis (SSc) is a connective tissue disease, with unknown aetiology and a variable and unpredictable outcome [7]. Due to its complex nature and the heterogeneity of the clinical outcome, finding an effective therapeutic approach is a challenging task [8]. The most common features of SSc are vascular injury, inflammation and extensive fibrosis, which most commonly affects the skin [9, 10]. Depending on the extent of skin involvement at diagnosis, the disease can be classified into limited cutaneous SSc, where skin fibrosis is mainly restricted to the hands, arms and face [10] and diffuse cutaneous SSc, in which a larger skin area (typically proximal limbs and trunk) is affected [8]. Other distinctive features of the two types of scleroderma are the increased vascular damage in patients with limited cutaneous SSc, while in diffuse cutaneous SSc inflammation and fibrosis are more prominent [8].

The underlying cause for SSc is not yet fully elucidated [11]. It is believed that exposure to several environmental factors such as viruses, drugs and other substances (e.g. silica or vinyl chloride) may trigger symptoms typical for scleroderma [12]. Other literature suggests that variations in the human genome can also enhance the susceptibility of an individual to the disease [13, 14]. For instance, particular alleles of the genes encoding transforming growth factor-β (TGFβ), monocyte chemoattractant protein-1 (MCP-1),

tumour necrosis factor (TNF), allograft inflammatory protein-1 (AIF-1), and angiotensin-converting enzyme (ACE) are likely to be responsible for systemic sclerosis [13].

Early manifestations of the disease are marked by macrovascular damage, which occurs in swelling and morphological abnormalities such as gaps between adjacent endothelial cells, vacuolization and duplications of the basement membrane [9, 10]. Alterations of vascular structure are due to intimal hyperplasia and fibrosis of the vessel media and adventitia as well as proliferation of smooth muscle cells in the vessel wall [8]. Endothelial cell damage is caused by hypoxia and the presence of reactive oxygen species. It has been suggested that endothelial injury is the cause of the Raynaud's phenomenon, which is usually the first symptom of scleroderma [15]. As a result of tissue injury, inflammatory processes are activated. Damage to endothelial cells evokes the secretion of factors that induce platelet aggregation and degranulation, followed by the release of cytokines that increase endothelial permeability [16]. This enables lymphocyte migration through the endothelial barrier, which leads to the recruitment of leukocytes to the site of injury. In healthy individuals, these events stimulate tissue repair and scar resolution during the process of wound healing [8]. In scleroderma patients however, the chronic nature of vascular injury causes alterations in the inflammatory and fibrogenic mechanisms observed during normal wound healing and repair [16].

One of the most characteristic features of the inflammatory response in scleroderma is the release into circulation of autoantibodies directed against nuclear proteins. These antibodies can serve as diagnostic biomarkers [17].

Several antibodies are associated with the type of scleroderma and are used to predict specific disease complications. For example, anti-centromere antibodies are characteristic of limited cutaneous scleroderma and development of pulmonary hypertension, while anti-RNA polymerase III antibodies are usually detected in diffuse cutaneous scleroderma patients with the risk of renal involvement [9]. The role of these antibodies in the pathogenesis of the disease is still unclear, although the presence of anti-endothelial cell antibodies (AECAs) has been shown induce endothelial cell apoptosis and therefore contribute to permanent tissue injury [18].

Alongside the generation of autoantibodies, activation of cells involved in inflammatory responses such as lymphocytes, macrophages and neutrophils, is also implicated in SSc [8]. There is extensive evidence showing that early inflammatory infiltrates produce transforming growth factor beta (TGFβ), platelet-derived growth factor (PDGF), interleukins, and connective-tissue growth factor (CTGF), which are necessary for the activation of fibroblasts and the transition of fibroblasts to myofibroblasts [9]. In response to vascular inflammation, increased production and aberrant accumulation of extracellular matrix (ECM), commonly referred to as fibrosis, is observed, and this is another characteristic component of SSc. The most universal feature of SSc is development of skin fibrosis which is seen as the main event triggering organ-based complications including pulmonary arterial hypertension (PAH), lung fibrosis, renal failure and disorders in the gastrointestinal tract [8]. The major molecular events of scleroderma pathogenesis are schematically depicted in Figure 1.1.

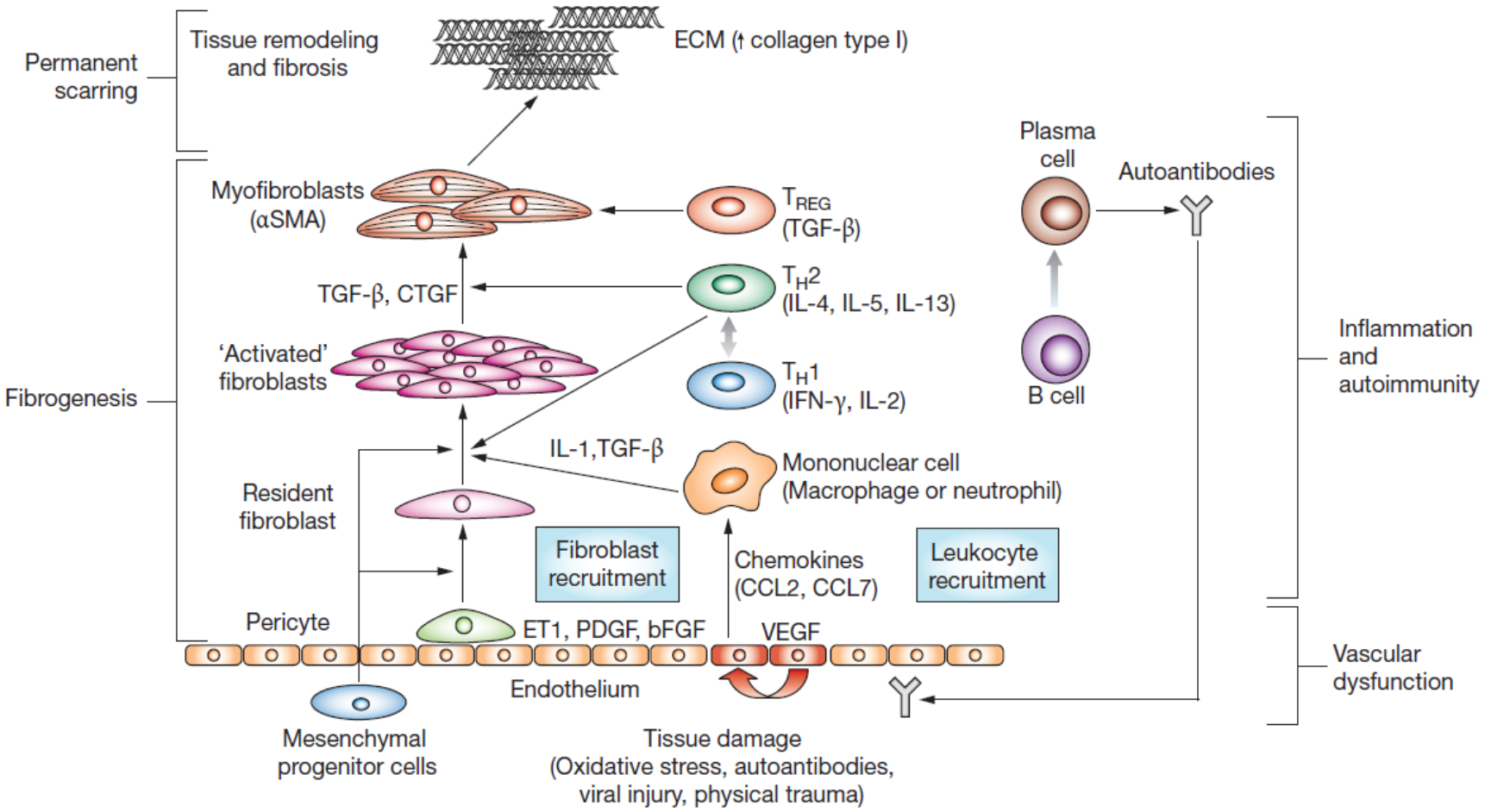

Permanent scarring
Tissue remodeling and fibrosis
ECM (↑ collagen type I)
Myofibroblasts (αSMA)
TREG (TGF-β)
Plasma cell
Autoantibodies
TGF-β, CTGF
TH2 (IL-4, IL-5, IL-13)
'Activated' fibroblasts
TH1 (IFN-γ, IL-2)
B cell
Fibrogenesis
IL-1,TGF-β
Mononuclear cell (Macrophage or neutrophil)
Resident fibroblast
Fibroblast recruitment
Chemokines (CCL2, CCL7)
Leukocyte recruitment
Inflammation and autoimmunity
Pericyte
ET1, PDGF, bFGF
VEGF
Endothelium
Vascular dysfunction
Mesenchymal progenitor cells
Tissue damage (Oxidative stress, autoantibodies, viral injury, physical trauma)

**Figure 1. 1 Key molecular pathways of scleroderma (adapted from [8]).** The pathogenic process starts with injury of the endothelium that provokes the release of the vascular endothelial growth factor (VEGF) and other chemokines that increase the endothelial permeability and the proliferation and recruitment of leucocytes. Macrophages and neutrophils are also responsible for the recruitment of blood cells. These mechanisms are characteristic of tissue repair occurring under homeostatic conditions. In SSc however, permanent vascular damage and the failure to resolve the inflammatory response lead to chronic inflammation and finally to excessive tissue scarring (fibrosis). The inflammation process includes the activation of T-cells that produce various pro-fibrotic cytokines (interleukin-4, 5 and 13). Plasma cells secrete antibodies that further promote endothelial injury. Endothelial growth factors recruit mesenchymal progenitor cells and resident fibroblasts. These cells are then activated by pro-fibrotic factors secreted by leucocytes and T-cells (TGFβ, CTGF, etc.) and differentiate into myofibroblasts that are responsible for the excessive deposition of extracellular matrix.
αSMA, alpha smooth muscle actin; bFGF, basic fibroblast growth factor; CCL, chemokine (C–C motif) ligand; ET1, endothelin-1; IFN, interferon; IL, interleukin; TH1, type 1 T helper cells; TH2, type 2 T helper cells; TREG, regulatory T cells.

### 1.3 Skin involvement in SSc

As previously mentioned, treatment of SSc provides a challenge due to its complexity and heterogeneity. Depending on the stage of disease presentation, effective approaches to the treatment of SSc would need to reduce the onset or progression of skin disease, or resolve organ based complications such as pulmonary, cardiac or renal involvement [19]. In general, a correlation is observed between the extent of skin involvement and the severity of disease. In this regard, patients suffering from diffuse cutaneous scleroderma, characterized by severe or widespread skin involvement in the proximal extremities or the trunk, are more likely to develop subsequent organ damage and organ failure leading to death. One clinical study demonstrated that in diffuse cutaneous SSc (dcSSc) patients progressive skin fibrosis is associated with decline in lung function and worse survival during follow-up [20]. Other studies have revealed that a spontaneous improvement of the skin state is linked to a lower risk of new organ damage, and reduced morbidity [21]. These findings suggest that skin fibrosis is a key indicator of disease outcome and the appearance of subsequent organ-based complications. Thus, finding new methods to resolve skin fibrosis could provide an effective disease-modifying therapy [8].

### 1.4 Fibrotic development in SSc

The fibrotic process in SSc has a detrimental effect on the overall outcome of the disease, since it leads to the most severe complications [22]. Apart from the skin, vascular structures, the musculoskeletal system and internal organs such as the lung and the heart can be affected by fibrosis during disease progression [8].

Skin fibrosis is the earliest fibrotic event in scleroderma beginning in the deep dermis and the subcutaneous fat tissue [10]. Enhancement of ECM production combined with inhibited ECM degradation and remodelling leads to accumulation of ECM structures in an irregular arrangement [23]. As a consequence, the dermal architecture is disrupted, and appears as a reduction of microvasculature and in later stages by a loss of appendages and rete ridges [9].

Fibrotic development is initiated by the transition of fibroblasts to myofibroblasts as a response to the release of interleukins, chemokines and growth factors such as TGFβ, PDGF and CTGF by inflammatory cells [9]: for example, macrophages are the major source of TGFβ-1 in fibrotic tissue [16]. Other pathways, such as Wnt, Notch, Hedgehog and Jak-Stat signalling, are also involved in fibrotic development [9]. Although the majority of activated myofibroblasts arise from cells present in the affected tissue, they can also arise from cells recruited from the circulation, such as mesenchymal

progenitor cells [9]. The presence of elevated levels of reactive oxygen species (ROS) damages fibroblast DNA and this can also lead to enhanced collagen synthesis [24]. The source of ROS is the membrane NAPDH oxidase system that is activated upon injury of vessel walls [25]. Therefore, this process is relatively independent of the immune response. Furthermore, it was demonstrated that scleroderma fibroblasts overexpress TGFβ receptors (TGFβR) and thereby sustain the fibrotic condition via autocrine signalling [26]. TGFβ signalling leads to the activation of Smad and non-Smad transcription factors that enhance the expression of genes encoding ECM proteins [7].

These processes are parts of a healthy response to tissue damage. Under acute conditions, the primary role of myofibroblasts is contraction of early granulation tissue, which is an early stage in the wound healing process. Subsequently, after wound contraction is completed, myofibroblasts are removed by apoptosis, which is an essential part of wound resolution [27]. In contrast, in the case of a chronic fibrotic disorder the constant release of TGFβ suppresses the apoptotic clearance of myofibroblasts, causing uncontrolled ECM production and therefore propagation of fibrosis [28]. Figure 1.2 illustrates common signalling pathways that lead to fibroblast activation in scleroderma.

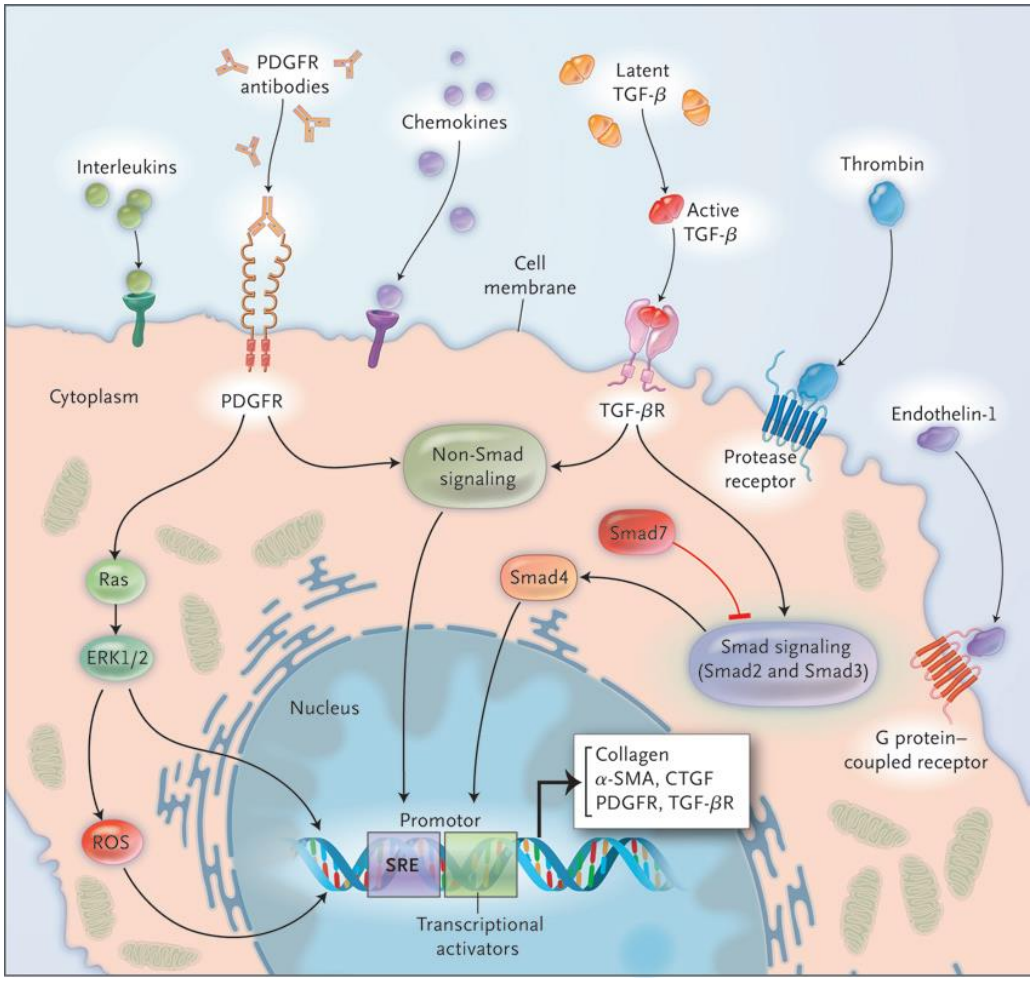



**Figure 1. 2 Fibroblast activation in SSc (figure adapted from [25]).** External factors such as interleukins, chemokines, growth factors, thrombin, endothelin-1, reactive oxygen species (ROS) and activating antibodies induce fibrotic signalling cascades. These events lead to the production of ECM proteins, cytoskeleton, cytokines and cytokine receptors that participate in autoregulatory signalling loops that help sustain fibroblast activation.

The increased ECM production in fibrotic tissues is accompanied by decreased ECM catabolism. The ECM is a dynamic structure that undergoes a constant rebuilding and remodelling via synthesis, degradation, reassembly and chemical modification [29]. These tightly regulated processes are important for development, wound healing and normal tissue homeostasis [29]. Indeed, dysregulated ECM remodelling is associated with

various pathologic conditions including fibrosis, cancer and osteoarthritis [30, 31]. Degradation is an important process during remodelling that controls the abundance, composition and structure of the ECM and is crucial for the release of different cell signalling molecules [29]. It is performed by various types of proteases such as matrix metalloproteinases (MMPs), adamalysins, meprins and cathepsins [29].

The main family of proteases involved in the ECM catabolism are the MMPs [32]. These enzymes target a wide range of ECM proteins; they are divided into six groups: collagenases, gelatinases, stromelysins, matrilysins, membrane-type MMPs, and other non-classified MMPs [33]. The mechanism of collagen I proteolysis by collagenases involves binding of the proteinase to the cleavage site, local unwinding of the collagen triple helical structure followed by hydrolysis of the $Gly^{775}$– $Ile^{776}$ and $Gly^{775}$–$Leu^{776}$ bonds [34].

Another large family of ECM proteases are the disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS) that are mainly responsible for the degradation of proteoglycans [35]. Other matrix proteins such as fibrin, fibronectin, and laminin are proteolysed by the serine protease plasmin [35].

In skin fibrosis, it is believed that while the amount of ECM proteins increases, the activity of MMPs is inhibited by the release of anti-MMP autoantibodies [36, 37]. Other literature suggests that MMP inhibition also results from increased levels of tissue inhibitors of metalloproteinases (TIMPs) that have been found in scleroderma skin fibroblasts [38, 39]. Another effect, which is thought to cause impaired degradation of ECM proteins is alterations in their cross-linking pattern [40]. Excessive cross-

linking changes the properties of the ECM by rendering it resistant to remodelling [41]. A possible explanation for this decreased digestibility, is that highly cross-linked ECM undergoes conformational changes that restrict the access of proteases to ECM cleavage sites [42]. In addition to impaired ECM degradation, it is believed that elevated cross-linking levels contribute to higher matrix stiffness [43], which promotes further production of matrisomal proteins. These changes are crucial for the development of SSc because cellular behaviour is influenced by the mechanical properties of the surrounding ECM. For example, this effect has been demonstrated by a study on collagen type I in which cross-linking by transglutaminase 2 (TG2) led to increased stiffness of collagen fibres that promoted cell attachment, spreading and proliferation of residing osteoblasts [44]. In another demonstration, cultivation of primary rat hepatic stellate cells (HSCs) on inert polyacrylamide supports with variable shear modulus (stiffness) revealed that an increase in polymer stiffness boosts HSC differentiation into myofibroblasts [43]. In conclusion, alterations in matrix stiffness due to elevated ECM cross-linking levels seem to play a key role in the fibrogenic process.

### 1.5. ECM cross-linking enzymes and their role in fibrosis

The optimal function of tissues is dependent on the activity of several cross-linking enzymes that catalyse the formation of covalent bonds between adjacent ECM proteins [45]. These bonds are essential for maintaining the structure and mechanical strength of the ECM [46]. In fibrotic tissue however,

hyperactivity of these enzymes plays a crucial role in fibrotic progression. There are different types of cross-links that are generated by different pathways. A common pathway for the formation of cross-links involves the modification of lysine by lysyl oxidases and lysyl hydroxylases giving raise to allysine and hydroxylysine respectively [47]. This allows subsequent reaction cascades to take place with other lysine, hydroxylysine and/or histidine residues to give raise to di-, tri- and tetravalent cross-links [48]. Common examples for chemical pathways leading to the formation of mature cross-links is depicted in Figure 1.3.

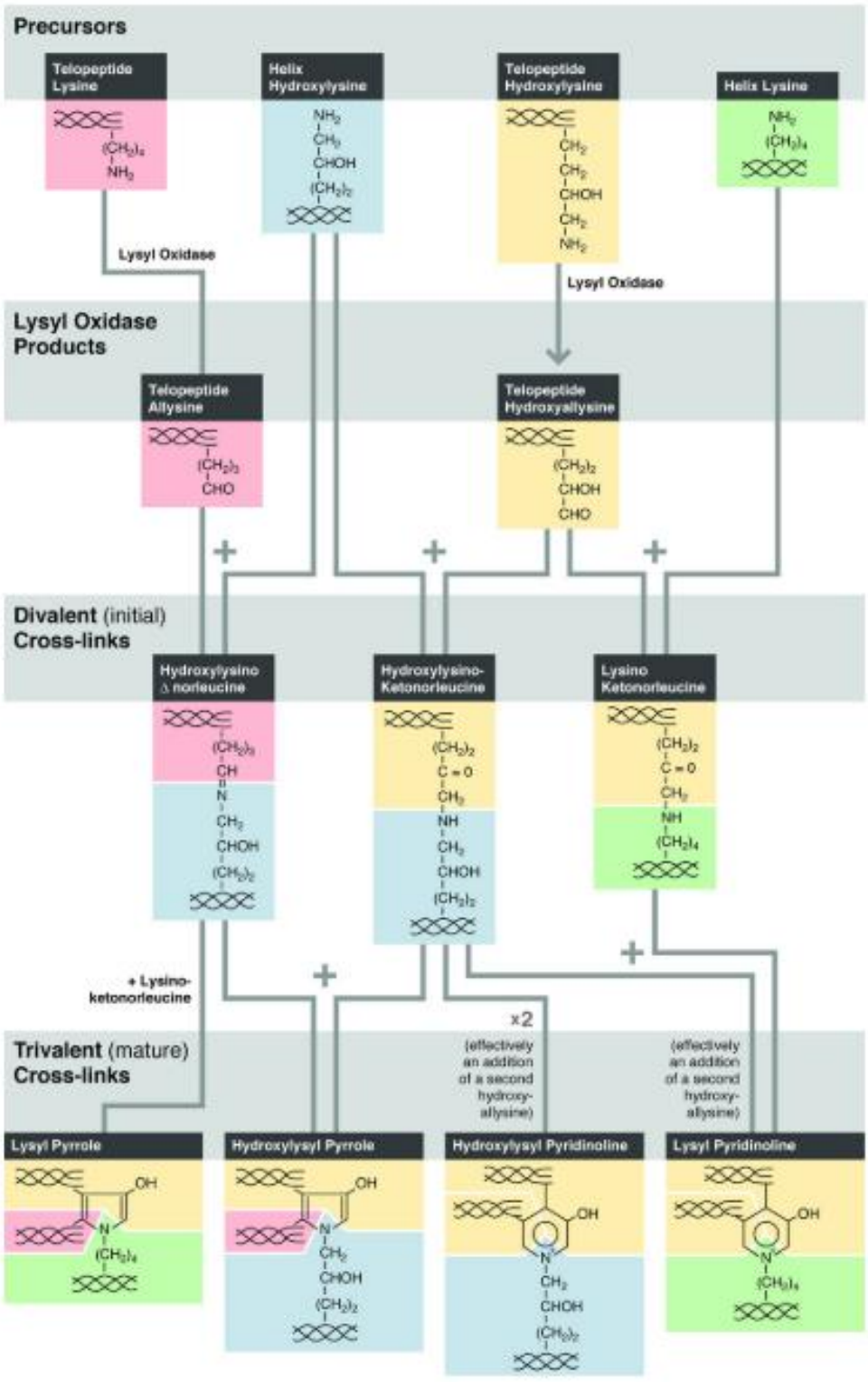



**Figure 1. 3 Formation of mature cross-links initiated by lysyl oxidase and lysyl hydroxylase** (figure adapted from [47])**.**

A typical example of a mature cross-link is the pyridinoline group. This tetravalent covalent cross-link can be formed between three adjacent collagen chains in a collagen fibril. Figure 1.4 depicts a pyridinoline cross-link between the helical region of a collagen chain and two non-helical regions (telopeptides) that can be either N-terminal or C-terminal.

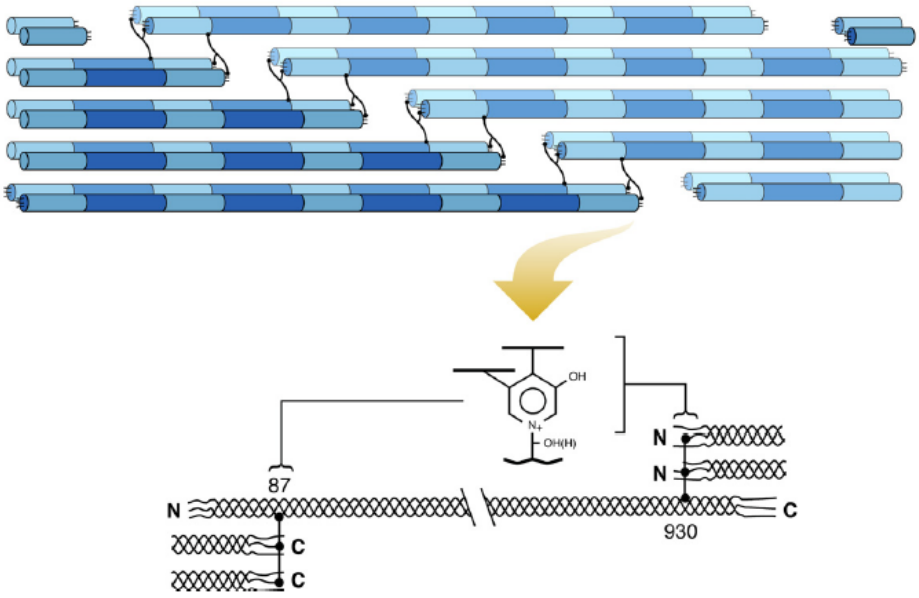



**Figure 1. 4 Schematic representation of an intermolecular pyridinoline cross-link between adjacent collagen molecules within a type II collagen fibril in human cartilage** (adapted from [47]). In this example, a cross-link between a helical region and two telopeptides is formed. N: N-terminal telopeptide; C: C-terminal telopeptide.

Under normal conditions, the allysine route predominates in skin, leading to a higher proportion of allysine-based cross-links and a lower proportion of hydroxyallysine-based cross-links [49]. In fibrotic skin diseases however, this equilibrium is shifted in favour of hydroxyallysine cross-links. Van der Slot et al. demonstrated that fibroblasts isolated from fibrotic SSc skin show

increased levels of the lysyl hydroxylase 2b enzyme (LH2b, encoded by the PLOD2 gene) [50]. This overexpression in SSc is accompanied by the accumulation of pyridinoline cross-links (hydroxylysyl pyridinoline (HP) and lysyl pyridinoline (LP), Figure 1.3) that are formed via the hydroxyallysine route [50, 51]. The increase in pyridinoline cross-links is associated with impaired ECM degradation as collagenous matrices containing higher levels of pyridinoline cross-links were more resistant to matrix remodelling by MMPs [52]. In contrast with these results, fibrotic skin isolated from bleomycin-treated mice, where this cross-link type is not increased, showed reversibility after omission of the pro-fibrotic agonist [52]. This observation further supports the theory that a change in cross-linking type is crucial to the irreversibility of fibrosis.

Alongside lysyl hydroxylase, evidence from different studies demonstrates that lysyl oxidase (LOX) is another cross-linking enzyme that is upregulated. For example, a significant increase in LOX concentration was found in serum obtained from SSc patients in comparison with healthy controls [53, 54]. Moreover, immunostaining revealed that LOX is particularly enriched in lung biopsies taken from SSc patients with pulmonary hypertension (PH), while no visible increase was found in the skin [54]. Further evidence for the role of LOX enzymes in lung fibrosis is provided by a study of a lysyl oxidase like 1 (LOXL1) deficient mouse model that proved to be protected from experimental lung fibrosis by possessing lower tissue stiffness and reduced activation of TGFβ-1 [55]. Such observations led to the suggestion that an effective anti-fibrotic therapy could be developed by inhibiting the activity of the lysyl oxidase like 2 (LOXL2) enzyme with the monoclonal antibody

Simtuzumab [56]. Using this agent, a reduction of activated fibroblasts, decreased production of growth factors and cytokines, and suppressed TGFβ signalling were demonstrated in models of liver and lung fibrosis [56]. The following mechanism explaining the role of LOXL2 in promoting fibroblast activation has been suggested. An increase of the ECM cross-linking, that leads to enhanced ECM stiffness, elevates local matrix tension [56]. This results in a conformational change in the large latent complex (LLC) – a TGFβ-containing molecular aggregate situated in the extracellular space [57]. As a consequence, TGFβ is released from the complex and the TGFβ signalling cascade is activated [57]. However, further drug development was not undertaken after the antibody did not improve the progression-free survival of patients with idiopathic pulmonary fibrosis (IPF) and liver fibrosis [58, 59]. One possible reason why treatment with Simtuzumab was not an effective anti-fibrotic therapy is that cross-linking by different enzymes is too redundant for inhibition of one isoform to cause significant inhibition of fibrotic progression in IPF [58]. Interestingly, lysyl oxidase has also been found to play an important role in the formation of osteolytic lesions during ER-breast tumour metastasis by disrupting normal bone homeostasis in favour of osteoclast-dependent resorption, thereby generating a niche environment for circulating tumour cell attachment and proliferation [60]. Whether or not LOX activity in this setting is related to its ECM cross-linking activity remains to be investigated.

Another important family of cross-linking enzymes are the transglutaminases [61]. These enzymes directly catalyse condensation of a glutamine and a lysine residue and are able to either induce polymer formation (*de novo* direct

polymerization) or stabilize non-covalent assemblies by introducing covalent bonds between adjacent polymers, known as enzymatic “spot welding” [61]. Typical targets for these enzymes are different types of collagen, fibronectin and elastin [61, 62]. Apart from their cross-linking activity, this class of enzymes has been reported to possess other enzymatic activities, including protein phosphorylation [63], glutamine deamidation, and amine incorporation [61]. The role of transglutaminases has been studied in relation to various fibrotic diseases [64-66]. Transglutaminase catalysed cross-links ($N^{\varepsilon}$(γ-glutamyl)lysine cross-links) that are undetectable in normal liver, were found in fibrotic liver via immunohistochemical staining [66]. Similar results were found in patients with IPF together with an increase of transglutaminase 2 (TG2) [64] that was found to be activated upon tissue injury leading to the cross-linking of many ECM proteins [67]. Moreover, the importance of TG2 in the development of pulmonary fibrosis was indicated by the observation that TG2-knockout mice developed significantly reduced fibrosis upon bleomycin treatment in comparison with wild-type controls [64]. Another study showed suppression of fibrotic development in cardiac tissues by direct inhibition of TG2 cross-linking activity [65]. It has been hypothesized that this enzyme plays a crucial role in inhibiting collagen turnover as a result of its collagen cross-linking activity [64]. In this regard, it has also been shown that increased TG2 activity leads to decreased collagen digestibility by MMP-1 [68]. Further evidence suggests that TG2 is able to activate TGFβ-1 signalling by increasing the deposition of large latent TGFβ-1 in the ECM before proteolytic release of the active, fibrogenic form [69, 70]. A combination of these two mechanisms might be responsible for the

accumulation of collagen in the ECM as a response to increased TG2 activity [71]. So far, there is no published evidence showing that transglutaminase-induced mechanisms are also relevant for the development of skin fibrosis in systemic sclerosis although a significant upregulation of TG2 mRNA was found in skin biopsies from patients with diffuse SSc [72]. Moreover, it was found that the TG2 protein level is increased in human dermal fibroblasts in response to stimulation with TGFβ-1 [73], which leads to the hypothesis that transglutaminases can also play a role in skin fibrosis.

A list ECM cross-linking enzymes commonly found in skin tissue together with their properties is given Table 1.1.

| *Protein* | Gene | Synonym | Subcellular location | Function |
|---|---|---|---|---|
| *TG2* | TGM2 | Protein-glutamine gamma-glutamyltransferase 2 | Cytosol, endoplasmic reticulum, extracellular region, mitochondrion, plasma membrane | Formation of a $N^{\varepsilon}$(γ-glutamyl)lysine cross-link |
| *LH1* | PLOD1 | Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 | Endoplasmic reticulum | Hydroxylation of lysine residues in collagen alpha chains |
| *LH2* | PLOD2 | Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 | Endoplasmic reticulum | Hydroxylation of lysine residues in -Xaa-Lys-Gly- sequences in collagens |
| *LOX* | LOX | Protein-lysine 6-oxidase | Extracellular region or secreted | Oxidative deamination of peptidyl lysine residues in precursors to fibrous collagen and elastin |
| *LOXL1* | LOXL1 | Lysyl oxidase homolog 1 | Extracellular region or secreted | |
| *LOXL2* | LOXL2 | Lysyl oxidase homolog 2 | Nucleus, endoplasmic reticulum, extracellular region or secreted | |

**Table 1. 1 Cross-linking enzymes found in skin tissue and their properties** (source: Uniprot)**.**

### 1.6 Analysis of cross-linking sites in ECM proteins

Several attempts have been made to determine the cross-linked positions in different ECM proteins in order to elucidate the exact mechanism of polymer formation. To achieve this various approaches have been adopted especially in the investigation of cross-linking sites induced by transglutaminases [74-76].

One study analysed the positions of cross-links induced by a microbial transglutaminase acting on purified calf skin type I collagen [74]. In brief, reactive glutamine residues were cross-linked to monodansylcadaverine (MDC) – a fluorescent compound containing a free amino group through which it could be enzymatically linked to glutamine. The collagen was then digested with cyanogen bromide (CNBr) and the fragments were separated from each other using standard chromatographic techniques. To estimate the number of cross-links in each fragment, the corresponding level of fluorescence was measured. While this method gave information about the protein fragments containing high cross-link densities, the exact positions of the cross-linked sites remained unresolved. Another disadvantage of this experiment was that the position of lysine cross-linking sites remained unknown, or even if suitable lysine residues existed within collagen I. Finally, purified proteins can undergo conformational changes or modifications in the course of extraction and purification that can influence the outcome of the cross-linking analysis. Consequently, no definite conclusions about any cross-linking positions relevant in vivo could be drawn from this study.

A different strategy for determining the position of transglutaminase cross-links was adopted by Purves et al. [75]. In this study, the cross-linked positions between two fibrin γ-chains were determined by first isolating γ-chain dimers from cross-linked fibrin $(\alpha\beta\gamma)_2$ polymers. Subsequently, the γ-γ dimers were digested and the cross-linked peptides isolated using chromatographic separation and then sequenced via gas-phase Edman degradation. A single cross-link was identified between lysine-406 and either glutamine-398 or -399. The identity of the cross link was then confirmed by coelution with the corresponding synthetic cross-linked dipeptide.

A further strategy for the identification of cross-linked peptides derived from protein digests is mass-spectrometric analysis [77]. The principle of cross-linked peptide detection is similar to the one used in the sequence analysis of linear peptides in proteomics experiments. Fragmentation of the analyte with an inert collision gas results in dissociation of the peptide backbone that gives rise to characteristic series of fragments, from which the amino acid sequence is deduced. The fragmentation spectrum of cross-linked peptides typically shows a series of cross-linked and non-cross-linked fragments from which one can determine the cross-link position as illustrated in Figure 1.5. This technique has been used for the identification of transglutaminase cross-links from complex protein digests [78-80]. Moreover, fragmentation analysis has been automated by using software specialized in the identification of cross-linked peptides from mass spectrometry data [78, 80]. This method has not yet been used for analysis of transglutaminase cross-linked positions in ECM proteins.

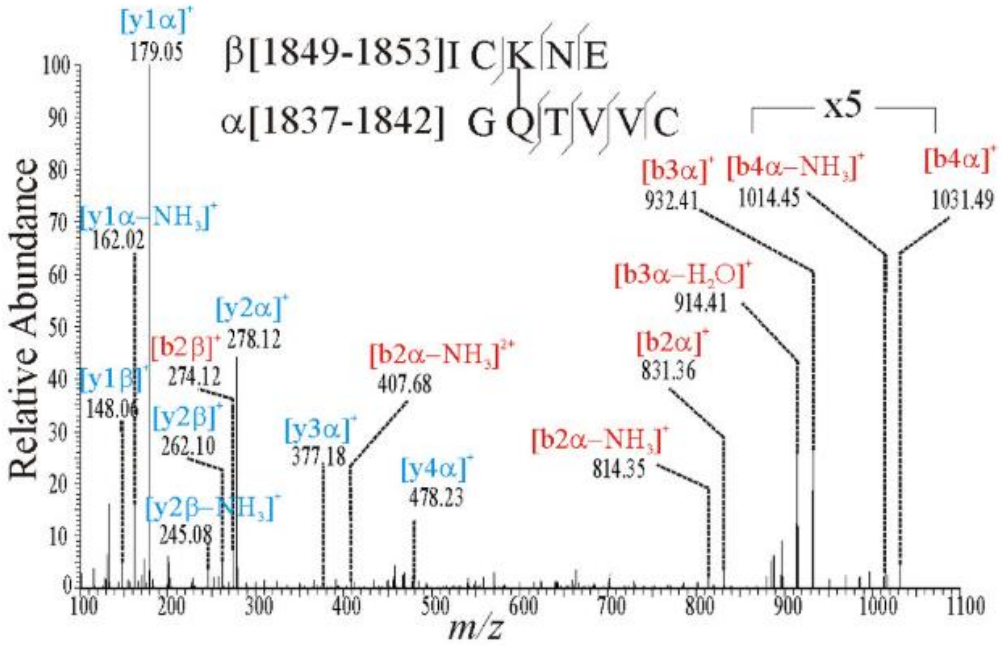



**Figure 1. 5 MS2 fragmentation of a transglutaminase cross-linked peptide pair (adapted from [78]).** Fragmentation of the backbone of both peptides involved in the cross-link (α and β) results in series of linear and cross-linked fragments that allow the position of the cross-link to be determined.

Mass spectrometry has also been used to identify the sites cross-linked by pyridinoline and pyrrole groups in collagen I in human bone [81]. In this study collagen I was digested with collagenase, resulting in a mixture of cross-linked and non-cross-linked peptides. The cross-linked peptides were enriched using molecular-sieve chromatography in which elution of the analytes was monitored by measuring the fluorescence signal from the pyridinoline group and the reactivity of pyrrole groups with Ehrlich's reagent. Individual cross-linked peptides were resolved using ion-exchange chromatography and reverse-phase, high performance liquid chromatography (HPLC). The structure of the peptides was established by combining the results of N-terminal microsequencing, cross-link analysis,

electrospray mass spectrometry and immunoassay. Mass spectrometry was used solely for determining the analyte mass without generating any fragmentation data. With the help of this method, cross-links between the triple helical region and telopeptides from the N- and C-terminal regions, respectively, were identified.

Further development of the mass-spectrometric analysis of pyridinoline cross-linked peptides was demonstrated in another study that analysed the cross-linking sites in type II collagen in human cartilage [47]. Pepsin-solubilized collagen II was digested with CNBr and the fragments were separated using gel electrophoresis. A band containing the cross-linked fragment was further digested in an excised gel piece using trypsin and analysed using liquid chromatography-mass spectrometry (LC-MS) (Figure 1.6). MS2 fragmentation of the +3-charged peptide showed fragments resulting from dissociation of the peptide backbones.

The mass analyser used in this thesis was a linear trap quadrupole (LTQ)-Orbitrap. In this system, ion trapping, ion selection, fragmentation and low resolution ion detection take place in the LTQ while the Orbitrap is used for high resolution and high mass accuracy ion detection [82]. Successful proteomic analysis of a complex protein digest using this instrument has been previously reported [83]. In brief, the mass of the peptides was first detected in the Orbitrap followed by isolation, fragmentation and low-resolution scan of the product ions in the LTQ. This system has also been used to identify transglutaminase cross-linked peptides from a protein digest [80].

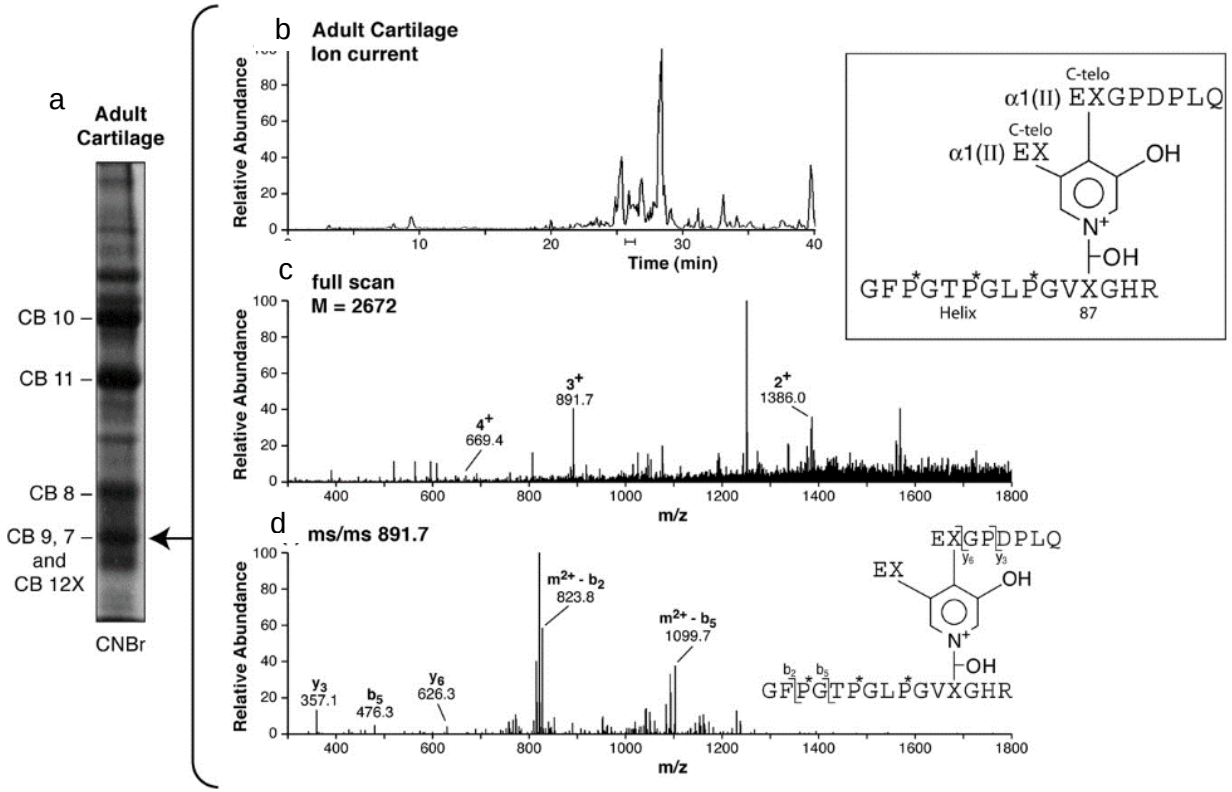



**Figure 1. 6 Identification of a hydroxylysylpyridinoline cross-link between two C-terminal telopeptide domains (C-telo) and a helical domain (Helix) in type II collagen from human cartilage** (adapted from [47]). The cross-linked sites are marked with X. a. SDS-PAGE of the type II collagen fragments after CNBr digestion. The fragment indicated with an arrow has been digested with trypsin and the resulting peptides were separated with HPLC (b). The MS1 spectrum of the cross-linked peptide and the fragmentation spectrum of the +3-charged parent ion are displayed in panels c and d, respectively.

### 1.7 Animal models of SSc

There is a large number of animal models of SSc and each of them primarily mimics one of the three different aspects of the disease: immunologic abnormalities, vasculopathy or fibrosis of the skin or internal organs [84-86]. There is no available model that covers the whole spectrum of pathogenic events that occur during SSc progression [84]. Nevertheless, animal models are a useful tool for resolving major pathologic mechanisms and for testing new therapeutic approaches [85]. There are several available SSc models showing the dermal fibrotic phenotype that are either inducible, where fibrosis results from a tissue injury with an injected agent or genetic models, in which genetic mutation can trigger the accumulation of ECM [85].

For example, skin and lung fibrosis can be induced by injection of bleomycin [87], hypochlorous acid (HOCl) [88] or topoisomerase-I and complete Freund's adjuvant [89]. In general, these reagents cause an inflammatory response that eventually leads to tissue fibrosis [85]. Another common mechanism that is also found in SSc patients is the production of ROS that can also increase collagen synthesis [88]. These models are suitable for analyses of the earliest disease course events, as the exposure to the pro-fibrotic agent typically lasts only several weeks [87, 89]. Another disadvantage provided by these models is that fibrosis is limited only to the site of injection, which differs from the systemic character of SSc [88].

Among the most popular genetic models showing a fibrotic phenotype are the tight skin (TSK) models [90, 91] and the Fra-2 transgenic model [92]. The two types of tight-skin mouse arise from a tandem duplication of the *fibrilin-1* gene (TSK1) [90] or a point mutation in the *Col3a1* gene (TSK2) [93]. These abnormalities subsequently lead to the activation of pro-fibrotic mechanisms that increase ECM deposition. The TSK1 model shows activation of the Wnt signalling cascade, which contributes to fibrotic development [94], whereas in TSK2, the dermal accumulation of TGFβ in the skin suggests that TGFβ signalling is responsible for skin fibrosis [95]. However, the exact mechanism behind these transgenic models remains unproven [85]. A major difference in these models in comparison with SSc is the occurrence of fibrosis in the hypodermis, which makes the relevance of these models to SSc questionable [96].

Another strategy for the induction of fibrotic features in mouse models is by the overexpression of the Fos-related antigen-2 (Fra-2) transcription factor, which is responsible for regulating several stress responses including inflammation and wound healing [97], and mediates TGFβ and PDGF signalling in SSc patients [92]. In addition to skin fibrosis, Fra-2 transgenic mice replicate a large spectrum of events relevant to SSc including pulmonary fibrosis, vasculopathy and inflammation [84].

A further model that relies on enhanced TGFβ signalling is a transgenic mouse line expressing a kinase-deficient type II TGFβ receptor (TβRIIΔk mice) [98]. Originally, this model was designed to study the role of disrupted downstream TGFβ signalling in fibrosis by modifying the type II high affinity receptor (TβRII). Contrary to the expected results, this model showed TGFβ hyperactivation that led to the development of pulmonary

and skin fibrosis with increased dermal thickness with a loss of the subcutaneous fat layer.

There is currently no known animal model that mimics the increased activity of the common ECM cross-linking enzymes as seen in SSc. As described in the previous section, increased ECM cross-linking is one of the major causes leading to decreased ECM remodelling. Therefore, a better understanding of the mechanism of altered cross-linking in SSc could enable the discovery of new insights into the causes of fibrotic irreversibility and the development of more effective anti-fibrotic therapies. In order to achieve this, an appropriate animal model would be extremely advantageous.

### 1.8 Current strategies for therapeutic approaches to SSc

Although treatments are available for many of the symptoms arising in the course of SSc, there is currently no approved curative therapy [99]. Moreover, treatment of some complications such as fatigue, calcinosis and anorectal dysfunction remain very challenging [99]. A major challenge in finding an appropriate way to resolve fibrosis is that, although fibrotic mechanisms in different diseases have similar clinical outcomes, they differ in terms of their underlying molecular pathways [2]. Thus, in order to find an appropriate target for a given fibrotic disease, the specific molecular mechanisms activated in that disease must be identified. In addition, as fibrotic development relies on a complex network of multiple and mostly redundant pathways, suppressing only one of them could be insufficient for the effective resolution of fibrosis. Hence, a combination of

different agents targeting different key pro-fibrotic pathways needs to be considered in the development of an anti-fibrotic therapy [100], for which further progress in the elucidation of fibrotic pathogenesis is required [9].

Due to the complexity and variety of mechanisms contributing to fibrosis, there is a large number of candidate anti-fibrotic targets [100, 101]. The development of immune suppressants targeting cells and cytokines essential for disease progression, such as interleukin-6 (IL-6) [102], interleukin-13 (IL-13) [103] and B-lymphocytes [104], has been investigated and drugs have been developed that attenuate fibrosis in SSc mouse models. For example, Rituximab (RTX) is a monoclonal antibody that depletes B-cells that are a major source of cytokine production involved in fibrotic progression [105]. So far, treatment using RTX has been demonstrated to improve the skin fibrosis score of SSc patients along with the improvement/stabilisation of lung fibrosis [106]. One suggested application of RTX is in the treatment of patients with very early symptoms of SSc to help prevent complications occurring at later disease stages [107]. However, it is currently impossible to predict which organs will be affected by fibrosis during disease progression [107]. Moreover, due to the heterogeneity of the early symptoms, a highly-sensitive and highly-specific method of diagnosis prior to skin involvement is still lacking [108].

A recent area of work on anti-fibrotic therapies concerns the critical role of myofibroblasts in fibrosis and the signalling pathways in these cells that regulate the transcription of genes involved in fibrosis. It has been suggested that a therapy that targets the transition of fibroblasts into pro-fibrotic myofibroblasts could inhibit fibrosis [109]. Since TGFβ is often seen to be an important mediator of this transition, inhibiting the TGFβ

pathway could be an effective therapeutic strategy. Possible approaches aiming to inhibit TGFβ production or activity by suppressing expression of its gene [110], blocking stimulators of TGFβ production [111], or inhibiting TGFβ itself [112] have been investigated. However, none of these studies has led to any promising anti-fibrotic agents. Other cell types including macrophages, monocytes and platelets are also considered to be possible targets for the development of novel anti-fibrotic therapies [100, 110]. For instance, successful inhibition of monocyte/macrophage recruitment during hepatic inflammation has been demonstrated by blocking the interaction of the chemokine (C-C motif) ligand 2 (CCL2) with its corresponding receptor CCR2 [113]. Targeting other events that contribute to fibrotic development in SSc, such as responses to vascular damage [114], activity of cytokines [115], production of reactive oxygen species [116], activation of morphogenic pathways (Wnt, Hedgehog and Notch) [117] and epigenetic modifications [118] provide additional potential targets of treatment strategies. An alternative approach to resolving fibrosis is to develop a polypharmacological drug that can suppress several pathologically-relevant pathways [100]. For example, the multi-tyrosine kinase inhibitor Nintedanib is able to block the VEGF receptor (VEGFR), PDGF receptor (PDGFR) and fibroblast growth factor receptor (FGFR) [119] and reduced fibrosis in several in vivo models of SSc [120]. Another anti-fibrotic strategy aims to prevent fibroblast activation in response to increased ECM stiffness. Fibroblasts are able to detect increased ECM stiffness via integrins such as the $\beta_1$ integrin that is upregulated in SSc fibroblasts [100]. Treatment with antibodies against the $\beta_1$ receptor ameliorated fibrosis, and normalized integrin expression and skin stiffness and

architecture in a transgenic mouse model [121]. A more selective approach has been developed by using an inhibitor of the $\alpha_v\beta_1$ heterodimer [122].

Although some of the strategies for anti-fibrotic therapies discussed above have the potential to prevent fibrotic progression of affected organs, none of them has been able to reverse the fibrotic process. The characteristic decrease in ECM digestibility and remodelling is a major event that renders fibrosis irreversible; thus, targeting this aspect is essential to finding a more effective, disease-modifying therapy.

## 1.9 Hypotheses

The hypotheses of this thesis are:

- First, that covalently cross-linked matrisome proteins in fibrotic and control tissues can be identified, and individual cross-linked sites in these tissues can be mapped using high-resolution mass spectrometry.
- Second, that the appearance of new cross-linked sites in the matrisome can be identified during disease progression, and that these play a role in impairing ECM remodelling.

## 1.10 Aims

As stated in Section 1.5, ECM cross-linking is one of the key pathways involved in the development of skin fibrosis in SSc. A better understanding of how cross-linking mechanisms impair ECM remodelling and contribute to

fibrotic progression could broaden the spectrum of therapeutic targets and open an opportunity for the development of novel therapeutic strategies. Although some progress has been made in the characterization of ECM cross-linking in scleroderma [40], more studies are needed to (i) comprehensively identify control and SSc skin ECM proteome (matrisome) components; (ii) identify the ECM proteins affected by the changes in cross-linking; (iii) map their cross-linking sites; and (iv) reveal the mechanism of how the identified changes cause impaired ECM remodelling during disease progression. Such studies could provide currently missing information required to find a corresponding disease-modifying therapy for scleroderma.

In this project, we aimed to address some of these needs. More specifically, we aimed to develop a method to analyse the ECM proteome (matrisome) and its cross-linking patterns in control and fibrotic skin samples. The methodological stages included decellularisation of skin biopsies to remove cell proteins, isolation of ECM proteins, and mass spectrometry-based proteomics. Molecular identification of different cross-linking-structures and peptide sequences were expected to allow the mapping and characterization of cross-linking sites in different ECM proteins. As already shown in Section 1.6, some progress in the mass-spectrometric identification of the ECM proteins and residues involved in different cross-link types has been made. Among them, transglutaminase and hydroxylysylpyridinoline (HP) cross-links were likely to have increased levels in the fibrotic skin of SSc patients [40, 72]. However, further research was required to develop a reliable procedure to extend the identification of cross-linking sites to multiple ECM proteins. our aim was to develop such a procedure and use it to compare cross-linking patterns in disease and control skin samples during different stages of scleroderma progression. In this way we expected to identify specific

changes in the position and abundance of multiple, molecularly-defined cross-links.

In addition to the specific objectives of this project, successful mapping of cross-links could lead to new biomarkers for skin fibrosis. For example, if specific cross-linked sites inhibit proteolysis, they may be able to distinguish reversible from irreversible disease, which is important for personalised treatment.

Successful establishment of a method to map ECM cross-links in SSc skin samples may also be applied to the analysis of cross-linking changes across different fibrotic tissues and thereby help identify novel therapeutic strategies for other fibrotic diseases.

### 1.11 Plan of investigation

At the outset we planned to optimise a bespoke method, developed in the group for the identification of fibrotic tissue matrisomes, for the high-resolution mass spectrometry (MS) analysis of skin matrisomes. It was important for the aims of this thesis to develop a reproducible and efficient protocol for extraction of ECM from skin samples. Furthermore, in order to develop a method suitable for the analysis of ECM cross-linking in skin tissues, simple models of cross-linked peptides were to be synthesised, followed by stepwise increases in sample complexity in order to find the required optimization steps. In parallel with the proteomic analysis, peptides linked by transglutaminase and HP cross-links were to be synthesized and a method for the identification of cross-linked residues based on MS fragmentation spectra would be developed. Success in these experiments would be followed by a comparison of ECM cross-linking in control and pro-

fibrotic fibroblast cultures, which contain less complex ECM than tissue and were therefore more suitable for the initial analytical stages. The experimental workflow consisted of isolation of ECM proteins from primary cell cultures, generation of tryptic digests, isolation of cross-linked peptides via HPLC, and MS-based analysis and qualitative comparison of the cross-linked peptides between control and fibrotic ECM. Finally, the method would be applied to analyse cross-linking in control and fibrotic mouse skin, as well as in human control and SSc skin tissues in order to investigate how the cross-linking pattern changes throughout disease progression.

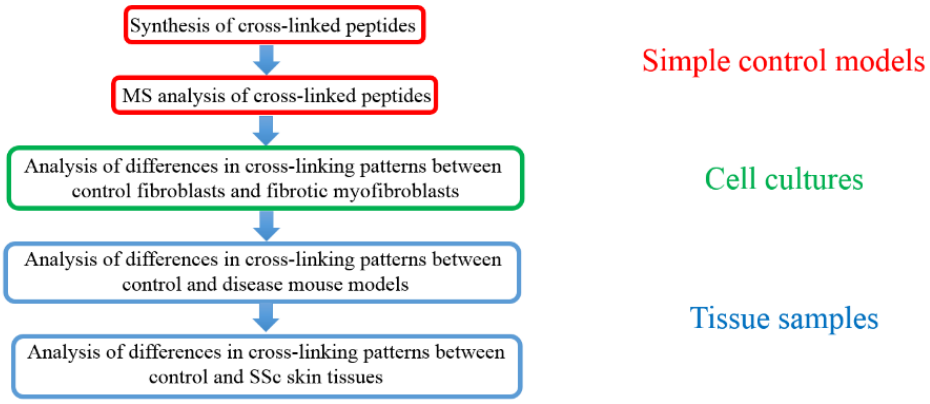



**Figure 1. 7 Schematic representation of the planned project workflow.**

# Chapter 2

# Materials and methods

All materials were purchased from Sigma unless otherwise stated.

## 2.1 Synthesis of cross-linked peptides

### 2.1.1 Synthesis of transglutaminase cross-linked peptides

#### *2.1.1.1 Chemical cross-linking*

##### *2.1.1.1.1 Chemical cross-linking using (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU)*

2.5 mg (8.5 x $10^{-7}$ mole) of Boc-Tyr(tBu)-Glu-Gly-2-chlorotrityl[1] resin (AltaBioscience) were placed on a 2 ml large solid phase peptide synthesis vessel (PeptideWEB) and swollen in 1 ml dimethylformamide (DMF) for 1 day. The solvent was removed by applying a vacuum through the vessel. The resin was washed 4 times with 1 ml DMF, followed by peptide activation with a solution of 0.5 M HBTU (about 235-fold molar excess) and 0.5 M N,N-Diisopropylethylamine (DIEA) in DMF. The reaction was performed for 7 h at room temperature with occasional stirring. The solution was removed and the resin was washed 7 times with 1 ml DMF. A solution of Boc-Ala-Lys-Thr(tBu)-OH (AltaBioscience) (0.02 M; 2 x $10^{-6}$ mole) and DIEA (0.05 M) in DMF was added to the resin followed by incubation at room temperature overnight. The solution was removed and the resin was washed 5 times with

[1] In this thesis the three-letter amino acid code is used only where confusion may arise from using the one-letter code.

1 ml DMF. Deprotection and cleavage of the product from the resin was achieved by treatment with 200 µL 95:5 (v/v) trifluoroacetic acid (TFA):water, for 2 h at room temperature. The solution was diluted with 200 µL water and the resin was removed by centrifugation through a 0.2 µm pore filter. The solvent was evaporated in a vacuum centrifuge (SPD SpeedVac, Thermo) at 20°C and the residue was dissolved in 200 µL 0.1% formic acid. The sample composition was analysed using mass spectrometry on an LTQ Orbitrap Velos system (Thermo Fisher) via direct injection (flow rate: 5µl/min, m/z range: 150-2000Da).

*2.1.1.1.2 Chemical cross-linking using 1,1'-Carbonyldiimidazole (CDI)*

The same amount of Boc-Tyr(tBu)-Glu-Gly-2-chlorotrityl resin as specified in Section 2.1.1.1.1 was swollen in 1 ml tetrahydrofuran (THF) for 1 day. After four washes with 1ml THF, 200 µl 0.5 M CDI ($1 \times 10^{-4}$ mole, 200-fold molar excess) in THF were added. The activation was performed for 2 h at room temperature with occasional stirring. After solvent removal the resin was washed 10 times with 1ml THF and a solution of Boc-Ala-Lys-Thr(tBu)-OH (0.01 M $2 \times 10^{-6}$ mole) and DIEA (0.05 M) in THF was added followed by an overnight incubation. Further sample preparation steps and MS measurement were performed the same way as described in the previous section.

In order to improve the reaction yield, the peptide coupling was extended to 30 h in a separate synthesis.

#### *2.1.1.2 Enzymatic cross-linking*

6 mM peptide stock solutions of Boc-GGAQLGV and Boc-GAKGSA, respectively, were prepared in an aqueous buffer of 400 mM Tris-HCl, 5 mM $CaCl_2$, pH 7.4. The solutions were mixed 1:1 to a final volume of 50 µl (1.5 x $10^{-7}$ mole of each peptide). 40 mU (0.78 mU/µl) of guinea pig liver TG2 was added followed by incubation on an orbital mixer at 800 rpm and 37°C for 15 h. The enzymatic reaction was terminated by adding 45 µL 0.085% TFA. The product was separated from the enzyme via centrifugation through a 3 kDa molecular weight cut-off (MWCO) centrifugal filter (Amicon Ultra) for 20 min at 14,000 x *g*. The peptide solution (filtrate) was purified with a stage tip (binding capacity: 50 µg). The tert-butyloxycarbonyl (Boc) protecting groups were cleaved in 1:1 TFA:DCM on an orbital mixer at 1000 rpm for 2 h at room temperature, followed by solvent evaporation in a vacuum centrifuge. The residue was dissolved in 100 µL 0.1% (v/v) formic acid and analysed using mass spectrometry via direct injection (Section 2.1.1.1.1).

For the cross-linking reaction the peptides Boc-EGGKGPR and Boc-SQDGGR (DC Biosciences) were used. The same reaction as described above was performed, but with the TG2 amount increased to 60 mU (1.18 mU/µl) and the reaction time extended to 3 days.

### 2.1.2 Synthesis of hydroxylysyl pyridinoline (HP) cross-linked peptides

A solution of 0.05 M Fmoc-GST (DC Biosciences) in 0.1 M 2-morpholin-4-ylethanesulfonic acid (MES) buffer (pH 5.6) was incubated on an orbital mixer with 1 M 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 0.115 M sulfo-NHS (N-hydroxysulfosuccinimide), for 30 min at 500 rpm and

20°C. The reaction mix was diluted with 0.1% formic acid followed by stage-tip purification. The sample was dissolved in 2 µL 0.2 M phosphate buffer (pH 7.4), and 1 µL of HP (TLC Pharmaceutical Standards) dissolved in water (4.16 mM) was added. The final reagent concentrations were 0.07 M Fmoc-GST and 1.4 mM HP. The solution was incubated at 500 rpm at 20°C and 0.5 µl aliquots were taken after 2, 4.5, 20, 24 and 28 h. Each aliquot was diluted with 0.1% formic acid followed by stage-tip purification. The samples were dissolved in 30% acetonitrile (ACN), 0.1% TFA and mixed mix with a saturated alpha-cyano-4-hydroxycinnamic acid (HCCA) solution in 30% ACN in a volumetric ratio of 1:1. Samples were pipetted onto a matrix-assisted laser desorption/ionization - time-of-flight (MALDI-TOF) sample plate (MTP 384 ground steel, Bruker) and dried. Sample composition was analysed via MS measurement on an Autoflex III Smartbeam MALDI-TOF instrument (Brucker Daltonics). The laser intensity was set to 70% and the mass range was 600-2000 Da.

## 2.2 Fragmentation analysis and software identification of transglutaminase cross-linked peptides

### 2.2.1 LC-MS analysis

An enzymatic transglutaminase cross-linking reaction with the peptides Boc-EGGKGPR and Boc-SQDGGR was performed by following the protocol from Section 2.1.1.2. After the reaction was terminated, the sample was split into two equal aliquots followed by deprotection and stage-tip purification. Each aliquot was then dissolved in 15 µl 0.1% formic acid. For each LC-MS run, 1.5 µl of the product was diluted with 0.1% formic acid to a final volume of 15 µl and centrifuged for 3 min at 16,000 x g. 11 µl were transferred into an MS

vial and 10 µl (10 nmol peptide) was used for LC-MS analysis. Samples were loaded using the autoinjector unit of a nanoAcquity™ ultra performance LC (Waters). Each sample was first loaded onto a pre-column (nanoAcquity™ 10K 2G V/M trap column, C18, 5 µm, 180 µm x 20 mm) with a flow rate of 10 µl/min. Peptides were then separated using a separating column (Nikkyo Technos Co.Ltd., C18, 5 µm, 100 µm x 15 cm) at a flow rate of 0.4 µl/min. Separations were performed at 20°C and the pressure range was 900-1000 psi. The composition of solvent A was aqueous 0.1% formic acid and the composition of solvent B was 0.1% formic acid in ACN. The solvent gradient was 1% B (0-1 min), 1-60% B (1-15 min), 60-90% B (15-15.01 min), 90% B (15.01-20 min), 1% B (20-30 min). MS analysis was performed using an LTQ Orbitrap Velos instrument. The MS1 m/z scan range was 300-2000 Da. The 20 most abundant precursor ions from each MS scan were isolated for MS/MS analysis and fragmented via collision-induced dissociation using helium gas. MS2 fragmentation was performed using collision induced dissociation (CID) or higher-energy C-trap dissociation (HCD) using a different collision energy in each run (30, 35, 40, 50, 60, 70, 80, 90 and 100). MS1 measurements and HCD fragmentation analysis were performed in the orbitrap with a mass accuracy window of ± 20 ppm while CID fragmentation analysis occurred in the linear ion trap with a mass accuracy window of ± 0.8 Da.

### 2.2.2 Software identification

The LC-MS results from the experiments with the ramped CID voltages were analysed using MassMatrix software [123]. The search was performed against a mouse collagen-1(I) amino acid sequence (source: Uniprot, downloaded on 23.04.2018, accession number: P11087). The cleavage

enzyme was trypsin with a maximum of 2 missed cleavages per peptide. No variable or fixed modifications were added. A decoy search was performed with the reversed protein sequence. The minimum and maximum peptide lengths were set to 6 and 40 amino acids, respectively. The minimum output pp and pp2 values were 5 and the minimum $pp_{tag}$ value was 1.3. The precursor and fragment mass tolerance were set to ± 20 ppm and ± 0.8 Da respectively. The search of a transglutaminase cross-link was performed by manually adding a new cross-link to the software database between a lysine and a glutamine residue with an ammonia loss (a mass shift of -17.03 Da), which results from the transglutaminase cross-linking reaction. The search was performed using the Exploratory Search mode and the maximal number of cross-links per peptide was set to 1. After the search was completed the results appeared with an automatic false discovery rate of 0%.

The same LC-MS data were also searched using MeroX software (version 2.0.0.3) that now contains StavroX – a software for identification of peptides with non-cleavable cross-links [124]. The set cleavage sites were lysine and arginine with a maximum of 2 missed cleavages. The minimum and maximum peptide length were set to 5 and 30 amino acids, respectively. No variable or fixed modifications were defined. A custom transglutaminase cross-linker was created *in silico* between a lysine and a glutamine residue with a single $-NH_3$ (ammonia) loss and a maximum distance of 0 Å. The defined precursor and fragment ion precisions were 20 ppm and 0.8Da respectively. The mass limits were set at 200 to 6000 Da and the signal-to-noise ratio to 2. The minimum number of fragment ions per peptide was 3 and the minimum charge of the MS1 precursor was 2. A prescore of 10%, a 5% false discovery rate (FDR) cut off and a score threshold of 50 were selected. The scoring speed was adjusted to standard scoring.

## 2.3 Cross-linking of purified proteins by transglutaminases

### 2.3.1 Fibronectin cross-linking

60 µg (1 mg/ml) of purified human fibronectin dissolved in 0.05 M Tris buffered saline and 20 mM $CaCl_2$ (pH 7.5) was mixed with 80 mU guinea pig TG2 (1.25 mU/µl). The solution was incubated at 37°C and 1000 rpm for 6 hours. To monitor the reaction progress, 1 µl aliquots were taken every hour. Each aliquot was mixed with loading buffer to the final concentration of 62.5 mM Tris-HCl, 2.5% SDS, 0.002% Bromophenol blue, 10% glycerol, 0.7135 M β-mercaptoethanol and incubated at 95°C for 5 min. Protein separation was performed via SDS-PAGE (Laemmli, [125]) using a 3-8% Criterion™ Tris-Acetate Protein Gel (Bio-Rad) on a Bio-Rad electrophoresis chamber. The electrophoresis conditions were 200 V, 100 mA and 100 W. The gel was fixed in 40% ethanol containing 10% acetic acid for 15 min and rinsed with distilled water for 5 min. The gel was shaken overnight in colloidal Coomassie (QC colloidal Coomassie Stain, Bio-Rad) at room temperature. Destaining was achieved via incubation with distilled water until a clear background was observed.

Another cross-linking reaction was performed by introducing several changes of the reaction conditions of the above described procedure. The temperature was lowered to 20°C and two separate reactions were performed with two different TG2 amounts: 2.5 mU (0.04 mU/µl) and 1.25mU (0.02 mU/µl). The total incubation time lasted 30 min and the progress of the reaction was monitored by taking an aliquot every 5 min. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis was performed as described above.

### 2.3.2 Fibrinogen cross-linking

#### *2.3.2.1 Fibrinogen cross-linking by guinea pig TG2*

The cross-linking reaction of purified human fibrinogen (Thermo Fisher) with guinea pig TG2 was performed in 0.4 M Tris-HCl buffer containing 20 mM $CaCl_2$ and 1 mM dithiothreitol (DTT) (pH 7.2). The fibrinogen amount used for the reaction was 7.8 µg (0.078 µg/µl) and the TG2 amount was 60 mU (0.6 mU/µl). The incubation was performed at 37°C in an orbital mixer at 1000 rpm for 70 min. 13 µl (1 µg protein) aliquots were taken after every 10 min of incubation. The protein composition of each aliquot was analysed via SDS-PAGE and Coomassie staining (Section 2.3.1) with a 4-15% TGX$^{TM}$ Precast Protein Gel (Bio-Rad). Selected gel bands were excised, cut into 1 $mm^2$ pieces and each band was transferred to an Eppendorf tube. The gel pieces were destained in 200 µl 25mM ammonium bicarbonate, 50% ACN for 30 min at 37°C and 1000 rpm. The samples were briefly centrifuged and the solution was removed, followed by repeating the destaining procedure for 5 min. The gel was rehydrated using three washes in 100 µl 25 mM ammonium bicarbonate for 5 min at 25°C and 1000 rpm. Samples were then treated with Peptide:N-glycosidase F (PNGase F, New England Biolabs, 500 U per sample) in 100 µl 25 mM ammonium bicarbonate (pH 8) at 37°C, 1000 rpm for 16 h. After brief centrifugation the samples were incubated with 10 mM DTT for 2 h at 37°C and 1000 rpm, followed by 25 mM iodoacetamide (IAA) for 30 min at room temperature in the dark. The solution was removed from the gel, followed by a washing step using 400 µl 25 mM ammonium bicarbonate and 400 µl ACN (10 min, 1000 rpm, 25°C). This procedure was repeated once. The samples were dried in a vacuum centrifuge (45°C, 15 min) and rehydrated in 100 µl 25 mM ammonium bicarbonate. The samples were digested with 0.08 µg endoproteinase Lys-C (Wako) for 2 h at 37°C and

1000 rpm and 0.08 µg trypsin (Promega) overnight at 37°C and 1000 rpm. After brief centrifugation, the solution was pipetted into a new tube and extraction of residual peptides from the gel was performed twice, each using 100 µl 5% formic acid (25°C, 1000 rpm, 15 min). The extracts were added to the corresponding solutions and the samples were evaporated on the vacuum centrifuge at room temperature. In order to estimate the peptide concentration each sample was dissolved in 20 µl 0.1% formic acid. 12 µl of each sample was diluted with 200 µl 0.1% formic acid and purified using a stage tip. Each eluate was dissolved in 60 µl 0.1% formic acid, centrifuged (3 min, 16,000x g) and 5 µl were used for LC-MS measurement. Samples were transferred to mass spectrometry (MS) vials (Thermo) that were then loaded into the system autoinjector unit of a nanoAcquity$^{TM}$ ultra performance LC (Waters). Each sample was first loaded onto a pre-column (NanoAcquity$^{TM}$ 10K 2G V/M trap column, C18, 5 µm, 180 µm x 20 mm) and peptides were separated using a separating column (Nikkyo Technos Co.Ltd., C18, 5 µm, 100 µm x 15 cm) at a flow rate of 0.4 µl/min. Separations were performed at 20°C and the pressure range was 900-1000 psi. Peptide elution was achieved using the following gradient: 1% B (0-2 min), 1-6% B (2-6 min), 6-31% B (6-79 min), 31-60% B (79-92 min), 60- 90% B (92-93 min), 90% B (93-98 min), 1% B (98-120 min). The composition of solvent A was aqueous 0.1% (v/v) formic acid and the composition of solvent B was 0.1% (v/v) formic acid in ACN. The LC-MS runs of each sample were followed by a series of washing steps with the following solutions: 0.1% (v/v) trifluoroacetic acid (5 µl), 10% (v/v) formic acid (10 µl), 0.1% (v/v) trifluoroacetic acid (10 µl), 1% (v/v) $NH_3$ in 50% (v/v) ACN (10 µl), 80% (v/v) methanol (10 µl), 80% (v/v) ACN (10 µl) as described by Fang et al [126]. Peptide identification was achieved using mass spectrometry. The 20 most abundant precursor ions from each MS scan were isolated for MS/MS

analysis and fragmented via collision-induced dissociation (CID 30) using helium gas. The m/z scan range was 400-2000 Da.

Protein identification was performed with the aid of MaxQaunt software (version 1.5.5.1). MS and MS/MS spectra were compared with theoretical spectra generated from tryptic in-silico digests of human protein sequences with no more than two missed cleavages. The following amino acid modifications that may occur naturally or arise during sample preparation were included in the search: carbamidomethylation [C] (fixed modification); carbamylation (N-terminal), deamidation [N], Gln → pyro-Gly (N-terminal), oxidation [M], oxidation [P], and acetylation (N-terminal) (variable modifications). Protein quantification was performed by calculating the iBAQ (intensity-based absolute quantification) value, which equals the sum of the peak intensities of all peptides belonging to one protein divided by the number of theoretically observable peptides [127]. Thus, the iBAQ value of each protein is designed to be proportional to its molar amount. The data were searched against a whole human protein database (source: Uniprot, downloaded on 19.07.2016, proteome ID: UP000005640).

Further fibrinogen cross-linking reactions were performed using a reduced amount of TG2. Prior to the reactions an SDS-PAGE analysis of the TG2 enzyme solution was made to assess its purity. 6 µl TG2 stock solution was mixed with 94 µl 0.4 M Tris-HCl buffer and 13 µl was used for the analysis (same TG2 amount as in the cross-linking reaction from above). In addition, a serial dilution of the prepared solution was made at volumetric ratios of 1/2, 1/5, 1/10, 1/25, 1/50 and 1/100. SDS-PAGE and Coomassie staining were performed according to the procedure described above.

Based on these results, which showed the absence of any contaminant bands at the 1/25 and 1/50 dilutions, another fibrinogen cross-linking reaction was performed with a TG2 amount (2.4 mU, 0.024 mU/µl) that was 25 times

lower than in the previous reaction, followed by SDS-PAGE analysis. Two cross-linking reactions were conducted with different incubation times: in the first reaction an aliquot was taken after every 10 min with a maximal incubation time of 110 min, while in the second reaction the total incubation time was extended to 4.5 h. Analysis of the reaction progress started after 2 h of incubation by taking an aliquot after every 15 min.

#### *2.3.2.2 Fibrinogen cross-linking by Factor XIIIa*

0.5 µg of human Factor XIIIa (Haematologic Technologies) was added to a solution of 4 mg/ml (200 µg) of human fibrinogen in 0.05 M Tris-HCl and 10 mM $CaCl_2$ (pH 7.6) to a final concentration of 10 µg/ml (0.56 mU/µl). The reaction mixture was incubated at 37°C for 110 min. The progress of the reaction was monitored by taking 1 µl aliquots every 10 min.

Another enzymatic reaction was performed using a 10-fold increase of the Factor XIIIa concentration (100 µg/ml; 5.6 mU/µl), while the rest of the reaction conditions remained identical to the previous reaction.

SDS-PAGE analysis of the reaction aliquots was performed as described in Section 2.3.1.

Another fibrinogen cross-linking reaction was performed in the presence of thrombin. 140.4 µg fibrinogen, dissolved in 0.01 M sodium phosphate, 0.01 M citric acid, 5 mM $CaCl_2$, 3 M NaCl, pH 7.3 (final concentration 7.02 µg/µl) was mixed with 0.1U (5 mU/µl) bovine plasma thrombin (RayBiotech) and 0.43 U (0.02 U/µl) Factor XIIIa. The reaction mixture was incubated for 18 h at 37°C. The solution was then briefly centrifuged and the Factor XIIIa activity was terminated by adding ethylenediaminetetraacetic acid (EDTA) to the final

concentration of 4.5 mM. The supernatant was carefully pipetted from the fibrin clot.

#### *2.3.2.3 Digestion and cross-linking analysis of the fibrinogen cross-linking products*

The product bands resulting from SDS-PAGE analysis of the cross-linking reactions of fibrinogen with TG2 and Factor XIIIa were excised and cut to 1 $mm^3$ cubes. Bands of the same apparent molecular weight from three different aliquots were pooled together followed by in-gel digestion. Gel pieces were then destained with 200 µl 25 mM ammonium bicarbonate on an orbital mixer (37°C, 30 min, 1000 rpm). The solution was removed and replaced with 100 µl ACN, followed by a second incubation for 5 min at 25°C and 1000 rpm. After solvent removal, the buffer was exchanged with 1x buffer 1A (5 mM sodium acetate, 2.5 mM $CaCl_2$, 25 mM NaCl, pH 6.8) by incubation with alternating buffer and ACN (100 µl, same conditions as above) until four washes had been made. Residual ACN was evaporated on the vacuum centrifuge (15 min at room temperature). Samples were then deglycosylated with chondroitinase ABC (Sigma, 0.05 U), heparitinase II (Amsbio, 0.05 U) and keratanase (Sigma, 0.125 U) in 100 µl 1x buffer 1A for 16 h at 37°C and 1000 rpm. Deglycosylation was completed by adding 0.05 volumes of 200 mM ammonium hydrogen carbonate and incubating with PNGase F (500 U, 4 h, 37ºC, 1000 rpm). Samples were then treated with 10 mM DTT (2 h, 1000 rpm, 37ºC), followed by 25 mM IAA (30 min, room temperature, in the dark), and finally washed with 1x buffer 1A and ACN (as above) and rehydrated in 100 µl buffer 1A. Protein digestion was achieved using 0.5 µg Lys-C (2 h, 37ºC, 1000 rpm) and 1 µg trypsin (37ºC, 1000 rpm, overnight). The resulting solution was pipetted to a new tube and residual

peptide was extracted three times with 40 µl 1x buffer 1A and three times with 5% formic acid (5 min, 1000 rpm, 25ºC). Each 5% formic acid extract was evaporated in a vacuum centrifuge at room temperature and combined with the other extracts from that sample. Each pooled extract was acidified with 0.1% formic acid to pH ≤ 3 and purified using a stage tip. To estimate peptide concentrations, samples were dissolved in 20 µl 0.1% formic acid and labelled the as shown in Chapter 3, Figure 28a. The samples were prepared and analysed using LC-MS followed by a MaxQuant search (as described in Section 2.3.2.1). 5 µl was used for each measurement.

Preparation of the product formed after Factor XIIIa cross-linking in the presence of thrombin (Section 2.3.2.2) was performed by removing the supernatant from the fibrin clot and adding 30 µl of denaturation buffer (8 M urea, 10 mM ammonium hydrogen carbonate). The sample was reduced and alkylated according to the previously described procedure. The urea concentration was lowered to 2 M by adding 90 µl 50 mM ammonium hydrogen carbonate followed by proteolytic digestion with Lys-C (2 µg) and trypsin (4 µg). A second portion of trypsin (2 µg) was added, followed by incubation for 2 h at 37°C and 1000 rpm. The sample was diluted with 0.1% (v/v) formic acid to pH ≤ 3 and purified with a stage tip. 1 µl of each sample was diluted with 0.1% formic acid in the ratio of 1:75, centrifuged (3 min, 16,000x g) and 5 µl (0.6 µg) was analysed using LC-MS.

A MassMatrix search for transglutaminase cross-linked peptides was performed with the FXIIIa and TG2_2 samples. The human fibrinogen gamma amino acid sequence (Uniprot, downloaded on 12.04.2019, accession number: P02679) was used for the search. The following modifications were added: carbamidomethylation [C] (fixed modification); and oxidation [M] and deamidation [Q] (variable modifications). All other parameters were the same as in Section 2.2.2.

A parallel search with these data using the MeroX software was performed using the same protein database and modifications as used for the MassMatrix searches. The prescoring option was deselected and the score threshold was set to -1. The rest of the search parameters remained identical to the ones given in Section 2.2.2.

## 2.4 Extracellular matrix (ECM) proteomics of human dermal fibroblast cell cultures and mouse skin tissue

### 2.4.1 Isolation of ECM proteins from human dermal fibroblast cell cultures

#### *2.4.1.1 Cultivation of primary dermal fibroblast cultures*

Primary dermal fibroblasts from a 50-year old female donor were cultured in a Dulbecco's Modified Eagle Medium (Gibco) containing 10% FBS (Labtech), 100 U/ml penicillin, 0.1 mg/ml streptomycin. Incubation conditions were 37°C, 95% air humidity, and 5% $CO_2$ content. Cells were passaged at 80% confluence. For the ECM preparation, cell cultures were allowed to reach 100% confluence, the medium was then exchanged for serum-free medium in which cells were incubated for 48 h. Cells between the 4th and the 7th passage were used for the experiments described in the following sections.

#### *2.4.1.2 Isolation of ECM from primary dermal fibroblast cultures*

##### *2.4.1.2.1 Trypsin method*

For this experiment cell cultures were grown in two 25 $cm^2$ flasks. After the serum-free medium was removed, cell cultures were washed three times with 4 ml PBS buffer (20°C), followed by minimal exposure to 0.25% trypsin-EDTA solution to detach the cells (3 min at 20°C). The cell suspension was removed and cell culture flasks containing ECM proteins were washed four times with phosphate-buffered saline (PBS) buffer (20°C, Gibco). 2 ml 8 M urea was added to each flask, followed by shaking (2 h, 160 rpm, 20°C). The proteins were scraped from the cell culture flasks and the solutions were concentrated by centrifugation through a 3 kDa-MWCO filter (Merck Millipore Ltd.) at 14,000x g and 20°C (Eppendorf centrifuge 5415D) until an approximate volume of 40 µl was reached. Samples were collected into tubes by inversion and centrifugation (4 min, 1,000x g). The solutions were diluted 4-fold with wash buffer to a final urea concentration of 2 M.

##### *2.4.1.2.2 Detergent method*

Confluent cell cultures were prepared as above in two 25 $cm^2$ flasks. The serum-free medium was removed from each flask, followed by four washes with 4 ml PBS buffer (20°C). Cell cultures were then shaken in 1 ml decellularisation buffer (1 mM EDTA, 25 mM $Na_2HPO_4$, 0.1 M $Na_2CO_3$, 1% (w/v) Tween 20, 0.5% (w/v) sodium deoxycholate, pH 11) for 1.5 h, at 160 rpm and 20°C. The decellularised samples were scraped from the cell culture flasks, transferred into tubes and centrifuged for 5 min at 12,000 rpm. Supernatants were carefully aspirated with a pipette until approximately 20 µl was left in the tube and the pellets were then washed three times with

wash buffer (5 mM sodium acetate, 2.5 mM $CaCl_2$, 25 mM NaCl, pH 6.8) by adding 1 ml to the pellet, shaking for 5 min at 1,000 rpm and 20°C (Eppendorf Thermomixer), spinning for 5 min at 12,000 rpm and removing the supernatant with a pipette, leaving approximately 20 µl buffer in the tube. After the last washing step, the remaining buffer was evaporated in a vacuum centrifuge at 20°C. 50 µl wash buffer was added to each sample, followed by incubation with collagenase (working concentration: 1 U/µl) for 24 h at 1000 rpm and 37°C. Digestion was terminated by adding 1.25 µl 0.1 M EDTA (final EDTA concentration: 2.5 mM).

*2.4.1.2.3 Modified detergent method*

Cells were grown in two 75 $cm^2$ flasks until 90% confluence was reached. After the serum-free medium was removed, the cell cultures were washed three times with 10 ml PBS buffer (20°C). 3 ml of decellularisation buffer containing 20 mM EDTA and a protease inhibitor cocktail (cOmplete™, Roche) at 1:1000 (v/v) was added to the cells. Flasks were shaken at 88 rpm and 20°C, and the progress of cell lysis was monitored under a microscope. After all the cells had lysed, the solution was carefully removed with a pipette and the flasks were washed five times with 10 ml PBS buffer (20°C). 4 ml 8 M urea was added to each flask and the ECM was scraped off the surface of the vessel. The sample volume was reduced to approximately 40 µl (Section 2.4.1.2.1). The urea concentration was lowered to 2 M by diluting with 1.3x wash buffer (1:4) followed by sample deglycosylation and digestion (Section 2.4.3).

In order to assess the effect of the 8 M urea extraction, another control sample was prepared using biological duplicates. Sample preparation was performed in the same way as described in Section 2.4.1.2.2 with the above

mentioned changes in terms of cell number and the addition of protease inhibitors during decellularisation.

After tryptic digestion, all 8 M urea extracts and control samples were fractionated on an OFFGEL system (Section 2.4.4).

A second set of 8 M urea extracts was prepared by introducing an additional purification step based on strong cation exchange (SCX) chromatography. After digestion, tryptic peptides were acidified with 0.1% (v/v) formic acid to achieve a final pH value of 2, followed by stage-tip purification using a tip with a binding capacity of 100 µg. Peptides were dissolved in 210 µl of a buffer containing 20 mM formic acid, 10% (v/v) ACN, pH 3.2. The resulting solutions were centrifuged (16,000x g, 5 min) and 200 µl was loaded onto a SCX column. SCX chromatography was performed on a SMART® system (Pharmacia Biotech) with a PC3.2/3 Mini S® precision column (column dimensions: 3.2 mm x 30 mm, particle size: 3 µm, charged group: $-CH_2-SO_3$, practical loading range: ≤200 µg protein). The composition of solvent A was 20 mM formic acid, 10% ACN, pH 3.2 and the composition of solvent B was 20 mM formic acid, 10% ACN, 1 M NaCl, pH 3.2. The flow rate was 100 µl/min and detection of the eluent was achieved via absorption measurements at 214 nm and 280 nm. The sample was loaded onto a 200 µl loop prior to injection onto the column (5 min after the method was started), which was followed by immediate collection of 100 µl fractions. Peptide elution was achieved using the following gradient: 0% B (0-10 min), 0-50% B (10-70 min), 50-100% B (70-75 min), 100% B (75-80 min). Fractions across each absorbance peak eluted within the linear gradient were combined to give 12 pooled fractions.

### 2.4.2 Isolation of ECM proteins from mouse skin samples

Two replicates of homogenized mouse skin (dorsal, frozen at -70°C) weighing approximately 75 mg per replicate were used. 1 ml decellularisation buffer was added to each replicate followed by incubation until the whole skin tissue was decellularised (3 h, 20°C, 1400 rpm). Replicates were pelleted by centrifugation for 2 min at 12000 rpm, and each supernatant was aspirated. Pellets were washed three times with wash buffer (Section 2.4.1.2.2) followed by lipid and detergent extraction in 200 µl 80% (v/v) ice-cold acetone (-20°C, overnight). Samples were then centrifuged (10 min, 15000x g) and the supernatants removed, followed by a second 80%-acetone extraction for 1 h at -20°C. The pellet was then washed once with 100 µl MS-grade water before 50 µl of wash buffer was added. Samples were incubated with collagenase (1 U/µl) for 24 h at 1000 rpm and 37°C and the collagenase activity was then quenched by the addition of EDTA to 2.5 mM.

### 2.4.3 Deglycosylation and tryptic digestion of ECM extracts

Proteins were deglycosylated with chondroitinase ABC (0.05 U), heparitinase II (Amsbio, 0.05 U) and keratanase (0.125 U) for 16 h at 37°C and 1000 rpm. After the addition of 0.05 volumes of 200 mM ammonium hydrogen carbonate (pH 8), samples were incubated with PNGase F (New England Biolabs, 1,250 U for fibroblast ECM, 2,500 U for skin ECM) for 4 h at 1000 rpm and 37°C. Reduction and alkylation were performed using 10 mM dithiothreitol at 1000 rpm and 37°C for 2 h, followed by 25 mM IAA at 20°C for 30 min in the dark. Proteins were digested using Lys-C (Wako, 1 µg for fibroblast ECM, 2 µg for skin ECM) by incubation at 1000 rpm and 37°C for 2 h, followed by addition of trypsin (Promega, 3 µg for fibroblast ECM, 6

µg for skin ECM) and incubation at 1000 rpm and 37°C overnight, and then a further addition of trypsin (1.5 µg for fibroblast ECM, 3 µg for skin ECM) and incubation at 1000 rpm and 37°C for 2 h.

### 2.4.4 OFFGEL peptide fractionation

After preparation of ECM protein digests from cell cultures and skin tissues, the resulting peptides were fractionated using liquid-phase isoelectric focusing (Agilent 3100). A pH gradient from 3 to 10 was employed, from which 12 fractions per replicate were obtained. OFFGEL fractionation was performed as described by the manufacturer using the following parameters: voltage – 4500 V, power – 200 mW, maximal current – 50 µA, energy – 22 kVh. After the isoelectric focusing was completed, liquid-phase fractions were removed to clean tubes. Peptides remaining in each well were then extracted using 0.1% formic acid for 1 h and added to the corresponding tubes.

Each fraction was purified with a stage tip (C18, binding capacity: 50 µg) as previously described for samples prior to LC-MS analysis (Section 2.1.1.2).

### 2.4.5 Mass spectrometry and MaxQuant protein identification

Purified fractions were dissolved in 25 µl 0.1% (v/v) formic acid followed by centrifugation for 3 min at 16,000x g. 20 µl of each supernatant was carefully aspirated and transferred to MS vials (Thermo) that were then loaded into the system autoinjector unit of a NanoAcquity™ ultra performance LC (Waters). The peptide concentration in each fraction was estimated via absorbance measurement at 280 nm (ε = 1). Based on the peptide concentration in each fraction, not more than 2 µg sample was used for each

LC-MS run. Each sample was first loaded onto a pre-column (NanoAcquity™ 10K 2G V/M trap column, C18, 5 µm, 180 µm x 20 mm) and peptides were separated using a separating column (Nikkyo Technos Co. Ltd., C18, 5 µm, 100 µm x 15 cm) at a flow rate of 0.4 µl/min. Separations were performed at 20°C and the pressure range was 900-1000 psi. Peptide elution was achieved using the following gradient: 1% B (0-2 min), 1-6% B (2-6 min), 6-31% B (6-79 min), 31-60% B (79-92 min), 60-90% B (92-93 min), 90% B (93-98 min), 1% B (98-120 min). The composition of solvent A was aqueous 0.1% (v/v) formic acid and the composition of solvent B was 0.1% (v/v) formic acid in ACN. The LC-MS runs of each replicate were followed by a series of washing steps (section 2.3.2.1). Peptide identification was achieved using mass spectrometry. The 20 most abundant precursor ions from each MS scan were isolated for MS/MS analysis and fragmented via collision-induced dissociation using helium gas. The m/z scan range was 400-2000 Da.

Mass spectrometry data were analysed using MaxQuant software (version 1.5.5.1). MS and MS/MS spectra were compared with theoretical spectra generated from tryptic in silico digests of human protein sequences with no more than two missed cleavages. The following amino acid modifications that may occur naturally or arise during sample preparation were included in the search: carbamidomethylation [C] (fixed modification); and carbamylation (N-terminal), deamidation [N], Gln → pyro-Gly (N-terminal), oxidation [M], oxidation [P], and acetylation (N-terminal) (variable modifications). Protein quantification was performed by calculating the iBAQ value. As previously described (Section 2.3.2.1), the iBAQ value of each protein is proportional to its molar amount.

## 2.5 Characterisation of systemic scleroderma (SSc) models

### 2.5.1 TGFβ-1 stimulated dermal fibroblast cell culture model

#### *2.5.1.1 Cell culture*

Human dermal fibroblasts from a 41-year old female donor (passages 2-5) were cultured in 75 $cm^2$ flasks until 100% confluence was reached (same culturing conditions as in Section 2.4.1.1). Cells were serum starved for 24 h followed by stimulation with 4 ng/ml TGFβ-1 (R&D Systems) in serum free medium for 5, 10 and 20 days, respectively. The medium was changed with fresh medium containing TGFβ-1 after every 2 days of incubation. Corresponding control cultures incubated with serum-free medium without TGFβ-1 were prepared in parallel. No biological replicates were prepared due to limited MS analysis time.

In a second experiment, similar cell cultures were grown in 75 $cm^2$ flasks until confluent and incubated in with TGFβ-1 for 20 and 30 days respectively. Biological triplicates were prepared for each condition (control and TGFβ-1 treated cells) of each time point experiment.

For TG2 activity assays, a suspension of the human dermal fibroblasts (donor details as above) in medium containing 10% (v/v) fetal bovine serum (FBS) was pipetted into a white Nunc MicroWell™ 96-well optical-bottom plate with a polymer base (Thermo Fisher). Approximately 4000 cells were seeded in each well. Cells were incubated in 10% (v/v) FBS until 100% confluence was reached. After serum starvation for 72 h, the medium was changed to serum-free medium containing TGFβ-1 (in 8 of the wells, conditions as above) or serum-free medium for the rest of the wells (3x8 wells). Cell stimulation lasted for 15 days. In addition, 30 days of TGFβ-1 treatment was performed with cell cultures grown in 25 $cm^2$ cell flasks

(conditions as above). The treated cells and corresponding controls were prepared in biological triplicates.

#### *2.5.1.2 ECM comparison between control and TGFβ-1 treated dermal fibroblasts using mass spectrometry based proteomics*

Cell cultures incubated with TGFβ-1 for 5, 10, 20 and 30 days and their corresponding controls (Section 2.5.1.1) were prepared for LC-MS proteomics measurements according to the procedure described in Section 2.4.1.2.2 with the introduction of several changes: the decellularisation buffer was supplemented with 20 mM EDTA (final concentration) and a protease inhibitor cocktail (cOmplete™, Roche) at 1:1000. After decellularisation of the samples from the 20- and 30-day TGFβ-1 incubation experiments, a 300 µl aliquot from the decellularisation supernatant of each replicate was transferred to fresh tubes and used for Western blot analysis.

After washing, residual detergent was removed from the ECM fraction with the aid of acetone extraction (Section 2.4.2) and the collagenase digestion was omitted from the protocol, as trypsin was found to be sufficient to digest the pellet. Sample preparation, MS analysis and protein identification otherwise followed the protocols described in Sections 2.4.3 – 2.4.5.

#### *2.5.1.3 Western blot analysis*

The 300 µl aliquots taken from the decellularised samples in the 20- and 30-day TGFβ-1 stimulation experiments and their corresponding controls (Section 2.5.1.2) were centrifuged for 5 min at 12,000 rpm. 200 µl of the supernatant was transferred to a fresh tube and 500 µl $H_2O$ was added to the pellet followed by vortexing and repeated centrifugation. 500 µl of the

supernatant was removed and the washing procedure was repeated twice. After the last washing step, the supernatant was aspirated until approximately 20 µl was left in the tube. 10 µl of 3x denaturing buffer was added to give final concentrations of 2.5% SDS and 62.5 mM Tris-HCl (pH 7.2). Samples were vortexed and incubated at 95°C for 5 min followed by centrifugation at 16,000x g for 3 min. 10 µl of the supernatant was used to estimate the protein concentration with the bicinchoninic acid (BCA) assay (VWR, Novagen) by using 1:50. The assay was performed by following the instructions of the manufacturer for a microplate procedure and the assay was performed on a 96-well plate (Costar). Each sample was diluted with water (1:50) and measured in technical duplicates. Protein concentration was estimated based on absorbance measurements at 562nm with the FLUOstar omega plate reader (BMG LABTECH). For this assay, the ECM extracts were diluted 1:50 with water, while the supernatants from the phase partitioning after decellularisation were measured at a 1:100 dilution. Based on the protein concentration results, 2 µg of ECM protein and 13 µg of the cell lysis supernatants were used for Western blot analysis. SDS-PAGE was performed as described in Section 2.3.2.1. After separation, the proteins were transferred from the gel onto a polyvinylidene fluoride (PVDF) membrane using the iBlot™ transfer stack, PVDF (Invitrogen, Thermo Fisher) via semidry transfer using an iBlot® Dry Blotting System (Invitrogen). The transfer conditions were 10 V and 9 min. For the TG2 detection the membrane was conditioned with tris-buffered saline-tween (TBST, 20mM Tris-HCl, 150mM NaCl, 0.1% TWEEN-20) buffer for 5min with rocking, followed by blocking with 5% milk powder in TBST for 1h at room temperature. The membrane was washed three times with TBST (5min with rocking at room temperature) followed by a primary antibody incubation. The membrane was incubated with the primary TG2 rabbit antibody (15100-1-AP, Proteintech) diluted 1:500 dissolved in of 5% milk powder in TBST at

4°C overnight. After the primary antibody incubation, the membranes were washed 3 times with TBST or 0.45mM TWEEN-20 in PBS respectively, followed by a secondary horseradish-peroxidase-linked antibody incubation (1:2000, #7074 Cell Signalling) in TBST or 0.45mM TWEEN-20 in PBS. The membranes were washed again three times followed by chemiluminescence detection with a luminol reagent (GE Healthcare) on the FluorChemM detector (ProteinSimple). The membrane was washed four times with TBST and incubated with stripping buffer (2% SDS, 62.5 mM Tris-HCl, 0.11 M β-mercaptoethanol) for 30 min at 50°C, followed by 6 washes with TBST and reblocking. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) detection was achieved using a GAPDH rabbit antibody (10494-1-AP, Proteintech) in a ratio of 1:2000.

Another SDS-PAGE and Western-blot analysis was performed to detect lysyl hydroxylase 2 using a 1:1000 diluted procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 (PLOD2)-specific rabbit antibody (21214-1-AP, Proteintech).

Quantitative comparison of the TG2 and PLOD2 levels between the control and TGFβ-1 treated replicates was achieved by determining the signal intensity based on the absorbance peak area using ImageJ software [128]. Prior to analysing statistical significance, TG2 or PLOD2 signal intensities in replicate blots were normalised to the corresponding GAPDH intensity. The statistical significance of the differences between the control and the TGFβ-1 treated sample protein levels was assessed using a one-tailed, heteroscedastic t-test.

#### *2.5.1.4 TG2 activity assay*

The TG2 activity assay was performed with cell cultures incubated in a 96-well plate (15-day TGFβ-1 experiment, Section 2.5.1.1). The medium was

removed from the cell cultures and exchanged for serum-free medium (blank); serum-free medium containing 10 mM EDTA (negative control); or serum-free medium containing 250 µM biotin-cadaverine (ThermoFisher) and 1 mM DTT (control and TGFβ-1 stimulated samples). Each condition was measured in eight biological replicates. Cells were incubated at 37°C for 2 h, and media were then removed and replaced with 2 mM EDTA in PBS buffer. The cells were decellularised with PBS buffer containing 0.1% (w/v) deoxycholate and 2 mM EDTA (5min at room temperature with rocking). The decellularisation buffer was removed and the plate was washed three times with TBST (5 min, room temperature, with rocking). The samples were the incubated with horseradish peroxidase (HRP)-conjugated streptavidin (ThermoFisher) diluted 1:1000 in TBST for 1 h at 37°C. The plate was washed again three times with TBST, and detection was performed by adding a luminol reagent (GE Healthcare). Chemiluminescence was detected at 440 nm (max. absorbance of luminol: 428 nm) on a FLUOstar omega plate reader (BMG LABTECH) using a gain setting of 4000.

The corresponding cells cultivated in 25 $cm^2$ cell culture flasks and stimulated with TGFβ-1 for 15 days were decellularised (as described in Section 2.5.1.2), followed by Western blot analysis of the TG2 and GAPDH levels (Section 2.5.1.3).

#### *2.5.1.5 Analysis of transglutaminase cross-linked peptides via MassMatrix*

A MassMatrix search was performed using the LC-MS data from the 30-day experiment with control and TGFβ-1 treated dermal fibroblasts. A separate search was performed with each fraction of each replicate. A human database was used containing the following proteins: collagen alpha-1(I),

collagen alpha-2(I), collagen alpha-1(III), collagen alpha-1(V), collagen alpha-2(V), collagen alpha-3(V), fibrinogen (α, β and γ) and fibronectin (Uniprot, downloaded on 19.04.2019, accession numbers: P02452, P08123, P02461, P20908, P05997, P25940, P02671, P02675, P02679, P02751). The selected variable modifications were deamidation of Q, oxidation of M and oxidation of P. The rest of the settings were identical with those given in Section 2.2.2.

### 2.5.2 TSK2 mouse model

#### *2.5.2.1 Histologic analysis of TSK2 mouse skin sections*

Dorsal mouse skin from 23-week-old control and TSK2 mice was cut into 3 mm slices and fixed in 10% formalin at room temperature for 80 days. Samples were then dehydrated in four ethanol solutions (70%, 80%, 95% and 100%) for 3 h, and incubated in three xylene solutions for 3 h each and 2 wax solutions, one for 3 h and one overnight, using a Leica TP1020. Samples were then embedded in wax using a Slee Mainza MPS/P1 and 4 µm sections cut using a Leica RM2135 microtome. Sections were dewaxed and rehydrated using 3 changes of xylene, 3 changes of alcohol (100%, 90% and 70%) and water (5 min, room temperature). Sections were stained with 1% (v/v) Harris Haematoxylin for 10 min, washed with water and differentiated in acid alcohol (1% (v/v) concentrated HCl, 70% (v/v) ethanol). Sections were then stained with 1% (v/v) aqueous eosin for 2.5 min followed by dehydration and mounting. Picrosirius-red staining was performed by incubation with 0.01% (v/v) Sirius Red F3B in saturated aqueous picric acid for 20 min.

#### *2.5.2.2 ECM analysis of the SSc mouse models TSK and DNR using mass-spectrometry-based proteomics*

For the ECM proteomic analysis of the TSK mouse model, homogenised skin tissue samples from 10- and 23-week-old female TSK2 mice and their corresponding controls were used. Each condition was measured in biological duplicates and each biological replicate was measured in technical duplicates. For the analysis of the TSK2 samples, 10 mg skin tissue per replicate was used, while for the DNR samples the amount was 30 mg per replicate. The DNR samples (23-weeks-old, female) were measured only in biological duplicates. The same experimental procedure was followed as described in Sections 2.4.2 – 2.4.5.

#### *2.5.2.3 Western blot analysis of the TSK2 and DNR mouse models*

For the Western blot analysis, the mouse skin samples prepared as described in Section 2.5.2.2 were used. As, Western blot analysis typically gives lower technical variations than proteomics experiments, only one technical sample for each biological replicate was prepared for the analysis. 1.5 mg of homogenised mouse skin was decellularised in 300 µl decellularisation buffer containing 20 mM EDTA and a protease inhibitor cocktail (cOmplete™, Roche) at 1:1000 (v/v) for 30 min at 1400 rpm and 20°C. The rest of the sample preparation and Western blot analysis was performed according to the procedure described in Section 2.5.1.3. For the SDS-PAGE analysis of the TSK2 samples, a 10% (w/v) Tris-HCl gel (Bio-Rad) was used, while the DNR samples were loaded on a 4-15% TGX™ Precast Protein Gel (Bio-Rad). 4 µl of each ECM extract and 5 µl of each supernatant solution were used for sample loading.

#### *2.5.2.4 Analysis of transglutaminase cross-linked peptides using MassMatrix*

MassMatrix searches of the LC-MS data obtained from the 10-weeks-old control and TSK2 samples were performed as described in Section 2.5.1.5 using a database containing the corresponding mouse protein sequences (Uniprot, downloaded on 19.04.2019, accession numbers: P11087, Q01149, P08121, O88207, Q3U962, E9PV24, Q8K0E8, Q8VCM7, P11276).

# Chapter 3

# Synthesis, characterisation and mass-spectrometric identification of cross-linked peptides

## 3.1 Introduction

Similar to other fibrotic diseases such as liver and lung fibrosis, a change in protein cross-linking pattern is a process that is also relevant to the development of skin fibrosis in systemic sclerosis (SSc) [52]. During SSc progression the skin undergoes changes in cross-linking, however this information is insufficient for developing an appropriate anti-fibrotic strategy as cross-linking is effected by diverse enzymes and acts on diverse proteins. Consequently, detailed information is needed to determine appropriate and specific therapeutic targets. For example, finding the positions of cross-links that arise only skin fibrosis could provide a valuable information for targeted anti-fibrotic therapy. For this purpose, mass spectrometry is currently the best method to determine the specific position of a given modification in different ECM proteins.

Mass spectrometry is a powerful tool for identifying peptide sequences from their fragmentation spectra. In proteomic studies this technique has been widely used to identify complex mixtures of proteins. It has also been used to analyse experimentally and naturally occurring cross-linked peptides in protein structure analysis experiments, which proved to be more challenging due to the higher complexity of the fragmentation spectra. Several approaches to identifying the positions of naturally occurring cross-links with the help of mass spectrometry have been made. These studies focused on

specific transglutaminase cross-links [78] and pyridinoline cross-linked peptides from tryptic collagen digests [81]. However, no comprehensive study of the identification and fragmentation behaviour of these species has been reported so far. A thorough investigation of the behaviour of naturally occurring cross-linked peptides in response to different fragmentation techniques could help develop a method to identify them with high confidence in complex protein digests. To achieve this, we hypothesized that the study of simple cross-linked peptide models with known sequences would help predict the fragmentation pattern of these species and thereby allow their identification in complex mixtures. The first aim of this chapter was to synthesize transglutaminase and hydroxylysylpyridinoline (HP) cross-linked peptides. We chose to study these types of cross-links as they are products of the enzymatic activity of lysyl oxidase, lysyl hydroxylase and transglutaminase - enzymes whose activity had previously been shown to be upregulated in fibrotic tissue [50, 53, 64]. Therefore, these cross-links were very likely to be upregulated in systemic sclerosis. The next aim was to use these peptides for different fragmentation experiments with various fragmentation methods (collision induced dissociation (CID) and higher-energy C-trap dissociation (HCD)) and collision energies that would help me develop an optimal fragmentation technique for cross-link identification. In addition to identifying series of linear and cross-linked fragments we planned to look for other potential fragments that could help confirm the identification of a cross-linked peptide. Finally, we aimed to develop an automated way to identify the cross-links of interest by optimising the use of bespoke software developed primarily for the analysis of chemically cross-linked peptides. This would enable me to search for naturally cross-linked peptides from complex samples such as ECM tryptic digests, which is the aim of the subsequent chapters of our thesis.

## 3.2 Results

### 3.2.1 Synthesis of transglutaminase cross-linked peptides

#### 3.2.1.1 *Chemical synthesis*

Transglutaminases cross-link peptides by catalysing the formation of an isopeptide bond between a lysine and glutamine residue [61]. For the chemical synthesis of this type of cross-linked peptide, a standard procedure for solid phase peptide synthesis was used. our objective was to form the isopeptide bond by chemically coupling the ε-amino group of a lysine residue with the side chain (γ) carboxyl group of a glutamic acid residue.

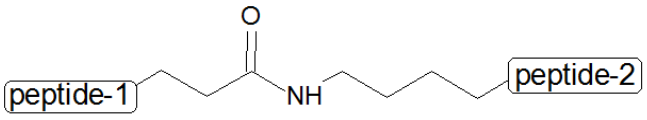



**Figure 3. 1 Structure of an intermolecular isopeptide bond between a lysine and a glutamic acid residue.**

For this reaction two peptides with the sequences YEG and AKT were used. For the choice of peptides, sequences that were hydrophilic enough to be dissolved in an aqueous solution for the MS analysis yet hydrophobic enough to allow the use of organic solvent during the synthesis were considered. The peptide containing a glutamic acid residue was attached to a resin that enabled a thorough wash to remove the excess activation reagents prior to coupling with the second peptide. Side reactions were prevented by protecting both N-termini with tertiary butoxycarbonyl (t-Boc) protective groups and all side-chain hydroxy groups with tertiary butyl (tBu) groups. The

glutamic acid containing peptide was attached to the resin via its C-terminus to direct the activation only towards the side-chain carboxyl group. At first, the activation was performed using (2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and N,N-diisopropylethylamine (DIEA) but this did not lead to any product formation (Figure 3.3). The activating reagent was then changed to carbonyldiimidazole (CDI) and this led to a positive result, as shown in Figure 3.2. A possible explanation why the two activation reagents delivered different results is that because the group attached to the C-terminus of the activated intermediate by CDI activation is smaller than the group induced by the HBTU reaction, CDI may have induced a more sterically favourable and therefore a more stable, activated peptide (Figure 3. 4).

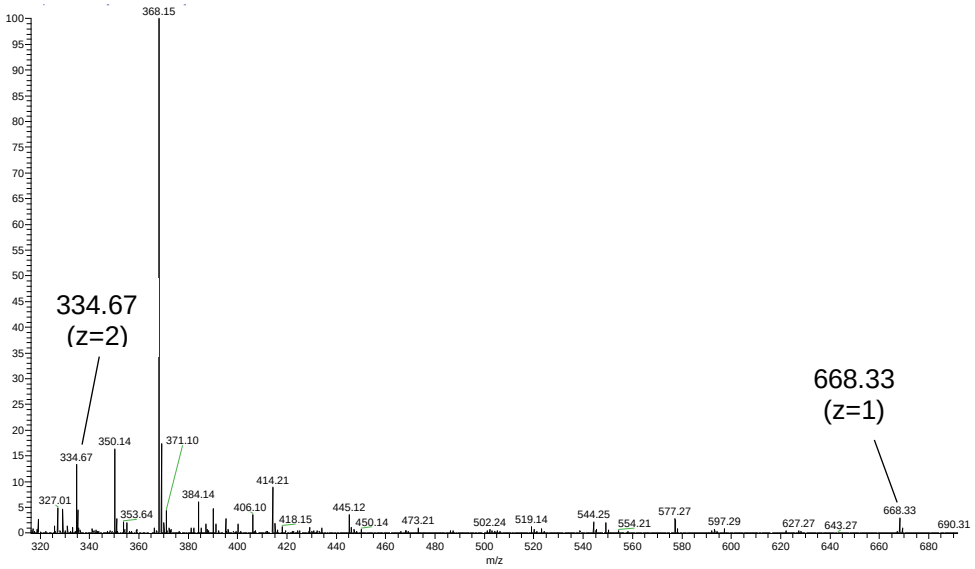



**Figure 3. 2 ESI-MS spectrum of the chemical transglutaminase-like, cross-linking reaction using the CDI activation method.** The indicated peaks with m/z values of 334.67 (z=2) and 668.33 (z=1) both arise from the cross-linked peptide (predicted m/z values: 668.32 (z=1); 334.70 (z=2)). The peak with the highest intensity at 368.15 Da (z=1) corresponds to the precursor peptide that was previously attached to the resin.

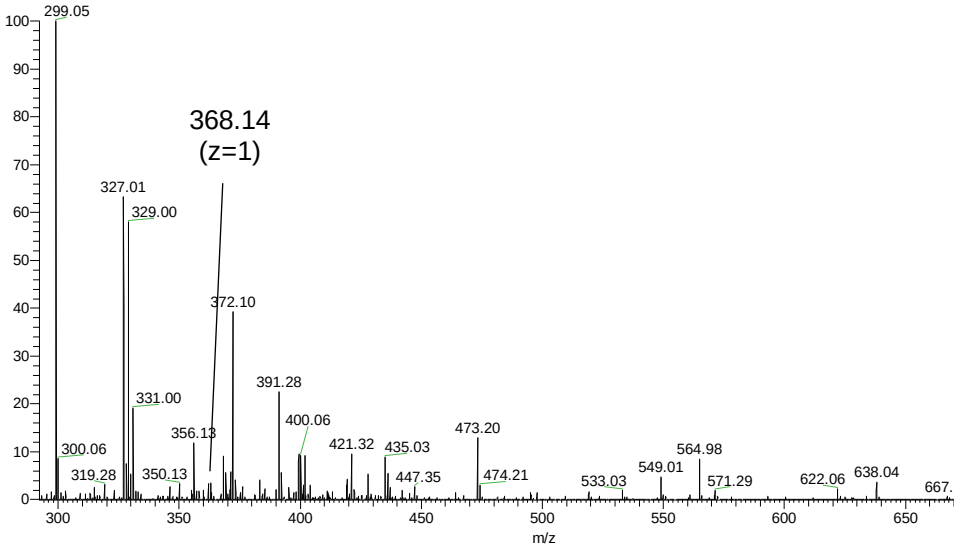



**Figure 3. 3 ESI-MS spectrum of the chemical, transglutaminase-like cross-linking reaction using the HBTU activation method.** The spectrum shows only the presence of the precursor peptide at m/z = 368.14.

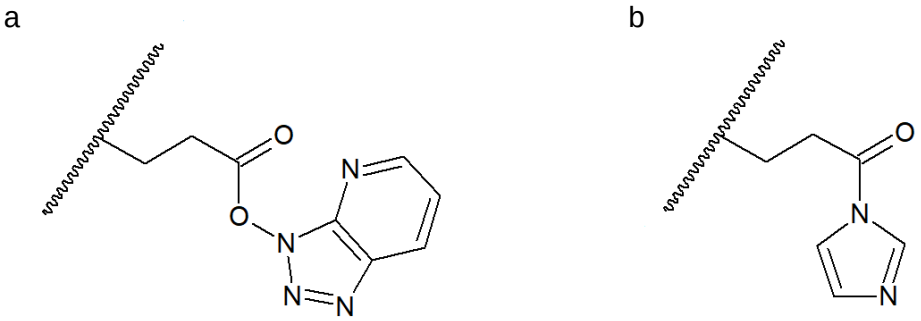



**Figure 3. 4 Activated side chain (γ) carboxyl group of a glutamic acid residue using HBTU (a) and CDI (b).**

In order to increase the product yield, the reaction with CDI was performed with an increased incubation time and the ratios between the product and the precursor intensity were compared. Table 3.1 shows that this change did not lead to an increase of the product yield. Increasing the concentration of the activating reagent might have helped reach higher product yields, however this was not possible under the conditions we employed, due to limitations in CDI solubility. Further reaction optimisation was not studied as the aim of this

project was only to produce a sufficient amount of product for isolation and fragmentation studies.

| *Incubation time [h]* | Precursor intensity (z=1) | Product intensity (z=1) | Product/precursor intensity ratio |
|---|---|---|---|
| *16* | $6.0 \times 10^6$ | $8.0 \times 10^5$ | 0.13 |
| *30* | $1.4 \times 10^6$ | $1.4 \times 10^5$ | 0.10 |

**Table 3. 1 Product intensity change after increasing incubation time.** The change in the product yield was analysed by calculating the ratio between the product and the precursor intensity for each reaction time.

### *3.2.1.1 Enzymatic synthesis*

Synthesis of transglutaminase cross-linked peptides was also performed via an enzymatic route. This was achieved by incubation of selected pairs of peptides with a guinea pig TG2 enzyme that catalyses the condensation of the ε-amino group of a lysine residue and the amide group of a glutamine residue side chain. Side reactions with the N-termini of the peptides were eliminated by the presence of t-Boc protective groups that were cleaved after the reaction and prior to MS analysis. Interestingly, we observed a strong dependence of the reaction rate on the sequence of the precursor peptides, as some reactions required longer incubation times and higher enzyme amounts in order to detect the final product.

#### *3.2.1.1.1 Cross-linked collagen I peptides*

Figure 3.5 shows the cross-linked product of peptides with sequences GGAQLGV and GAKGSA. The chosen peptides are fragments of type we collagen, that is a known target for TG2 [44]. There is no evidence in the

literature that leads to the hypothesis that these collagen sites are indeed cross-linked by the enzyme. The aim of this experiment was not to predict the fragmentation behaviour of a known cross-linked sequence but to find non-sequence specific MS fragments that would help identify unknown cross-linking sites. Therefore, the choice of peptide sequences was mainly based on their hydrophilicity that would allow a cross-linking reaction in an aqueous solution. In addition, a peptide library assay found that the presence of hydrophobic amino acids adjacent to the glutamine residue (e.g. leucine) accelerates the TG2 cross-linking reaction rate [129]. Thus, we hypothesized that the chosen glutamine containing peptide will contribute to the formation of a large amount of product. Indeed, even though a significant amount of precursor remained unreacted, the product yield appeared to be sufficient for the aims of this project and the protocol was not further optimised. More specifically, the peak intensity of the cross-linked peptide with the +2 charge state was sufficiently high for MS/MS isolation and fragmentation.

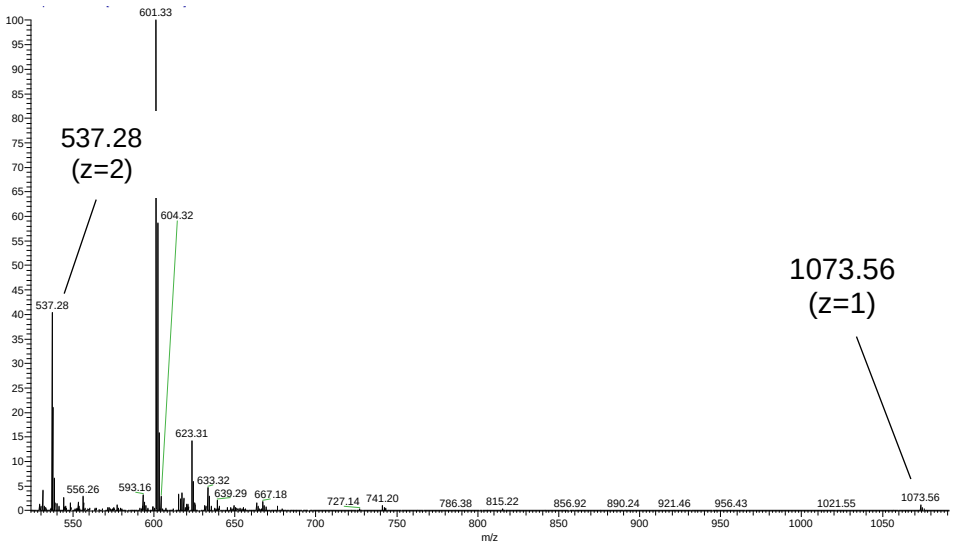



**Figure 3. 5 ESI-MS spectrum of the enzymatic reaction between the peptides GGAQLGV and GAKGSA with guinea pig TG2.** Product peaks were found at 537.28 Da (z=2) and 1073.56 Da (z=1). The predicted product m/z values are 537.27 (z=2) and 1073.54 (z=1). An unreacted precursor peak at 601.33 Da corresponding to the GGAQLGV peptide can also be seen. No peak is evident from the GAKGSA peptide as this has a m/z value below the range of the spectrum (m/z = 490.18 at z=1).

To further confirm that the cross-link position was indeed between the lysine and the glutamine residue and to obtain preliminary information about the fragmentation behaviour of a transglutaminase cross-linked peptide pair, the doubly-charged product was isolated and fragmented. Using standard instrument settings employed in the lab for fragmentation (CID 30) we was able to identify a series of cross-linked and linear fragments (Figure 3.6) from the most abundant peaks. These series had previously been observed by another group that used CID fragmentation to identify transglutaminase cross-linked peptides obtained from protein digests [80].

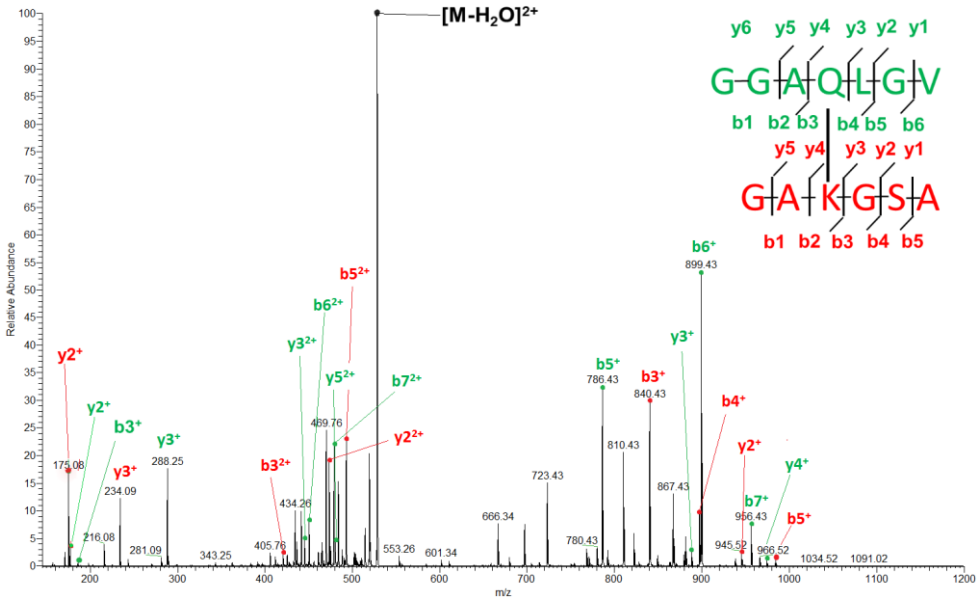



**Figure 3. 6 Fragmentation spectrum of a transglutaminase cross-linked peptide pair.** Collision induced dissociation (CID) with a standard collision energy of 30 was used. The insert indicates the positions at which fragmentation occurred in the cross-linked peptides to produce the b- and y-series ions labelled in the spectrum. The peak belonging to the precursor ion is labelled with M.

A similar fragmentation pattern showing the presence of linear and cross-linked fragment ions (data not shown) was obtained from the chemically cross-linked peptides (3.2.1.1 *Chemical* 0). These results are directly comparable with one study that used a similar instrument for peptide fragmentation and demonstrated the formation of a comparable set of linear

and cross-linked fragments using a slightly higher collision energy (CID 35) [80].

My results from two pairs of peptides, taken together with the results of other studies, suggested that our instrument and method would be able to identify the sequences of structurally-similar, unknown transglutaminase cross-linked peptides in subsequent studies in this project, although any large differences in physical properties, such as size and charge, were likely to affect their ionisation and fragmentation behaviour.

#### *3.2.1.1.2 Cross-linked tryptic collagen I peptides*

In the next stage of the method development, we decided to investigate the fragmentation of cross-linked peptides predicted to be produced by trypsin digestion of collagen I, a major protein in healthy and fibrotic skin ECM. For this purpose, tryptic peptide sequences were chosen from the collagen-alpha1(I) mouse sequence and chemically synthesised. Up to this point, our results had been limited to model cross-linked peptides with a total charge state of +2 at the working pH of approximately 2.7 (0.1% formic acid). Naturally occurring cross-linked peptides however are commonly found in higher charge states after trypsin digestion. As cross-linked peptides with higher charge states may show differences in their fragmentation behaviour compared with the previously-used model peptides, they were more appropriate for furthering the method development.

For this study, the peptide sequences EGGKGPR (893-899) and SQDGGR (1198-1203) were chosen from the mouse collagen 1α1 sequence. The criteria for peptide selection were (i) the presence of arginine instead of lysine at the C-terminus to avoid side reactions during the cross-linking; (ii) linear peptide sequence lengths of no more than 10 amino acids that would ensure

the efficient HPLC separation with a standard gradient used in our group prior to MS analysis, and (iii) sufficient hydrophilicity to enable solubilisation in water during the cross-linking reaction. The choice of sequences did not rely on predicted cross-linking positions, as there was no available literature from which we could have identified any known or probable cross-linking sites in other tissues or species. Accordingly, the main focus of this study was to investigate and characterise the general fragmentation behaviour of transglutaminase cross-linked peptides at higher charge states, rather than to analyse the fragmentation of a specific cross-link.

Unexpectedly, using identical reaction conditions to the previous enzymatic cross-linking experiment did not lead to any identifiable product formation. The product, termed TXLP, was instead detected after extending the incubation time from overnight to three days and adding 1.5-times more enzyme than in the previous reaction (Figure 3.7). Three different charge states of the cross-linked TXLP product were identified in the resulting spectrum.

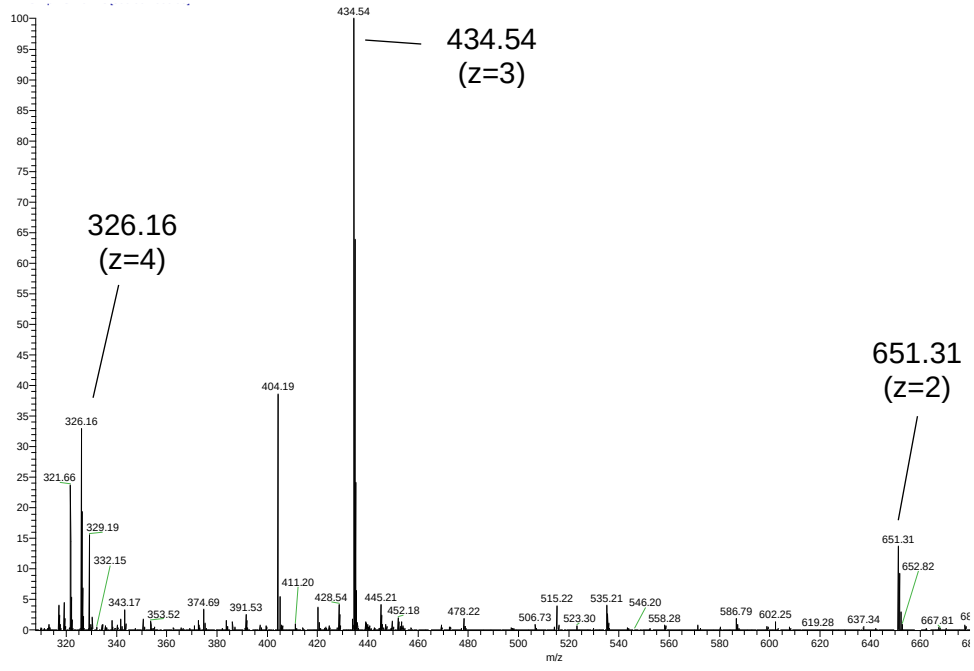



**Figure 3. 7 ESI-MS spectrum of TXLP synthesised using guinea pig TG2.** Product peaks were found at 326.16 Da (z=4), 434.54 Da (z=3) and 651.31 Da (z=2). No peaks were evident from the precursor peptides (m/z = 700.37 and 619.28 at z=1).

### 3.2.2 Synthesis of hydroxylysyl pyridinoline cross-linked peptides

The strategy for the synthesis of pyridinoline cross-linked model peptides was to couple synthetic peptides via commercially available hydroxylysyl pyridinoline (HP). The rationale for this approach (Figure 3.8) was to extend the three amino acid residues on HP using a standard peptide synthesis technique. The resulting structure would resemble a class of naturally-occurring, HP cross-linked peptides.

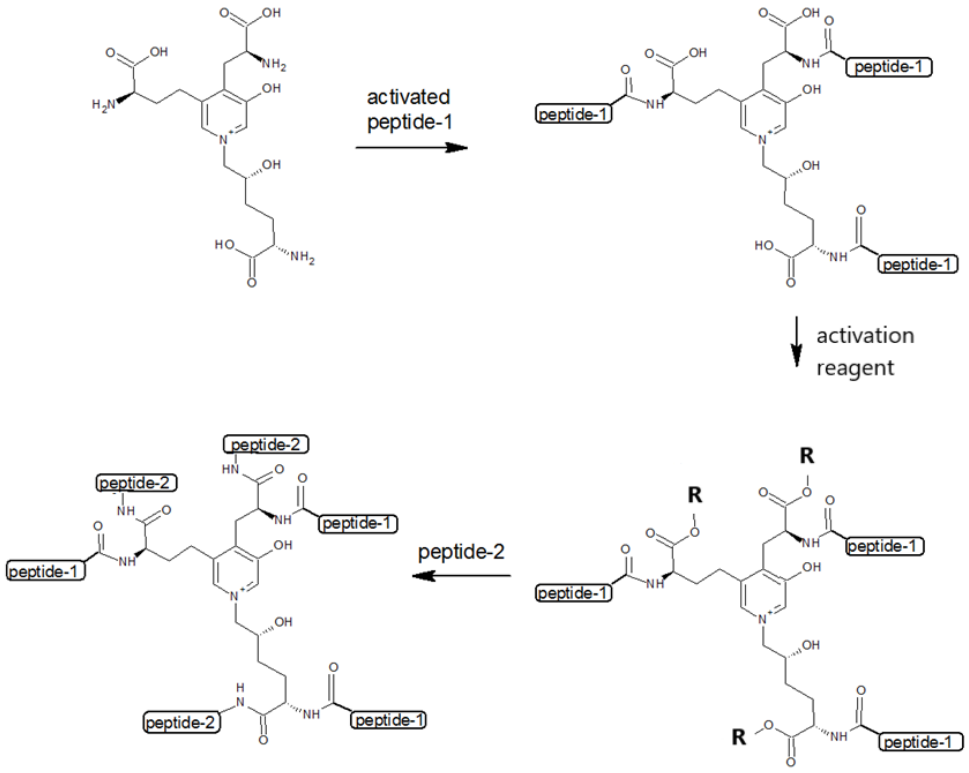



**Figure 3. 8 Strategy for the synthesis of a HP cross-linked tripeptide.** The plan consisted of a sequence of activation and coupling reactions. In the first step, the primary amine groups of the three amino acid residues on HP were coupled with an activated peptide (peptide-1) giving rise to the first intermediate. The N-terminus of peptide-1 was protected to suppress a side reaction. The three carboxyl groups of the first intermediate, equivalent to the amino acid C-termini, were then activated (activated groups are marked as R) to enable peptide bond formation between the second intermediate and the N-terminus of a second peptide (peptide-2).

In order to find a suitable medium for this reaction, the solubility of HP in different solvents was first tested. HP showed poor solubility in organic solvents widely used in peptide synthesis experiments, such as dimethyl sulfoxide (DMSO) and DMF (<0.2 mM), even at increased temperatures or after water bath sonication (data not shown). In contrast, HP demonstrated more hydrophilic properties, since it showed a good level of solubility (4 mM) in water without the need for any additional solubilisation techniques. This observation is consistent with previous literature that demonstrated a good level of water solubility for pyridinoline isolated from bovine bone after acid hydrolysis [130]. Dissolving pyridinoline in water allowed me to reach sufficiently high concentrations for the reaction compared with previously published peptide synthesis protocols [131]. Thus, a water-based peptide synthesis method was chosen for the experiment.

The sequences of the peptides used for the synthesis were an Fmoc-protected peptide with the sequence GST that would be used for the coupling to the amino acid residue N-termini of pyridinoline and a YSA peptide that would be attached to the C-termini. These sequences were again designed to include hydrophilic residues that would allow a good level of aqueous solubility. In the first stage of the synthesis, the objective was to attach the C-terminus of the Fmoc-GST peptide to the N-terminus of pyridinoline (the first reaction step in Figure 8). To avoid a side reaction, the N-terminus of the peptide was designed to include an Fmoc protective group. The peptide was first activated with EDC and sulfo-NHS according to the protocol described by G.T. Hermanson [131]. The activation was followed by a coupling reaction with HP. The course of the coupling reaction was monitored by analysing the sample composition at different incubation times using MALDI-TOF MS. Figure 3.9 shows the MALDI-TOF spectrum after 20 h of coupling. Multiple peaks were produced including one at 896.39 Da (monoisotopic,

deconvoluted mass data) corresponding to the mass of the first intermediate containing the GST peptide attached to HP.

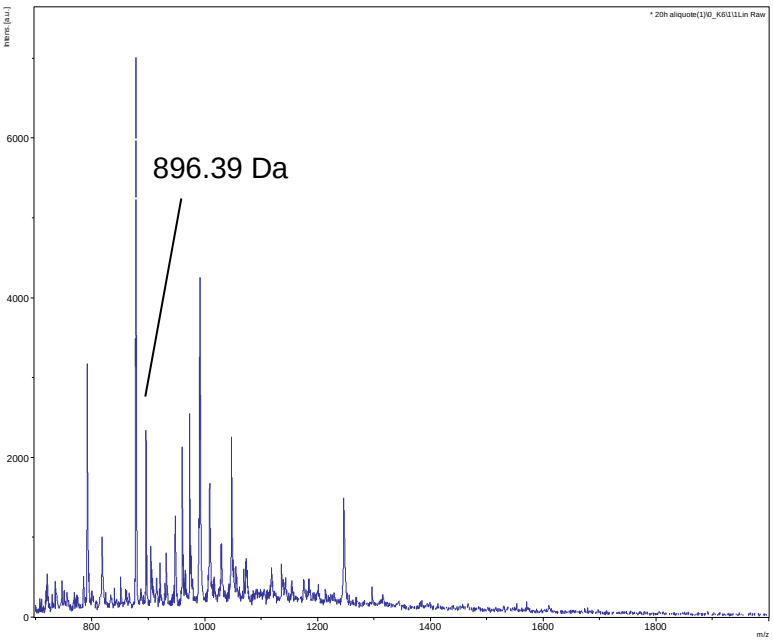



**Figure 3. 9 MALDI-TOF spectrum after a coupling reaction time of 20 h.** The indicated peak at 896.39 Da corresponds closely to the monoisotopic mass of the first intermediate (pyridinoline + one peptide), which is 896.37 Da (mass error: 0.02 Da). Standard mass error: 0.08 Da.

The formation of this intermediate was monitored by analysing the normalised peak intensity as a function of the reaction time. The total peak intensity at each time point was normalised to the intensity of a standard peptide that had been added at a constant level to each time-point aliquot prior to measurement. A diagram of the change in normalized peak intensities over time is depicted in Figure 3.10, where an increase of the intermediate during the first 24 h was observed. This is followed by a slight drop at 28h, which was probably due to the reaction of the intermediate with the second peptide to form the second intermediate (see below). We couldn't confirm this by detecting the peak of the second intermediate in the MALDI-TOF spectrum, as this seemed to be beyond the capacity of the aging instrument which was hampered by considerably reduced laser power. The calculated m/z value of the second intermediate is 1364.4 Da (z=1), while

the MALDI-TOF spectra showed only peaks in the lower mass range using samples and standards. Moreover, the results could not be further confirmed by showing a decrease of any of the precursors: pyridinoline did not bind to the stationary phase during stage tip purification and the N-terminally protected peptide does not carry a positive charge.

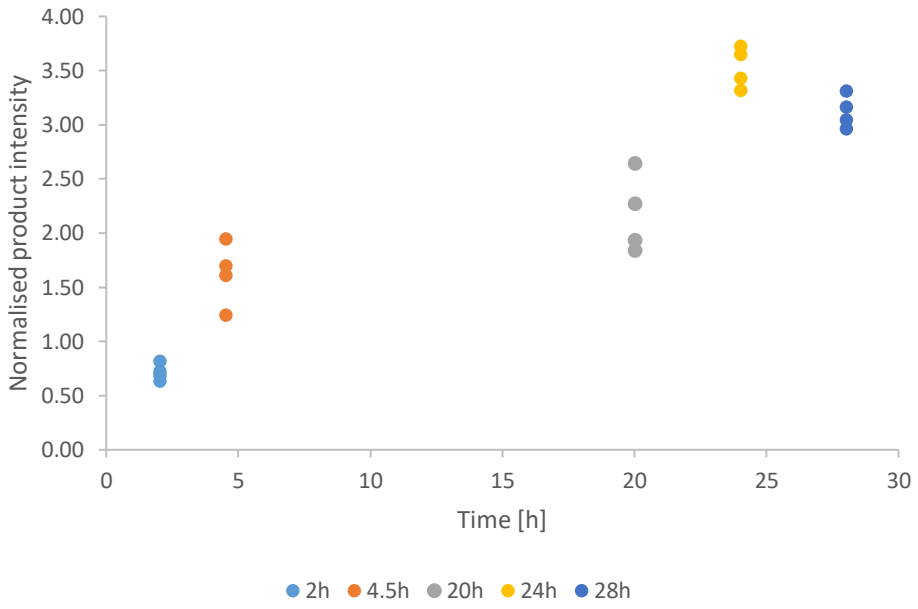



**Figure 3. 10 Development of the first intermediate product in the coupling of Fmoc-GST to pyridinoline.** (n=4)

Due to the restrictions of the MALDI-TOF instrument, further experiments were performed using an orbitrap electrospray ionisation (ESI)-MS system but this did not lead to the detection of a product (including intermediates). In order to investigate which reaction needed to be optimised, only the activation reaction was performed. Once again though, ESI-MS analysis did not show the activated peptide.

Because uncharged carbodiimides are faster activators than EDC (Dr P. Gaffney, personal communication), the peptide activation strategy was changed to using *N*,*N*′-diisopropylcarbodiimide (DIC) and *N*-hydroxysuccinimide (NHS) instead EDC/sulfo-NHS. The reason why sulfo-

NHS was omitted from the reaction is that it carries a negative charge. Thus, coupling of sulfo-NHS combined with an N-terminal protonation of the activated peptide prior to MS analysis would lead to a neutral compound that won't be detectable. Lower DIC concentrations than in previous publications were used [132] in order to maintain equimolar amounts of the peptide and the activation reagent. We reasoned that using excess activation reagent would have complicated the synthesis process, as purification from unreacted material would have been necessary prior to coupling. Nonetheless, lowering the DIC concentration did not lead to any detectable product formation. While extending the reaction time could have led to some product, this part of the project had become disproportionately time consuming and we did not proceed further due to the uncertainty of obtaining the required results and time limitations.

### 3.2.3 Fragmentation analysis and software identification of TXLP

#### *3.2.3.1 Fragmentation analysis*

##### *3.2.3.1.1 Comparison of CID and HCD fragmentation*

Following the ESI-MS analysis of TXLP, we wanted to confirm the identification of this ion and explore its fragmentation using ESI-MS/MS. Although most of the commonly used software for identification of cross-linked peptides are programmed to identify amino acid sequences from CID fragmentation spectra, HCD enables the formation of internal fragments that can't be seen in a CID collision experiment. Therefore, HCD fragmentation data could serve to confirm the type of crosslink in peptide sequences identified by CID fragmentation. As identification of cross-linked peptides within complex digests would prove arduous if all spectra had to be analysed,

one particular focus of the HCD analysis was to try to identify an internal fragment of the +4-charged peptide that could serve as a reporter ion for the presence of an isopeptide bond. Since HCD can lead to fragmentation at several sites within polypeptide chains giving rise to internal fragments [133], a +4-charged peptide could potentially fragment at each of the four peptide bonds next to the isopeptide bond, thereby forming a glutamine-lysine reporter fragment derived exclusively from the cross-linked residues (Figure 3.11a). We expected to find this fragment only in the MS2 spectrum of the +4-charged TXLP because only this charge state would allow fragmentation on all four sites that is required for its formation. If found in the fragmentation spectrum, this fragment could then be searched for in all fragmentation spectra independent of the surrounding amino acid sequence. This would accelerate and improve the accuracy of such searches. Another possible site of fragmentation was within the isopeptide bond itself, which would result in two linear fragments (Figure 3.11b). If the intensity of these fragments were sufficiently high, they could be isolated and fragmented in an MS3 experiment. Sequence analysis could be then performed with the aid of commonly used proteomics software such as MaxQuant or Mascot.

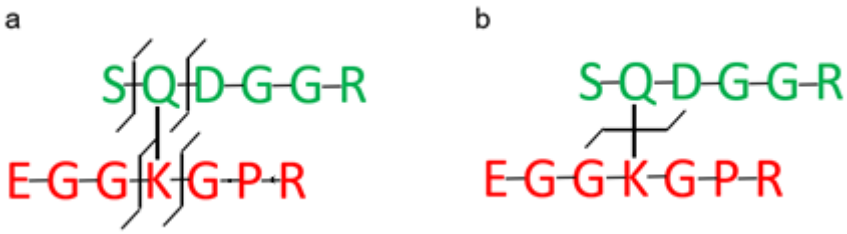
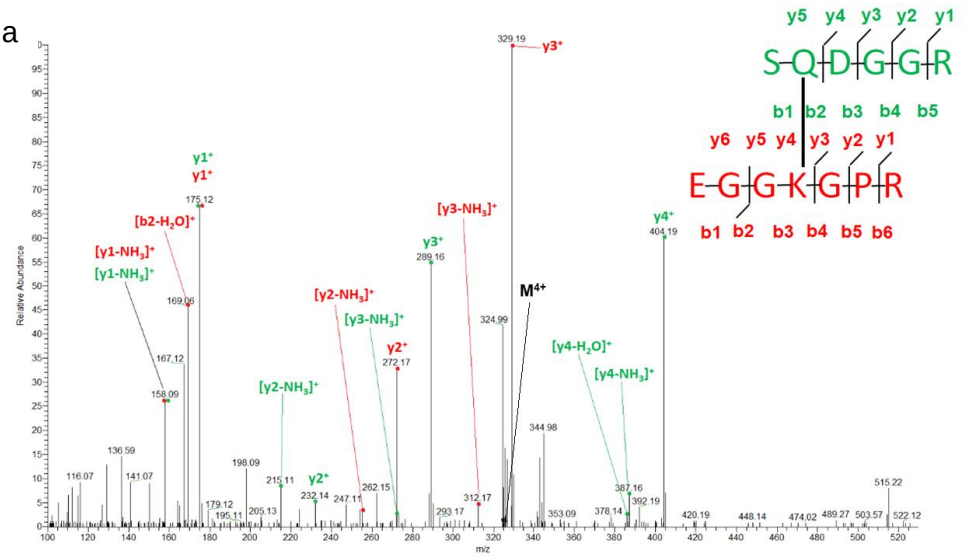



**Figure 3. 11 Potential fragmentation sites of the isopeptide bond cross-linked TXLP.** a. Formation of a reporter ion; b. Isopeptide bond cleavage.

To find the optimal fragmentation energy for each method, CID and HCD voltage ramping were performed with TXLP. For each collision mode, the spectrum with the highest number of fragments was selected for analysis.

The results of the +4-charge state of TXLP are illustrated in Figures 3.12 and 3.13. Surprisingly, we found that the HCD experiment showed much poorer fragmentation than the CID experiment. In addition, only singly-charged linear fragments were identified using HCD, whereas in the CID spectrum multiply-charged linear and cross-linked fragments were present. These observations could be explained by the fact that cross-linked fragments are generally more highly charged and may therefore be less stable. No cross-linked fragments were found even in the spectrum produced by the lowest applied HCD energy. Moreover, no additional internal fragments were identified in any of the spectra in the HCD ramping experiment including the expected reporter ion.

a

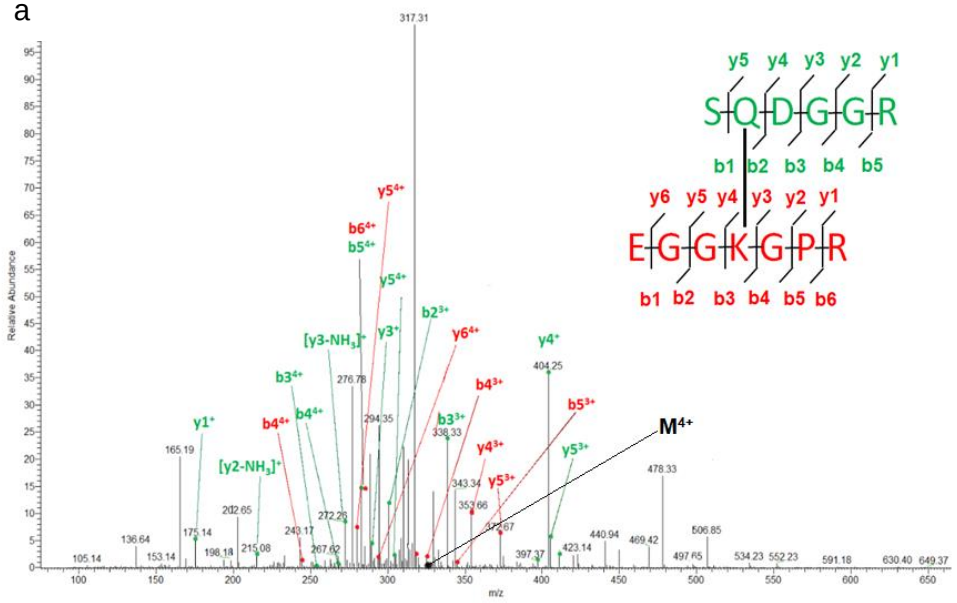



b

| *Fragment* | m/z theoretical | m/z experimental |
|---|---|---|
| $y4^+$ | 404.19 | 404.19 |
| $[y4\text{-}NH_3]^+$ | 387.16 | 387.16 |
| $[y4\text{-}H_2O]^+$ | 386.18 | 386.18 |
| $y3^+$ | 289.16 | 289.16 |
| $[y3\text{-}NH_3]^+$ | 272.14 | 272.14 |
| $y2^+$ | 232.14 | 232.14 |
| $[y2\text{-}NH_3]^+$ | 215.11 | 215.11 |
| $y1^+$ | 175.12 | 175.12 |
| $[y1\text{-}NH_3]^+$ | 158.09 | 158.09 |
| $y3^+$ | 329.19 | 329.19 |
| $[y3\text{-}NH_3]^+$ | 312.17 | 312.17 |
| $y2^+$ | 272.17 | 272.17 |
| $[y2\text{-}NH_3]^+$ | 169.06 | 169.05 |
| $y1^+$ | 175.12 | 175.12 |
| $[y1\text{-}NH_3]^+$ | 158.09 | 158.09 |
| $[b2\text{-}H_2O]^+$ | 169.06 | 169.05 |

**Figure 3. 12 HCD fragmentation spectrum of the +4-charge state of TXLP.** Spectra were collected using a normalised collision energy level of 35. a, HCD fragmentation spectrum (precursor peak is marked with M); b, List of the theoretical and experimental m/z values of all identified fragments (standard mass error of the orbitrap: 20 ppm).

a




b

| *Fragment* | m/z theoretical | m/z experimental |
|---|---|---|
| $y5^{3+}$ | 405.24 | 405.52 |
| $y4^{+}$ | 404.19 | 404.25 |
| $b3^{3+}$ | 338.49 | 338.33 |
| $b2^{3+}$ | 300.14 | 300.34 |
| $y5^{4+}$ | 304.04 | 304.4 |
| $y3^{+}$ | 289.16 | 289.31 |
| $b5^{4+}$ | 282.63 | 282.64 |
| $[y3\text{-}NH_3]^{+}$ | 272.14 | 272.26 |
| $b4^{4+}$ | 268.37 | 267.62 |
| $b3^{4+}$ | 254.12 | 254.26 |
| $[y2\text{-}NH_3]^{+}$ | 215.11 | 215.08 |
| $y1^{+}$ | 175.12 | 175.14 |
| $y5^{3+}$ | 372.52 | 372.67 |
| $y4^{3+}$ | 353.52 | 353.66 |
| $b5^{3+}$ | 344.16 | 344.51 |
| $b4^{3+}$ | 325.15 | 325.1 |
| $y6^{4+}$ | 282.63 | 282.64 |
| $y5^{4+}$ | 279.64 | 279.84 |
| $b6^{4+}$ | 282.63 | 282.64 |
| $b4^{4+}$ | 244.11 | 244.28 |

**Figure 3. 13 CID fragmentation data of the +4-charge state of TXLP.** Spectra were collected using a normalised collision energy level of 35 (CID 35). a, CID fragmentation spectrum (precursor peak labelled with M); b, List of the theoretical and experimental m/z values of all identified fragments (standard mass error of the linear ion trap: 0.8 Da).

To investigate our hypothesis that cross-linked fragments are less stable than linear fragments, we compared the ion intensities of different types of fragments, linear and cross-linked, at two different HCD energies. Both types of fragments were found in the MS2 spectrum of the +2 charge state of TXLP (Figure 3.14). If the hypothesis was correct, we would have expected a simultaneous drop in the intensities of the cross-linked fragments and rise in the intensities of the linear fragments with increasing HCD energy. As shown in Figure 3.15, there was indeed an overall increase in the intensities of the linear fragments, while a much reduced increase or decrease was observed for cross-linked peptides. This result could not be explained by the higher charge state of cross-linked fragments, since only singly- and doubly-charged peptides were used for both types of fragments.

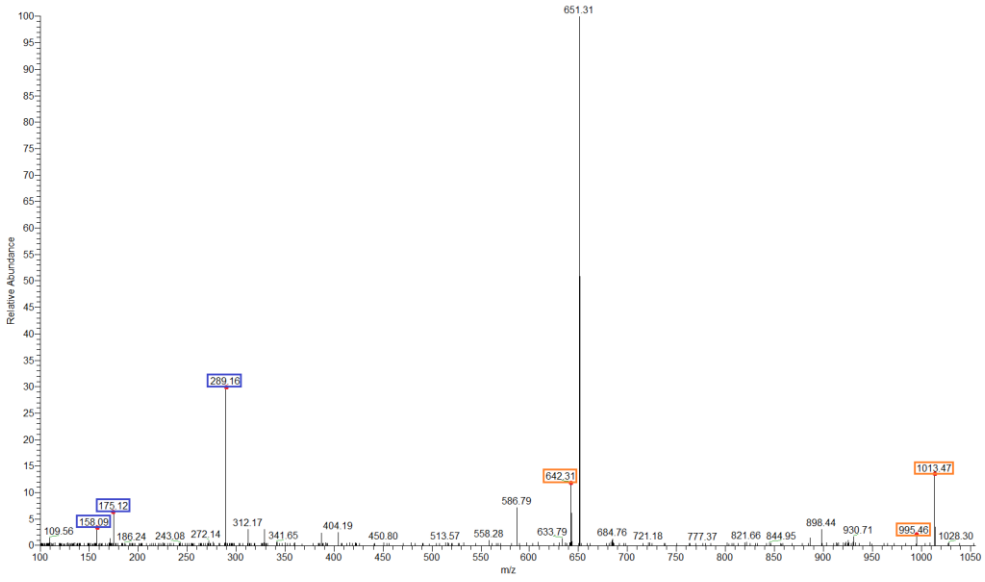



**Figure 3. 14 HCD fragmentation spectrum of +2-charged TXLP.** The peak at m/z = 651.31 belongs to the precursor ion. The fragment ions used to compare fragment intensities are marked with blue (linear) and orange (cross-linked) borders.

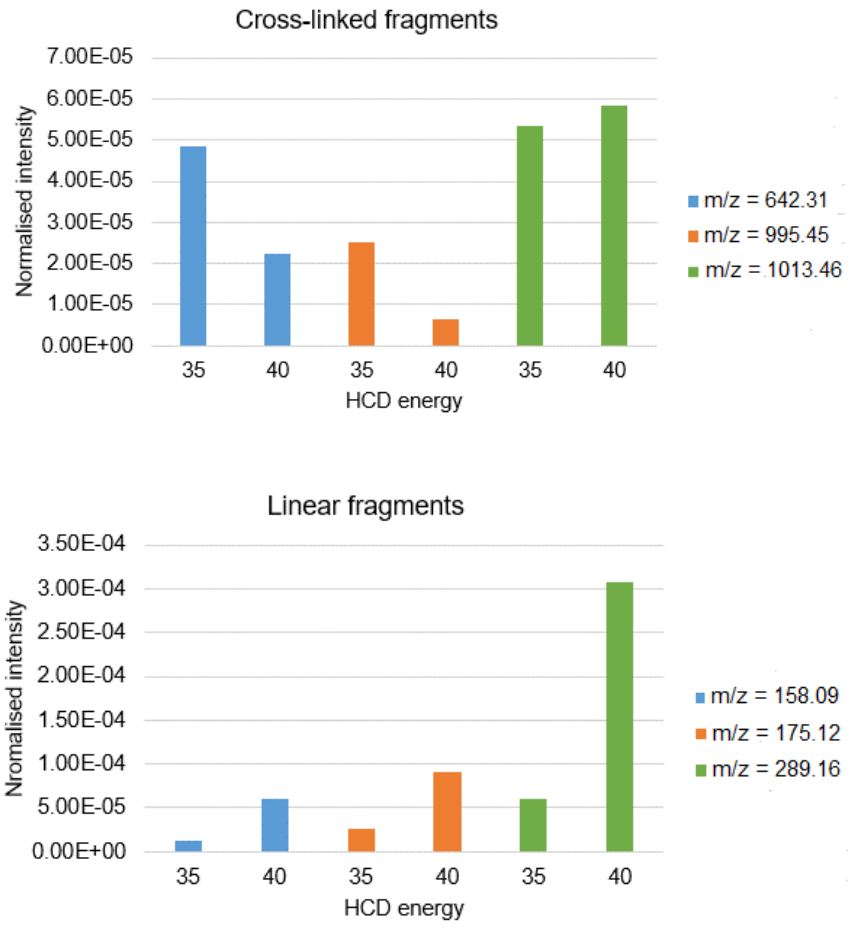



**Figure 3. 15 Normalised intensities of linear and cross-linked fragments of the +2-charged TXLP using two different HCD energies.** Fragment ion intensities were normalised to the corresponding precursor ion intensity. The analysed fragments are shown in Figure 3.14.

Taking these results together, we concluded that the HCD fragmentation method provided no additional information that could be useful in future experiments. Consequently, further method developments were done only using CID with the aim of combining precursor mass and MS2 fragmentation (i.e. amino acid sequence) data to identify cross-linked peptides.

*3.2.3.1.2 Collision energy*

In the next stage, we aimed to find an optimal CID collision energy for fragmentation. In addition to good sequence coverage of both peptides, an

important criterion was to obtain the highest possible fragment intensity. To address this question, we selected several high-intensity, cross-linked fragments from the TXLP CID 35 spectrum and investigated how the intensity changed with increasing CID energy. The results, illustrated in Figure 3.16, showed that the normalised intensities reach a maximum at CID 35 followed by a sudden drop caused by over fragmentation.

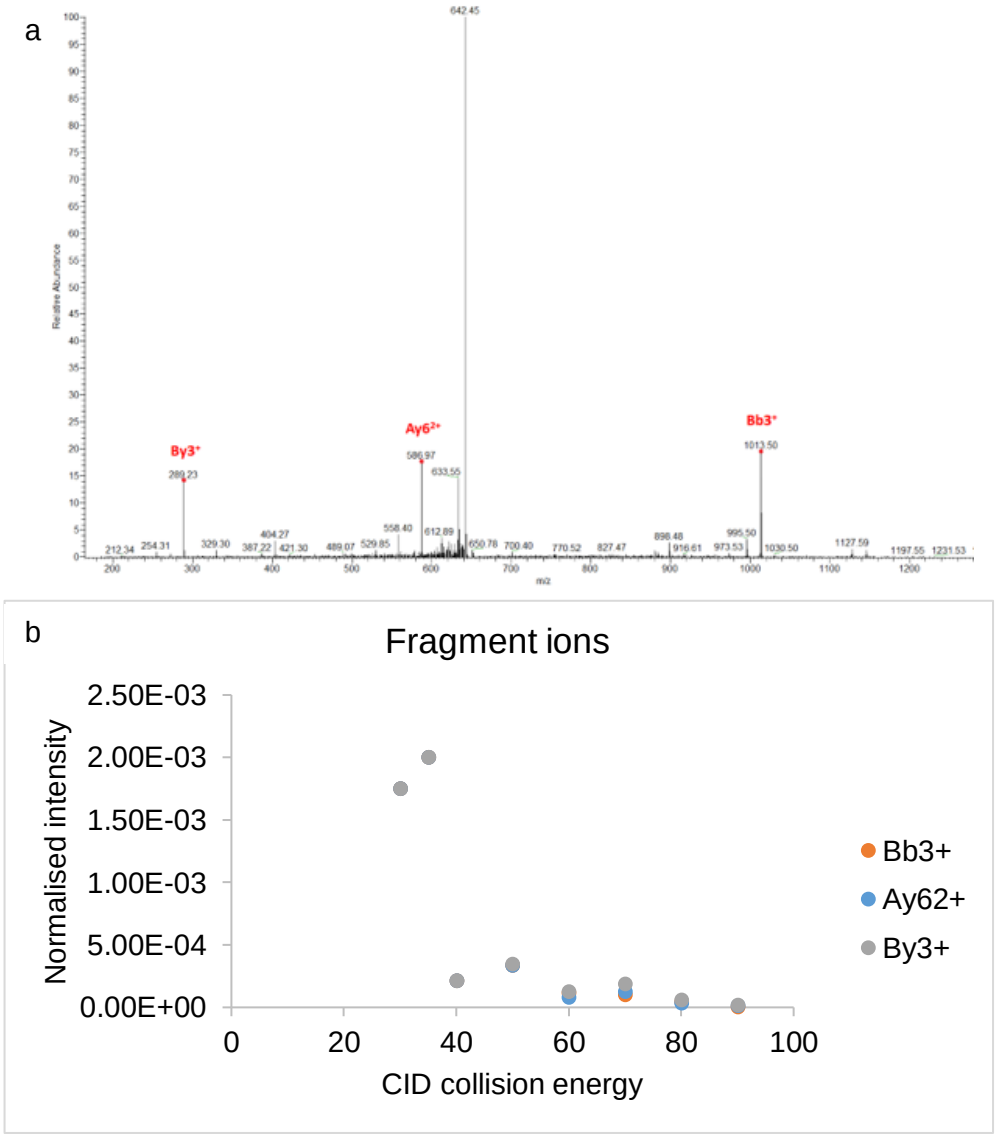



**Figure 3. 16 Analysis of the fragmentation intensity of the three most intense TXLP fragments.** The selected fragment ions are shown in panel a; a graph showing the dependence of normalised peak intensity on CID collision energy is depicted in panel b. The fragment peak intensities were normalised against the precursor intensity from the MS1 spectrum. Data were selected from the fragmentation spectra of +2-charged TXLP. The fragments are labelled with A or B to indicate the fragmented peptide (A peptide: EGGKGPR, B peptide: SQDGGR).

In addition, this effect seemed to be independent of the precursor charge state, as the intensity of each fragment reached a maximal value at the same CID energy for three different charge states (Figure 3.17).

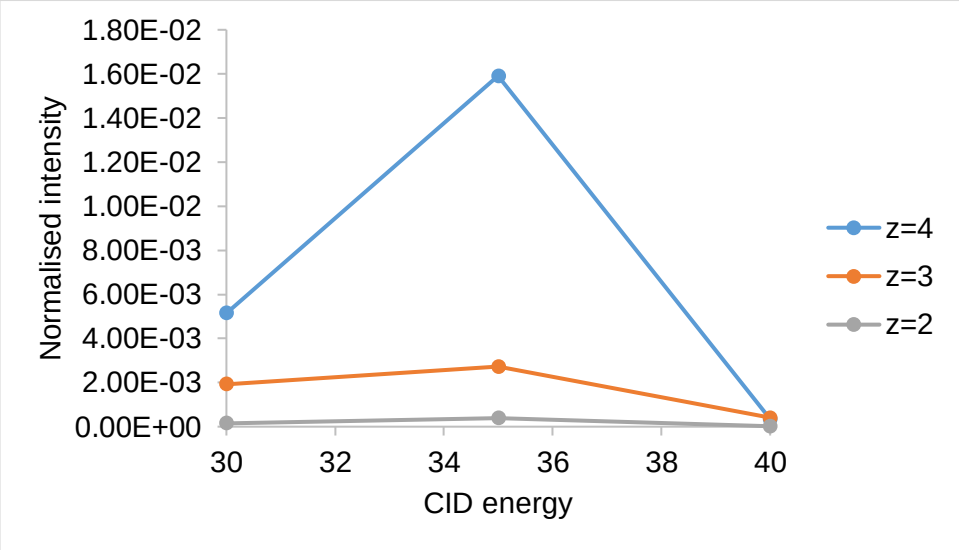



**Figure 3. 17 Intensity change of the fragment $y4^+$ from Figure 3.13 (m/z = 404.25) with increasing CID energy for the different precursor charge states.** A CID energy of 35 produced the optimal fragmentation of TXLP at three different charge states.

Unlike the HCD experiment, the CID experiment also showed linear fragments resulting from the cleavage at the isopeptide bond. Their fragmentation intensity curve (Figure 3.18) shows a very similar pattern to the previously-described, cross-linked fragments (Figure 3.16). This suggested that the peaks produced by isopeptide and peptide bond cleavage behave similarly with respect to optimal collision energy. Subsequent isolation and MS3 fragmentation to characterise the linear fragments was not successful due to the very low intensity of these species. We tried to manage this problem by increasing the maximal injection time of the instrument to allow more MS2 fragment ions to accumulate prior to MS3. However, this measure also showed no positive results. Since there was consequently insufficient evidence to confirm that the TXLP had fragmented at the

isopeptide bond, we did not look any further for isopeptide bond cleavage products in the course of method development.

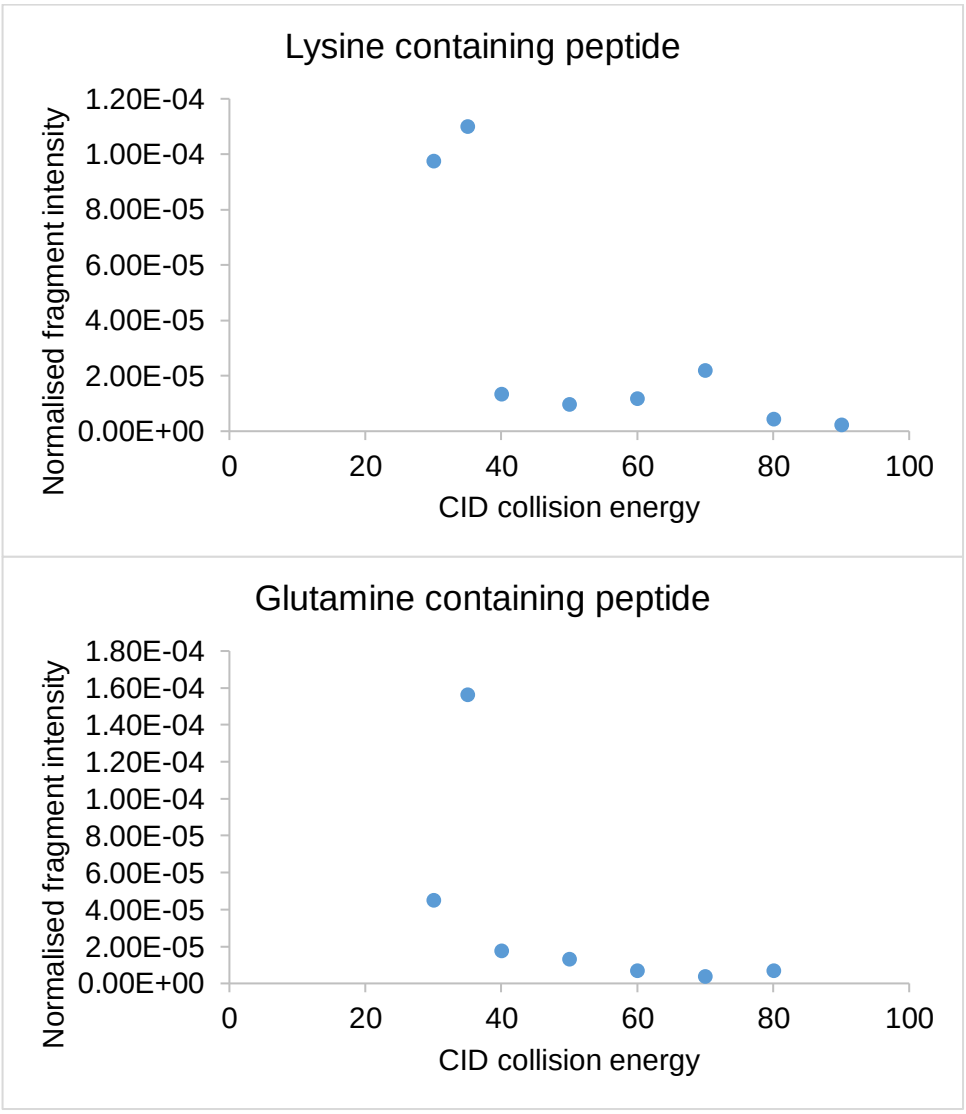



**Figure 3. 18 Dependence of normalised intensities of the isopeptide bond cleavage fragments on CID activation energy.** Precursor charge state: +2.

### *3.2.3.2 Identification of a tryptic cross-linked peptide using MassMatrix*

In order to screen digests of complex protein mixtures prepared from biological samples such as cell culture and tissue, the next step in the method development was to find an automated way to identify transglutaminase cross-linked peptides using software. The TXLP model peptide from the previous section was used to develop this assay. As the

amino acid sequences of this synthetic peptide were designed to correspond to two mouse collagen-alpha1(I) tryptic peptides, we planned to search the mass spectrometry data from the previous section for sequences derived from the whole protein. our primary objective was therefore to input the collagen-alpha1(I) sequence and optimise the software parameters in different programs to allow each program to identify the cross-linked peptides.

Standard proteomic software for the analysis of trypsin digests was unable to identify cross-linked peptides. In addition, much of the available bespoke software for cross-linked peptide identification had been designed only for chemical cross-linking and did not encompass the isopeptide cross-link. Only two suitable programs were identified, MassMatrix [123] and MeroX [124], that allowed the user to define the properties of the cross-link. Searches were therefore performed using these programs.

A secondary objective of this study was to gather preliminary information about what range of score can be achieved by a true-positive identification. This information was expected to be crucial for discriminating true positives from false-positive identifications in complex biological samples.

I started by investigating the highest score that could be achieved from the voltage-ramped, CID-fragmentation spectra of the model peptide. Figure 3.19 shows a typical report of a positive hit produced by a MassMatrix search. It can be seen that both cross-linked peptides were correctly identified within the collagen-alpha1(I) protein sequence.

a Mouse collagen-alpha1(I) protein sequence

```
001 MFSFVDLRLL LLLGATALLT HGQEDIPEVS CIHNGLRVPN GETWKPEVCL ICICHNGTAV CDDVQCNEEL DCPNPQRREG ECCAFCPEEY VSPNSEDVGV  100
101 EGPKGDPGPQ GPRGPVGPPG RDGIPGQPGL PGPPGPPGPP GPPGLGGNFA SQMSYGYDEK SAGVSVPGPM GPSGPRGLPG PPGAPGPQGF QGPPGEPGEP  200
201 GGSGPMGPRG PPGPPGKNGD DGEAGKPGRP GERGPPGPQG ARGLPGTAGL PGMKGHRGFS GLDGAKGDAG PAGPKGEPGS PGENGAPGQM GPRGLPGERG  300
301 RPGPPGTAGA RGNDGAVGAA GPPGPTGPTG PPGFPGAVGA KGEAGPQGAR GSEGPQGVRG EPGPPGPAGA AGPAGNPGAD GQPGAKGANG APGIAGAPGF  400
401 PGARGPSGPQ GPSGPPGPKG NSGEPGAPGN KGDTGAKGEP GATGVQGPPG PAGEEGKRGA RGEPGPSGLP GPPGERGGPG SRGFPGADGV AGPKGPSGER  500
501 GAPGPAGPKG SPGEAGRPGE AGLPGAKGLT GSPGSPGPDG KTGPPGPAGQ DGRPGPAGPP GARGQAGVMG FPGPKGTAGE PGKAGERGLP GPPGAVGPAG  600
601 KDGEAGAQGA PGPAGPAGER GEQGPAGSPG FQGLPGPAGP PGEAGKPGEQ GVPGDLGAPG PSGARGERGF PGERGVQGPP GPAGPRGNNG APGNDGAKGD  700
701 TGAPGAPGSQ GAPGLQGMPG ERGAAGLPGP KGDRGDAGPK GADGSPGKDG ARGLTGPIGP PGPAGAPGDK GEAGPSGPPG PTGARGAPGD RGEAGPPGPA  800
801 GFAGPPGADG QPGAKGEPGD TGVKGDAGPP GPAGPAGPPG PIGNVGAPGP KGPRGAAGPP GATGFPGAAG RVGPPGPSGN AGPPGPPGPV GKEGGKGPRG  900
901 ETGPAGRPGE VGPPGPPGPA GEKGSPGADG PAGSPGTPGP QGIAGQRGVV GLPGQRGERG FPGLPGPSGE PGKQGPSGSS GERGPPGPMG PPGLAGPPGE 1000
1001 SGREGSPGAE GSPGRDGAPG AKGDRGETGP AGPPGAPGAP GAPGPVGPAG KNGDRGETGP AGPAGPIGPA GARGPAGPQG PRGDKGETGE QGDRGIKGHR 1100
1101 GFSGLQGPPG SPGSPGEQGP SGASGPAGPR GPPGSAGSPG KDGLNGLPGP IGPPGPRGRT GDSGPAGPPG PPGPPGPPGP PSGGYDFSFL PQPPQEKSQD 1200
1201 GGRYYRADDA NVVRDRDLEV DTTLKSLSQQ IENIRSPEGS RKNPARTCRD LKMCHSDWKS GEYWIDPNQG CNLDAIKVYC NMETGQTCVF PTQPSVPQKN 1300
1301 WYISPNPKEK KHVWFGESMT DGFPFEYGSE GSDPADVAIQ LTFLRLMSTE ASQNITYHCK NSVAYMDQQT GNLKKALLLQ GSNEIELRGE GNSRFTYSTL 1400
1401 VDGCTSHTGT WGKTVIEYKT TKTSRLPIID VAPLDIGAPD QEFGLDIGPA CFV 1453
```

b

| Charge | score | pp | $pp_2$ | $pp_{tag}$ | m/z | MW(obs) | MW | delta | sequence |
|---|---|---|---|---|---|---|---|---|---|
| +2 | 19 | 5.9 | 4.9 | 1.8 | 651.3138 | 1301.6204 | 1301.6193 | 0.0011 | EGGK($1)GPR SQ($1)DGGR |
| +2 | 17 | 5.9 | 5.0 | 2.4 | 651.3145 | 1301.6218 | 1301.6193 | 0.0025 | EGGK($1)GPR SQ($1)DGGR |
| +2 | 19 | 6.8 | 5.1 | 2.8 | 651.3145 | 1301.6218 | 1301.6193 | 0.0025 | EGGK($1)GPR SQ($1)DGGR |

c

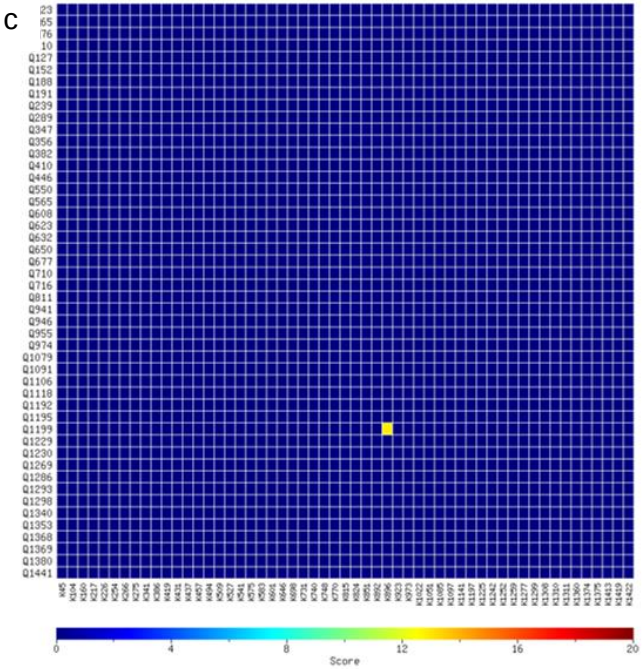



**Figure 3. 19 MassMatrix search of the TXLP LC-MS data obtained using CID 30.** The positions of both cross-linked peptides are highlighted in the sequence using red text in panel a. The scans in the original MS data that contained the single, high-scoring pair are shown in panel b. The colour coding of each amino acid indicates in which series of the MS2 spectrum it was found. The green colour indicates that the amino acid was found in the y-ion series, the blue in the b-ion series and the red in both series. The amino acids that were not found in any of the series are shown in bold type. The exact position of the cross-link is marked with the symbol ($1) after the cross-linked amino acid. Further information about the precursor m/z value, the theoretical (MW) and the observed (MW(obs)) molecular weights of the precursor as well as the corresponding mass difference are reported in Da in the search result. A 2D heatmap of the overall score for each glutamine (Q) and lysine (K) residue pair is depicted in panel c, in which only the single positively-identified cross-link position (Q1199-K896) possesses an overall score greater than zero.

The scoring system in MassMatrix consists of three different parameters - three statistical values (pp, pp2 and $pp_{tag}$) and one descriptive value (score) for each spectrum [134]. An overall score (displayed in Figure 3.19b) is derived from the ratio of the abundance of all matched product ions to the total spectrum abundance. This value is not independent on peptide length. The pp value is the most reliable parameter for each spectrum, as it considers the number of predicted product ions that match the peaks in that spectrum. This value implicitly depends on the mass accuracy. The pp2 value is calculated from the total abundance of the matched fragments, while the $pp_{tag}$ value evaluates the match based on colour tags of the peptide (see Figure 3.19b). Hence, the latter two quantifications provide complementary, additional information to the pp value. Another overall score displayed in the heat map (Figure 3.19c) is derived from the values of all statistical parameters of a peptide hit (pp, pp2 and $pp_{tag}$) and the number of spectral matches.

To confirm the previous assessment of optimal CID energy, we aimed to investigate how the MassMatrix scoring would change with increasing CID energy. We assumed that the change of the CID energy was the key variable affecting the MassMatrix score, as it was the main parameter that determined the number and abundance of formed fragment ions. To assess the optimal CID energy required to maximise the MassMatrix score, we performed a search using LC-MS data obtained at ramped CID voltages and monitored the change of the different scores for TXLP identification. Table 3.2 shows the dependence of MassMatrix scores on CID energy. Although there was only a modest change in pp value between CID 35 and 50, and although the number of identified fragments increased across this range from only from 49 to 52, CID 40 gave the highest pp value and also the highest number of

identifying scans. This suggested that CID 40 was the optimal fragmentation energy for this peptide.

| *CID energy* | max score | max pp value | max pp2 value | max $pp_{tag}$ value | number of MS/MS scans |
|---|---|---|---|---|---|
| *30* | 19 | 6.8 | 5.1 | 2.8 | 3 |
| *35* | 30 | 9.9 | 6 | 3.9 | 5 |
| *40* | 25 | 11 | 6.4 | 2.4 | 6 |
| *50* | 49 | 10.9 | 9 | 3 | 1 |
| *60* | 24 | 8.3 | 6.4 | 2.2 | 1 |
| *70* | 28 | 7.5 | 9.7 | 2.5 | 1 |

**Table 3. 2 Dependence of MassMatrix score on CID fragmentation energy.** Although the table contains data from all of the detected charged forms of TXLP, only the +2-charged peptide was reproducibly identified by the software. The +3 charge state appeared only occasionally in the search results and the +4 charge state was not reported at all.

The spectra from which the identifications were made revealed a poor amount of fragmentation relative to the precursor peak at CID 30 and 40, but CID 50 appeared to produce multiple fragment peaks with comparable abundance to the precursor peak (Figure 3.20). However, increasing the CID energy to 50 also appeared to result in a relatively high level of over fragmentation, as the number of productive MS2 scans decreased from six at CID 40 to only one (Table 3.2). The labelled precursor peak had a m/z shift of approximately -9 units compared to the theoretical m/z value of the doubly charged precursor and was identified by MassMatrix as the precursor with a loss of ammonia, for which the expected m/z shift is -8.5 units at z=2. Interestingly, a main peak with the same m/z shift was also found in the fragmentation spectra of the more highly charged TXLP precursors. This can be seen in the fragmentation spectrum of the +4-charged peptide (Figure 3.13) where a main peak at 317.31Da was detected (precursor m/z =

326.16Da). This characteristic feature could be used for future identification of transglutaminase cross-linked peptides. However, more research is needed to explain the formation of this fragment by revealing its structure.

The MassMatrix search also delivered several false-positive matches to other tryptic peptides predicted from the collagen alpha-1(I) sequence. These peptides were identified based on only one MS/MS scan (data not shown), while multiple scans were assigned to TXLP (Figure 19b). This observation led me to conclude that the number of scans was crucial for discrimination between true- and false-positive matches. For this reason, we rejected using CID 50 for the analysis.

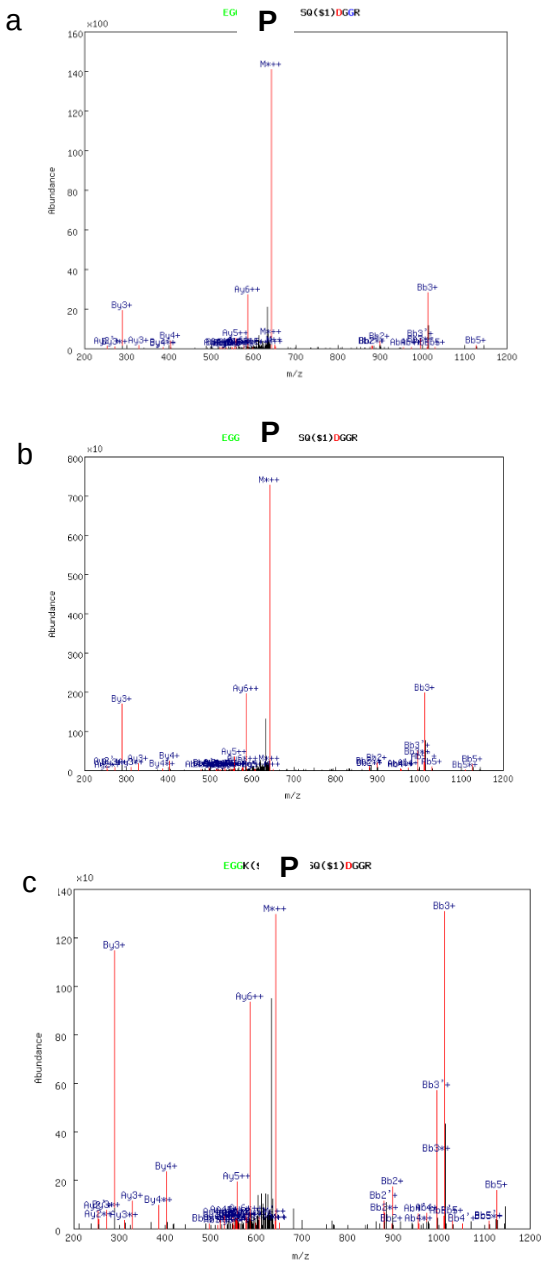



**Figure 3. 20 Fragmentation spectra produced at different CID energies of the +2-charge state of the precursor ion identified by MassMatrix.** Spectra were recorded at CID energies of 30 (panel a), 40 (panel b) and 50 (panel c). (**P**, precursor ion peak).

I had expected cross-linked peptide ions with higher charges than +2 would show good levels of fragmentation at lower energies, as previously demonstrated at z=4 and CID 35 (Figure 3.13). Although the +4-charge state of the TXLP was not identified by MassMatrix, the good level of sequence coverage arising from the fragment ions identified at CID 35 using a precursor charge state of +2 suggested that this fragmentation mode would

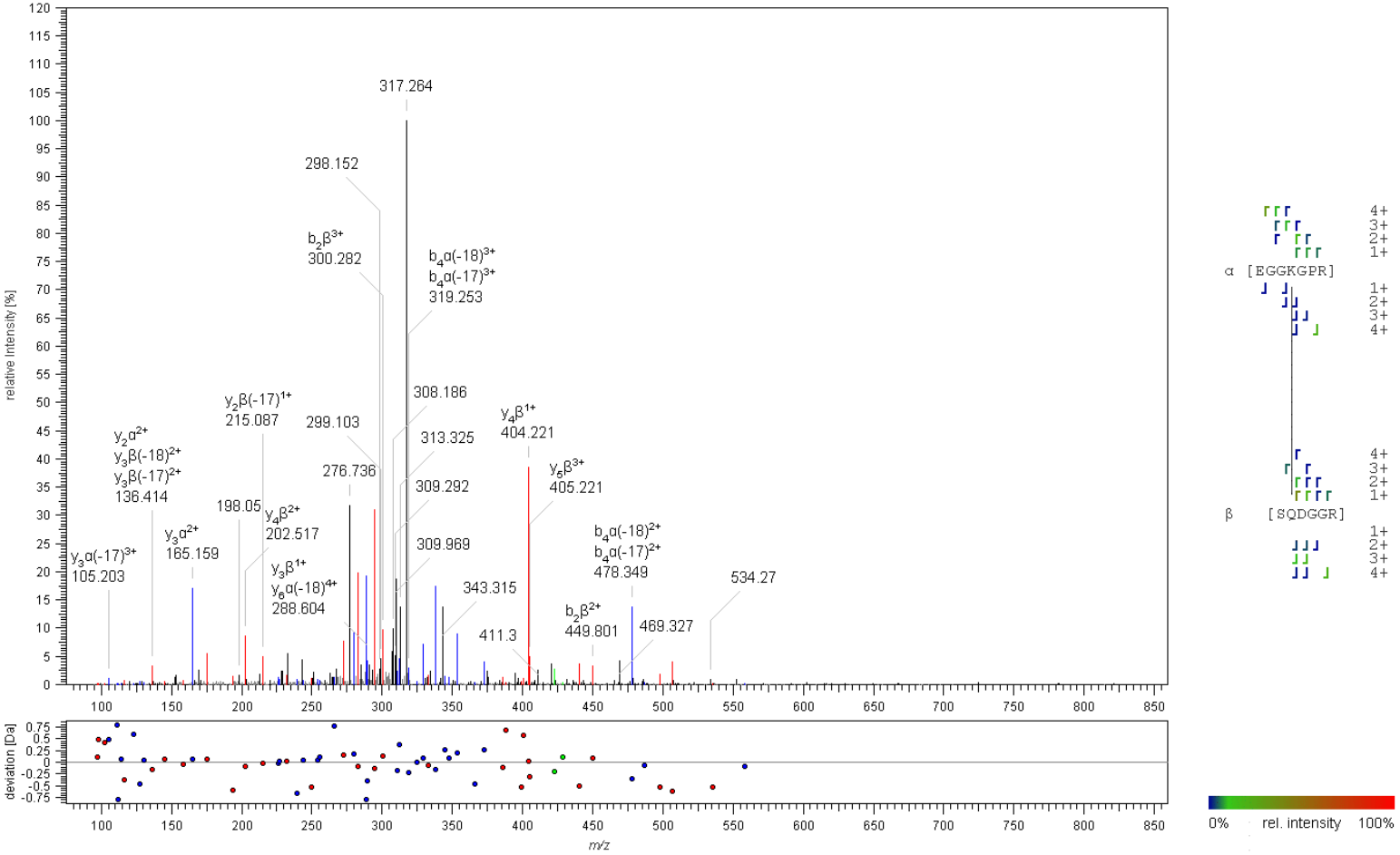
be sufficient for the identification of other cross-linked peptides at this charge state. Therefore, we concluded that CID 35 was the optimal energy for charge states from +2 to +4. However, in tryptic protein digests, missed cleavages might lead to the formation of cross-linked peptides with even higher charge states than +4, that may be less stable than the analysed TXLP charge states. Thus, in addition to the optimal energy of CID 35, we also decided to search for higher-charge-state, cross-linked peptides in LC-MS data generated using CID 30.

#### *3.2.3.3 Identification of a tryptic cross-linked peptide using MeroX*

MeroX is another software program previously used to identify transglutaminase cross-linked peptides from tryptic protein digests [78]. The advantage of this software relies on several scoring settings that enable the selection of cross-linked peptide IDs with high confidence. For instance, it allows the user to define a signal to noise ratio threshold for the MS2 spectra and to select the number of MS2 spectra used for the calculation of the score (known as score precision). In addition, it is suitable for data analysis generated with high-precision MS instruments, while MassMatrix does not have this option. Based on these advantages, we reasoned that MeroX would be a reliable tool for identifying transglutaminase cross-linked peptides and we investigated the optimal search settings using the data previously generated using TXLP. Searches of the voltage-ramped LC-MS analysis of TXLP showed a significant hit only using CID 40, 50 and 60. Lower energy MS2 spectra (30 and 35) did not allow identification of the cross-linked peptide even after applying less stringent settings, including eliminating the pre-scoring option, or lowering the signal to noise ratio threshold or the score threshold. The CID 40 fragmentation data allowed the identification of the +4-

charged peptide, whereas the higher fragmentation energy data allowed only the +2-charged peptide to be found (Figure 3.21). The scores were 115 for the +4-charged cross-linked peptide and 72 for the +2-charged cross-linked peptide (CID 50). Increasing the CID energy from 50 to 60 led to a modest increase of the latter score to 76. However, the energy increase seemed to have led to over fragmentation, as some highly abundant fragments present in the CID 50 spectrum were absent from the CID 60 MS2 spectrum (data not shown). The observation that the model peptide starts to be identified in the MeroX search at higher CID energies than the MassMatrix search may be explained by the higher requirements of the former software in terms of the MS2 spectrum quality. MeroX seemed to require more efficient fragmentation in order to report a positive match. However, the results for the fragmentation of the +2-charged peptide pair were also affected by the high abundance of the fragment identified as the precursor ion after water or ammonia loss (Figure 2.21b: precursor(-18)$^{2+}$, precursor(-17)$^{2+}$) that lowered the relative intensity of the rest of the fragments. This may also explain why no identifications were made using the data generated at CID 30 and 35.

a


relative Intensity [%]
317.264
298.152
b2β3+
300.282
b4α(-18)3+
b4α(-17)3+
319.253
y2β(-17)1+
215.087
308.186
y4β1+
404.221
299.103
313.325
y2α2+
y3β(-18)2+
y3β(-17)2+
136.414
276.736
y5β3+
405.221
309.292
198.05
y4β2+
202.517
y3α2+
165.159
309.969
b4α(-18)2+
b4α(-17)2+
478.349
y3α(-17)3+
105.203
y3β1+
y6α(-18)4+
288.604
343.315
534.27
b2β2+
449.801
411.3
469.327
deviation [Da]
m/z
α [EGGKGPR]
β [SQDGGR]
0%
rel. intensity
100%

b

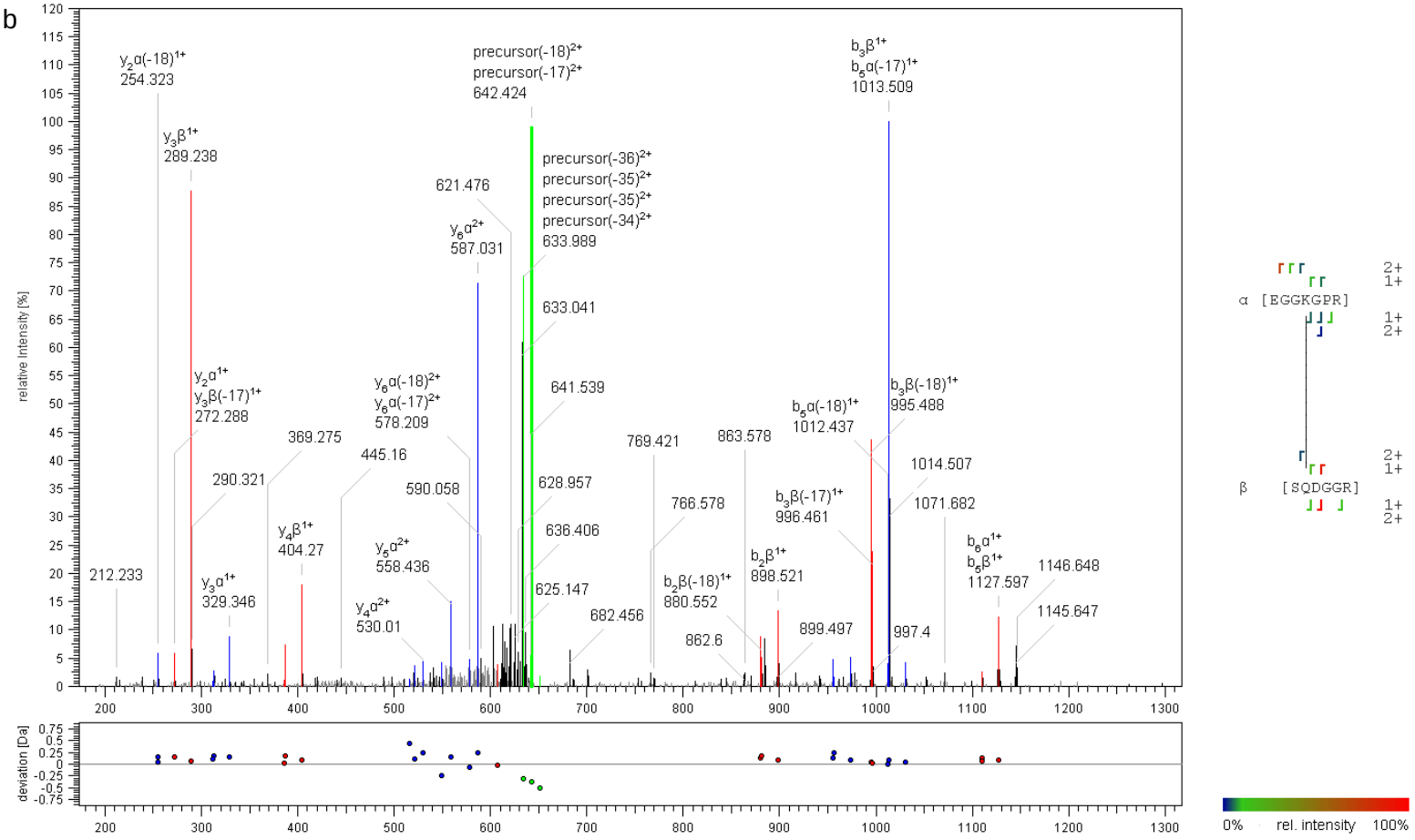

relative intensity [%]
deviation [Da]
y2α(-18)1+ 254.323
y3β1+ 289.238
precursor(-18)2+ precursor(-17)2+ 642.424
b3β1+ b5α(-17)1+ 1013.509
precursor(-36)2+ precursor(-35)2+ precursor(-35)2+ precursor(-34)2+ 633.989
621.476
y6α2+ 587.031
633.041
y2α1+ y3β(-17)1+ 272.288
y6α(-18)2+ y6α(-17)2+ 578.209
641.539
b3β(-18)1+ 995.488
b5α(-18)1+ 1012.437
769.421
863.578
369.275
445.16
290.321
590.058
628.957
1014.507
766.578
b3β(-17)1+ 996.461
1071.682
y4β1+ 404.27
636.406
b2β1+ 898.521
b6α1+ b5β1+ 1127.597
y5α2+ 558.436
625.147
b2β(-18)1+ 880.552
1146.648
212.233
y3α1+ 329.346
682.456
y4α2+ 530.01
862.6
899.497
997.4
1145.647
α [EGGKGPR]
β [SQDGGR]
2+ 1+ 1+ 2+
0% rel. intensity 100%

**Figure 3. 21 MeroX analysis of TXLP LC-MS/MS data.** The identified +4-charged peptide is shown in panel a (CID 40) at m/z = 317.264, and the +2 charged peptide (CID 50) is shown in panel b at m/z = 642.424. The schemes on the right the spectrum show the fragmentation positions of the identified fragments and the fragment intensities. The mass deviation of each fragment is depicted in the graphs below the spectra.

### 3.2.4 Identification of transglutaminase cross-linked peptides from protein digests

Up to this point our results had left unresolved the question of how to set a scoring threshold for a range of cross-linked peptides with higher charges. While this was relatively straightforward using a single model peptide, sequence variation in cross-linked peptides and the presence of numerous linear peptides had to be addressed to achieve our aim of identifying cross-linked peptides in complex protein digests.

I decided to increase sample complexity incrementally in three stages, which assessed purified proteins, cell culture ECM, and finally tissue ECM. In the first stage cross-linking experiments were performed with selected, purified proteins using different transglutaminases. This was followed by digestion and mass spectrometry to identify cross-linked sites. For this part of the project, only proteins known to be good targets for transglutaminases, specifically fibronectin and fibrinogen, were investigated [135, 136].

#### *3.2.4.1 Cross-linking of purified proteins by transglutaminases*

##### *3.2.4.1.1 Fibronectin cross-linking*

The molecular weight of human fibronectin is approximately 263 kDa, therefore a bimolecular cross-linking reaction catalysed by TG2 would give rise to a dimer with a molecular weight of 526 kDa. A typical initial result is shown in Figure 3.22, where no higher molecular weight bands above the monomer were detected. Instead, staining appears after the cross-linking reaction only on the top of the gel. This indicated that all of the monomer had

reacted by the first time point. We deduced that the cross-linking reaction had been too fast and that any intermediate dimer had completely polymerised under these conditions. Thus, the experiment was repeated with reduced temperature, incubation time and enzyme amount.

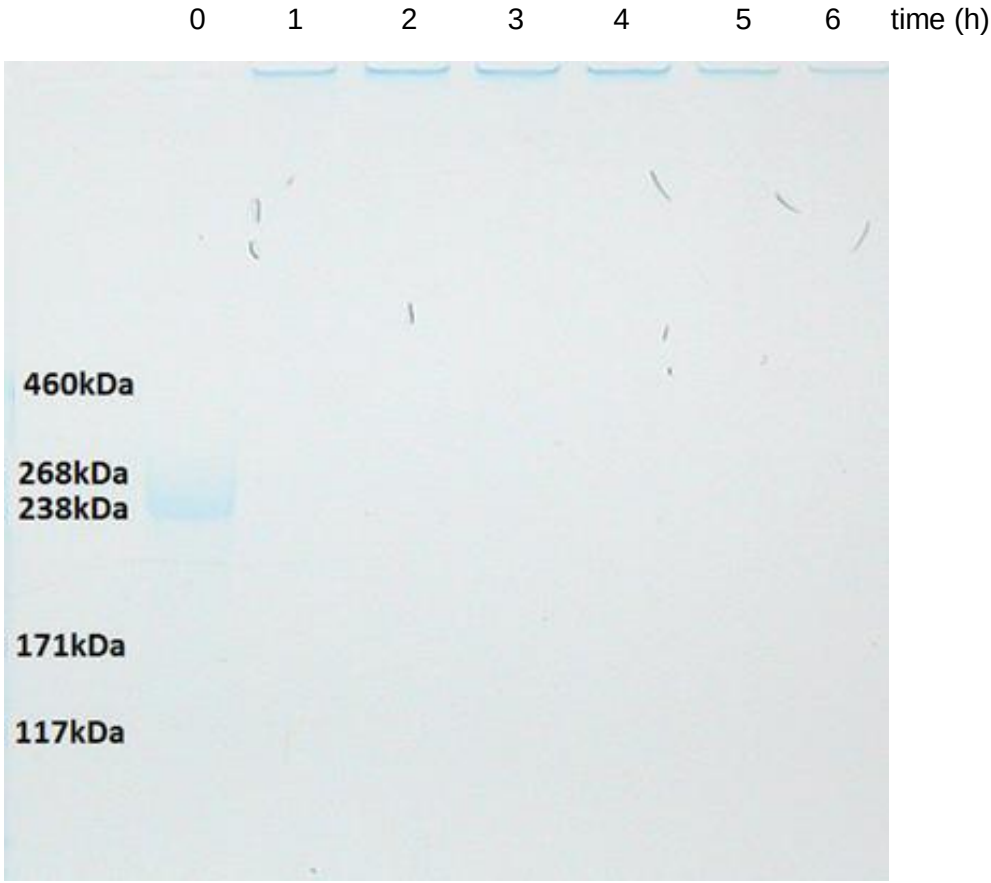



**Figure 3. 22 SDS-PAGE analysis of the time course of fibronectin cross-linking by TG2.** A single stained protein band that corresponds to monomeric fibronectin was detected before incubation with TG2. The protein was cross-linked after the incubation into a polymer that was not able to enter the gel (blue staining of the top of the gel that is absent at the zero time point).

However, reducing the TG2 concentration from 20mU to 5mU per sample, and reducing the temperature from 37°C to 25°C again revealed no dimer band, although the experiments with 20mU TG2 showed a time-dependent decrease of the monomer (data not shown). The absence of a dimer band in

these experiments can be explained either by a very low amount of dimer formation that was below the detection limit of the Coomassie staining or by the size of the dimer or higher order polymer being too large to enter the pores of the gel.

To address the possibility that dimer was being formed at levels that were too low for detection, we tried to detect the dimer by increasing the protein loading on the gel, however once again this did not lead to any dimer detection. Instead, only the blue stain at the top of the gel could be seen (data not shown).

Finally, to eliminate the possibility that the TG2 activity in the previous experiments was still too high to allow any dimer to be detected, we repeated the reaction using even less TG2 activity. As demonstrated in Figure 3.23, staining of the top of the gel was reduced, but the dimer was again not found. Based on these results, we concluded that either TG2 catalyses the formation of higher order polymers of fibronectin far more quickly than the rate of dimer formation, or the dimer is too large to be separated with a 4-15% SDS-PAGE gel. Since a gel with larger pore size could not easily be used, we changed the TG2 substrate to fibrinogen, whose subunits have lower molecular weights.

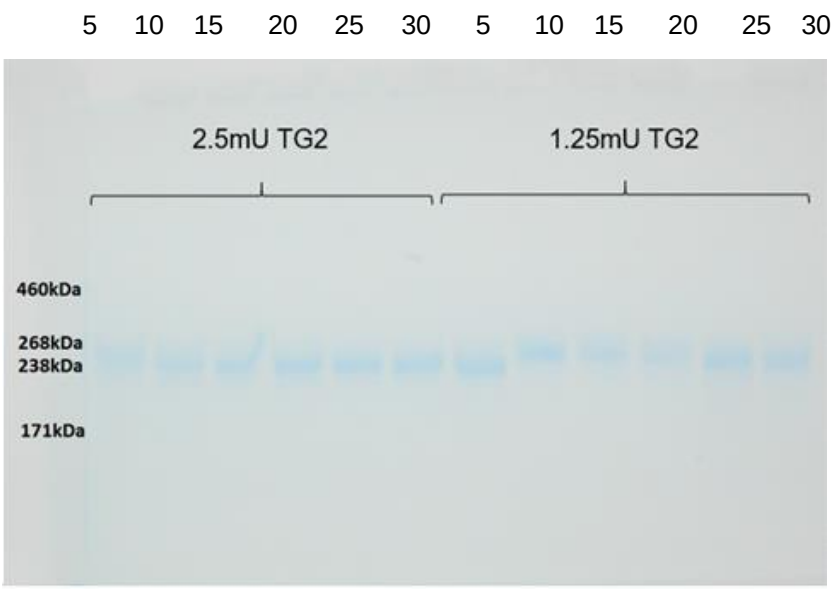



**Figure 3. 23 SDS-PAGE analysis of fibronectin cross-linking with a further decrease in the amount of TG2.**

*3.2.4.1.2 Fibrinogen cross-linking*

Fibrinogen is a heterohexamer (αβγ dimer), whose subunits have molecular weights of approximately 64 kDa (α), 56 kDa (β) and 47 kDa (γ) [137]. All the chains within a hexamer are held together by disulphide bridges. Fibrinogen cross-linking by TG2 was performed using similar reaction conditions to those described by Murthy and Lorand [137]. After treatment with TG2, any cross-linked chains were separated from the non-cross-linked chains by reduction and SDS-PAGE to monitor the progress of the reaction. SDS-PAGE analysis of the TG2 preparation revealed that the enzyme was contaminated by multiple proteins with similar molecular weights to the excised protein bands. These bands were not visible at the 1/25 dilution, suggesting that using this enzyme dilution for fibrinogen cross-linking would allow only fibrinogen products to be seen.

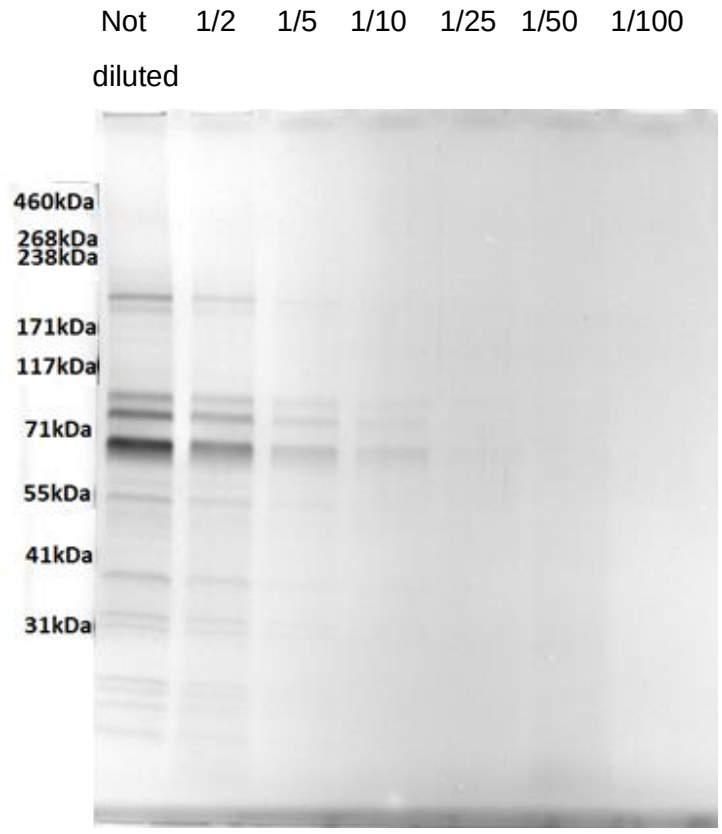



**Figure 3. 24 SDS-PAGE analysis to determine the dilution of TG2 enzyme solution to be used for fibrinogen cross-linking.** The band below the 71kDa marker corresponds to the enzyme.

All subsequent experiments therefore used TG2 at 1:25 dilution. A time course using this reduced TG2 concentration showed the formation of two higher molecular weight bands with a simultaneous decrease of all three monomer band intensities (Figure 3.25). Based on their observed molecular weights, the new species were most likely to be alpha-gamma, beta-gamma or gamma-gamma dimers. Extending the incubation time did not lead to any visible increase in the intensity of the smaller dimer band or formation of any new bands, however the intensity of the larger dimer band decreased with time (Figure 3.25b).

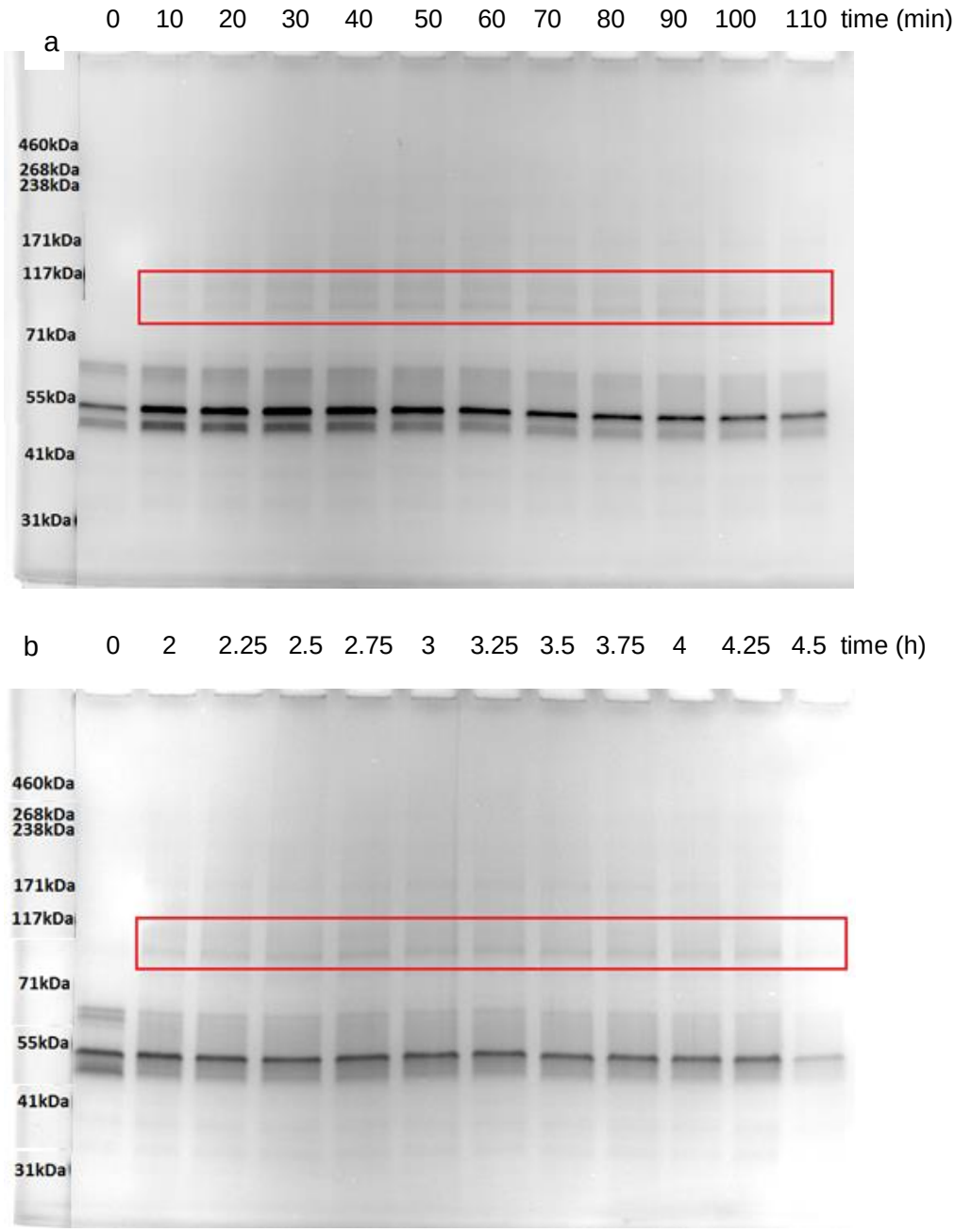
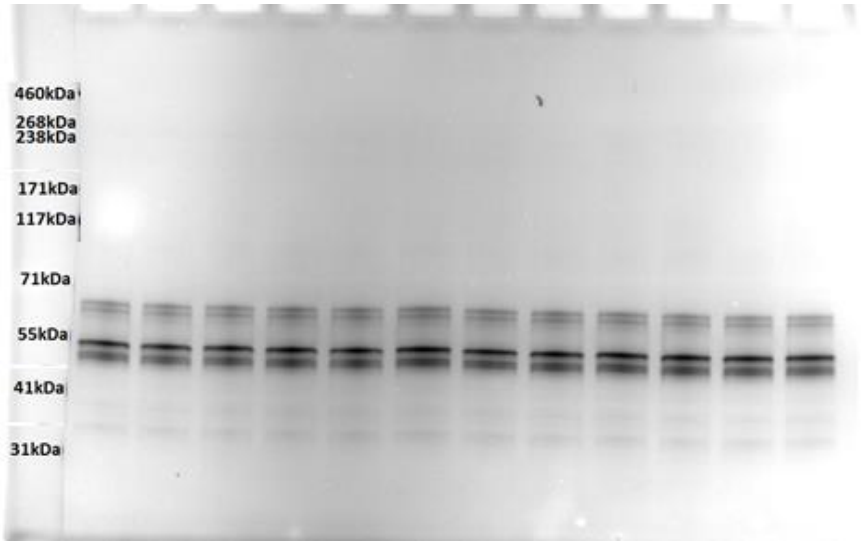

a
0 10 20 30 40 50 60 70 80 90 100 110 time (min)
460kDa
268kDa
238kDa
171kDa
117kDa
71kDa
55kDa
41kDa
31kDa
b
0 2 2.25 2.5 2.75 3 3.25 3.5 3.75 4 4.25 4.5 time (h)
460kDa
268kDa
238kDa
171kDa
117kDa
71kDa
55kDa
41kDa
31kDa
c
Peak area
6000
5000
4000
3000
2000
1000
0
0
20
40
60
time (min)
TG2_2
TG2_1

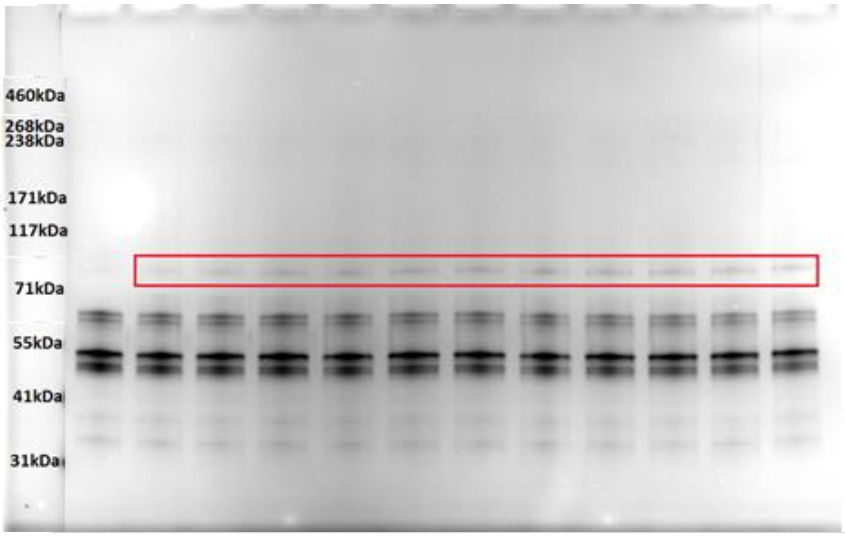
**Figure 3. 25 SDS-PAGE analysis of fibrinogen cross-linking with a lower amount of TG2.** Two major high molecular weight bands appeared between the 71 kDa and 117kDa marker bands (outlined in red in panels a and b). The sizes of these bands corresponded with dimers formed with the fibrinogen gamma chain: alpha-gamma (112 kDa); beta-gamma (103 kDa) or gamma-gamma (94 kDa). Panel c shows the increase of the band intensities of the products over time seen in panel a. TG2_1: larger dimer band; TG2_2 smaller dimer band.

A set of fibrinogen cross-linking experiments was also performed using Factor XIIIa, which is a transglutaminase that had shown a high cross-linking activity with fibrinogen [136]. Previous studies on the cross-linking activity of this enzyme showed an increased preference of Factor XIIIa for fibrinogen gamma chains [137]. Thus, we expected this reaction to result in the formation of a gamma-gamma dimer. In addition to comparing the substrate specificity of different transglutaminases, using Factor XIIIa was of particular value to our project because previous studies had provided evidence for the identity of the fibrinogen residues cross-linked by this enzyme [75]. This would allow a direct comparison with our methods and results. Fibrinogen was cross-linked using Factor XIIIa at two different enzyme concentrations. Figure 3.26 shows a gradual increase in the abundance of a new band only in the reaction with the higher enzyme amount. The apparent size of this band was consistent with the expected gamma-gamma dimer (Figure 3.26b).

a 0 10 20 30 40 50 60 70 80 90 100 110 time (min)




b 0 10 20 30 40 50 60 70 80 90 100 110 time (min)




c

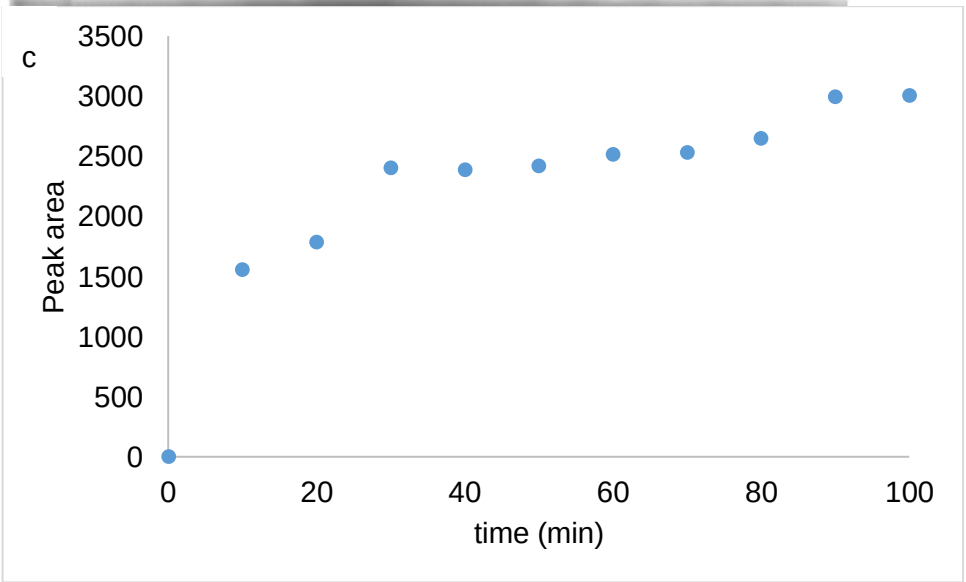
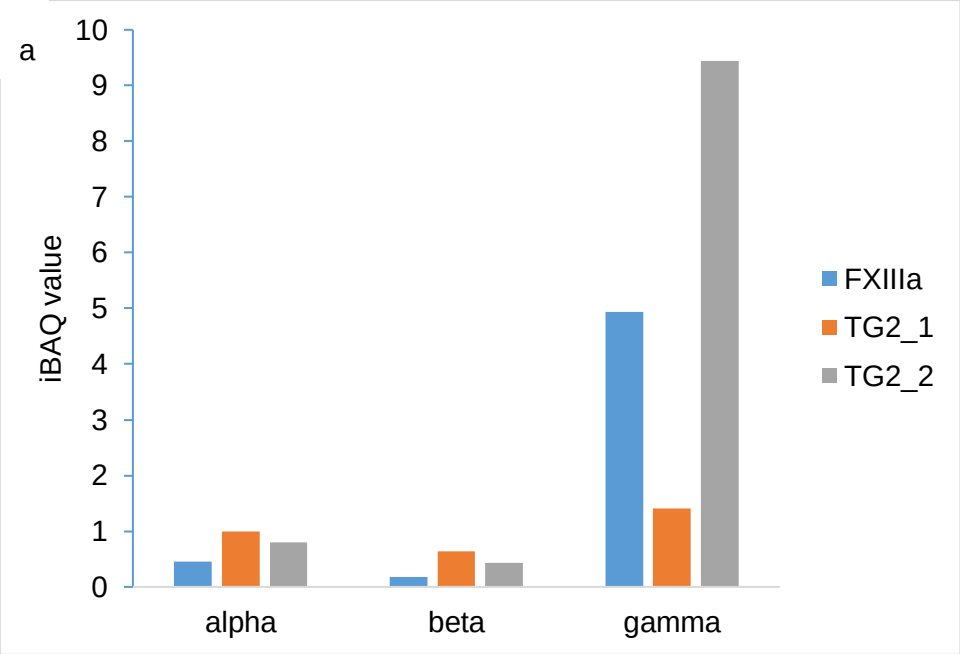


**Figure 3. 26 SDS-PAGE analysis of fibrinogen cross-linking by Factor XIIIa.** Incubation with 0.5 µg Factor XIIIa did not lead to any visible product formation (a), but a specific reaction product was seen after increasing the enzyme amount to 3.8 µg (b). The molecular weight of this product is in the same range as the expected gamma-gamma dimer (94 kDa). Panel c shows the increase of the product band intensity over time.

The bands resulting from the TG2 and Factor XIIIa cross-linking experiments were excised followed by in-gel digestion, ESI-MS/MS and a MaxQuant search to identify the cross-linked proteins. Figure 3.27 shows the results of these searches, which identified all three fibrinogen subunits as the most abundant proteins in each of the three bands. It can be concluded that the Factor XIIIa cross-linking product and the TG2 cross-linking product with the lower molecular weight (TG2_2) are most probably gamma-gamma cross-linked subunits as the iBAQ values of the gamma chain are more than 10-fold higher than the iBAQ values of the other two subunits. This inference is supported by the higher number of gamma-subunit peptides identified compared to the number identified for the larger alpha and beta subunits. In contrast, the composition of the TG2_1 sample remained unclear since the iBAQ and peptide values showed no clear difference between subunits, and the size of the TG2_1 band was inconsistent with a heterotrimeric composition. One possible explanation was that this band contained a mixture of alpha-gamma and beta-gamma dimers, but there was insufficient time to pursue this further.

a




b

| *Fibrinogen subunit* | Ratio to gamma (FXIIIa) | Ratio to gamma (TG2_1) | Ratio to gamma (TG2_2) | FXIIIa band (peptides) | TG2_1 (peptides) | TG2_2 (peptides) |
|---|---|---|---|---|---|---|
| *alpha* | 0.09 | 0.71 | 0.09 | 5 | 6 | 6 |
| *beta* | 0.04 | 0.46 | 0.05 | 5 | 8 | 6 |
| *gamma* | 1 | 1 | 1 | 6 | 4 | 8 |

**Figure 3. 27 MaxQuant search results for the Factor XIIIa and TG2 cross-linked products.** The product bands from the 10, 20 and 30-minute incubation times were pooled and prepared for analysis (Figures 25a and 26b). The bar chart in panel a depicts the iBAQ values for the different fibrinogen subunits found in each sample. The calculated ratios of the iBAQ values to the iBAQ value of the gamma chain and the peptide numbers identified for each subunit are shown in panel b.

### *3.2.4.2 Identification of transglutaminase cross-linking sites from purified protein digests*

Having established clear evidence that Factor XIIIa and TG2 both catalysed the formation of cross-links between fibrinogen gamma subunits, our next objective was to identify the cross-linking positions in each dimer by applying the MassMatrix and MeroX searches that we had tested and developed using a synthetic peptide.

MassMatrix searches were performed for each dimer against the fibrinogen gamma sequence. The search performed for the dimer formed by Factor XIIIa activity identified one pair of cross-linked peptides at CID 30, which contained residue Q154 cross-linked to residue K380. The identified cross-link was reliable based on the criteria that we had previously defined (Figure 3.28). The scores for this identification were higher than the scores achieved using the synthetic cross-linked peptide and this peptide pair was identified in three MS2 scans for an ion with charge +8. The higher pp value can be explained by the longer peptide length in comparison with the model peptide that leads to an increase in the number of fragments found in the MS2 spectrum. The fragmentation spectrum shows that all of the major peaks had been assigned to a predicted fragment from the identified cross-linked tryptic peptide. The number of abundant peaks was low indicating poor fragmentation efficiency, as previously found for the model peptide (Figure 3.20). Repeating the LC-MS experiment with the collision energy increased from 30 to 35 did not lead to identification of this specific site. Interestingly, the precursor generated a completely different fragmentation pattern after increasing the fragmentation energy and we were unable to identify these ions based on well-established peptide fragmentation patterns. A possible explanation for these unexpected results is that the fragments produced at CID 30 were too unstable at CID 35 and underwent further fragmentation and/or rearrangement. With our synthetic peptide models, we were able to characterise the fragmentation behaviour of a transglutaminase cross-linked peptide with a charge state of no higher than +4. Further studies are needed to predict the stability and fragmentation patterns of more highly charged ions as a function of CID energy.

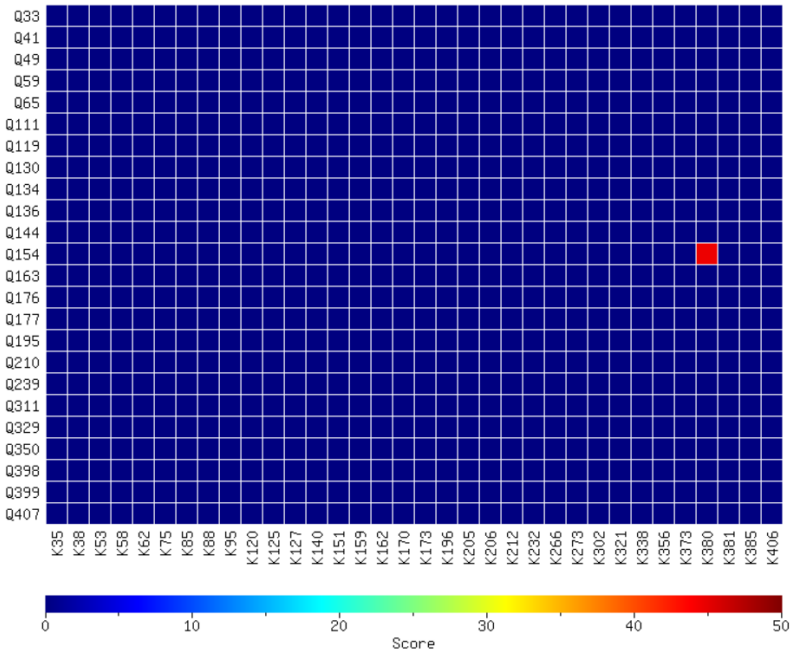
A MeroX search led to the identification of the same peptide pair (data not shown), however although the corresponding score of 31 passed the 5%

FDR cut off, it was considerably lower than the scores reached with the TXLP that were 115 and 72 for the +4- and the +2-charged peptide, respectively.

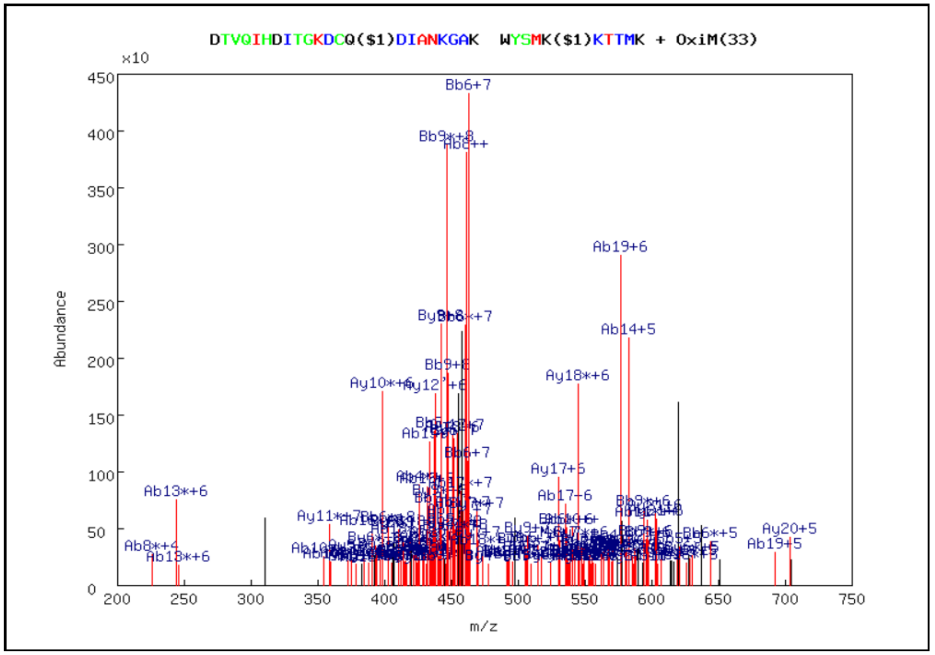
a Fibrinogen gamma sequence:

```
001 YVATRDNCCI LDERFGSYCP TTCGIADFLS TYQTKVDKDL QSLEDILHQV ENKTSEVKQL IKAIQLTYNP DESSKPNMID AATLKSRKML EEIMKYEASI  100
101 LTHDSSIRYL QEIYNSNNQK IVNLKEKVAQ LEAQCQEPCK DTVQIHDITG KDCQDIANKG AKQSGLYFIK PLKANQQFLV YCEIDGSGNG WTVFQKRLDG  200
201 SVDFKKNWIQ YKEGFGHLSP TGTTEFWLGN EKIHLISTQS AIPYALRVEL EDWNGRTSTA DYAMFKVGPE ADKYRLTYAY FAGGDAGDAF DGFDFGDDPS  300
301 DKFFTSHNGM QFSTWDNDND KFEGNCAEQD GSGWWMNKCH AGHLNGVYYQ GGTYSKASTP NGYDNGIIWA TWKTRWYSMK KTTMKIIPFN RLTIGEGQQH  400
401 HLGGAKQVRP EHPAETEYDS LYPEDDL
```

b




c



d

| charge | score | pp | $pp_2$ | $pp_{tag}$ | m/z | MW(obs) | MW | delta | sequence + modifications |
|---|---|---|---|---|---|---|---|---|---|
| +8 | 87 | 18.0 | 12.6 | 7.6 | 466.9731 | 3728.7337 | 3728.8081 | -0.0743 | **DTVQIHDITGKDCQ($1)DIANKGAK WYSMK($1)KTToxMK** |
| +8 | 46 | 14.5 | 7.9 | 3.8 | 466.9732 | 3728.7345 | 3728.8081 | -0.0736 | **DTVQIHDITGKDCQ($1)DIANKGAK WYSoxMK($1)KTTMK** |
| +8 | 55 | 20.0 | 8.5 | 14.2 | 466.9737 | 3728.7386 | 3728.8081 | -0.0695 | **DTVQIHDITGKDCQ($1)DIANKGAK WYSMK($1)KTToxMK** |

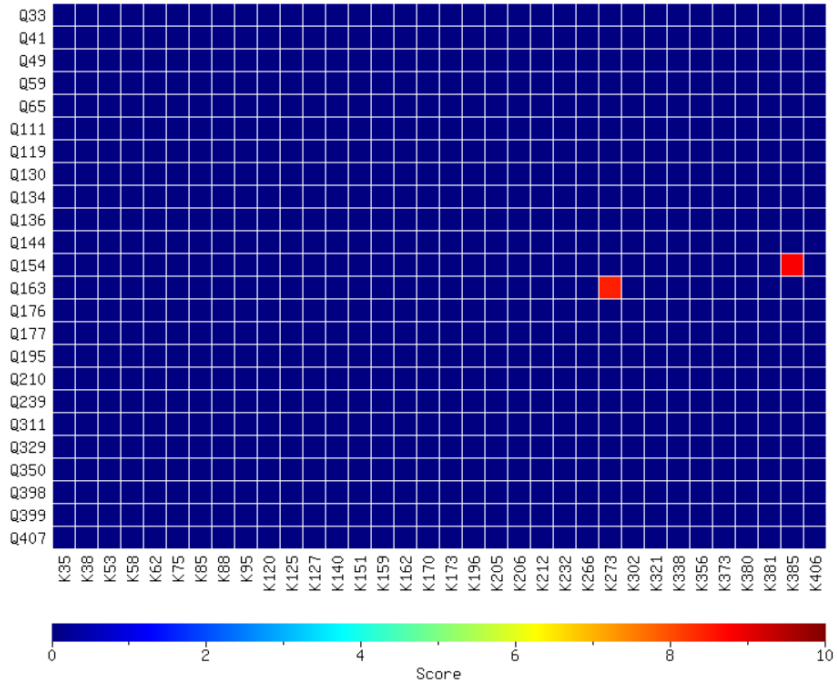
**Figure 3. 28 MassMatrix search result of the Factor XIIIa cross-linked fibrinogen product.** A single cross-link was identified by this search and the positions of the two cross-linked peptides in the fibrinogen gamma sequence are shown in red text (panel a). The heat map from the search result and the MS2 fragmentation spectrum are shown in panels b and c, respectively. The calculated scores for each MS2 scan of the cross-linked peptide are shown in panel d. The exact position of the cross-link is marked with the symbol ($1) after the cross-linked amino acid. Oxidised methionine residues are shown as oxM and MW(obs) and MW values are shown in Da.

The MassMatrix analysis of the TG2_2 ESI-MS/MS data identified two pairs of cross-linked fibrinogen gamma subunit residues (Figure 3.29), but with lower certainty than the Q154-K380 cross-link formed by Factor XIIIa (Figure 3.28). This lower certainty was due to the lower scores and because only one scan contained a positive identification for each of these peptides. In addition, both fragmentation spectra showed the presence of highly-abundant unidentified peaks that raise the probability of a false-positive identification. Furthermore, one of the peptides has been identified by fragmenting a $^{13}C$-containing isotope instead of the monoisotopic peak (Figure 3.29e, second identification). This further increases the probability for a false-positive identification, as the MS instrument was programmed to fragment only monoisotopic species. At this stage there was insufficient information to justify a score threshold to separate true from false identifications. Furthermore, the presence of these cross-links could not be confirmed with a MeroX search.

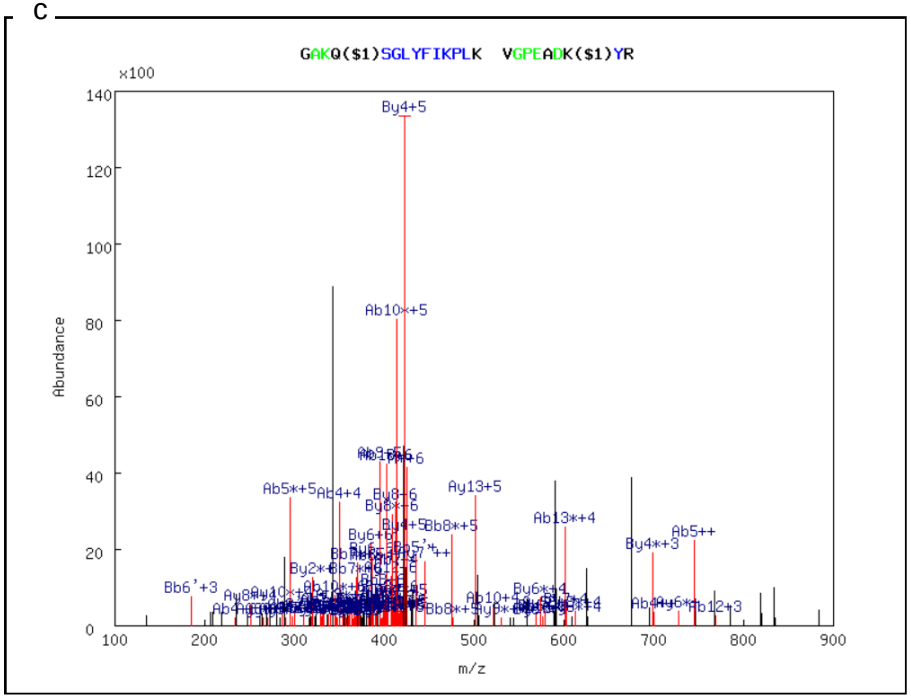
a Fibrinogen gamma sequence:

```
001 YVATRDNCCI LDERFGSYCP TTCGIADFLS TYQTKVDKDL QSLEDILHQV ENKTSEVKQL IKAIQLTYNP DESSKPNMID AATLKSRKML EEIMKYEASI  100
101 LTHDSSIRYL QEIYNSNNQK IVNLKEKVAQ LEAQCQEPCK DTVQIHDITG KDCQDIANKG AKQSGLYFIK PLKANQQFLV YCEIDGSGNG WTVFQKRLDG  200
201 SVDFKKNWIQ YKEGFGHLSP TGTTEFWLGN EKIHLISTQS AIPYALRVEL EDWNGRTSTA DYAMFKVGPE ADKYRLTYAY FAGGDAGDAF DGFDFGDDPS  300
301 DKFFTSHNGM QFSTWDNDND KFEGNCAEQD GSGWWMNKCH AGHLNGVYYQ GGTYSKASTP NGYDNGIIWA TWKTRWYSMK KTTMKIIPFN RLTIGEGQQH  400
401 HLGGAKQVRP EHPAETEYDS LYPEDDL
```

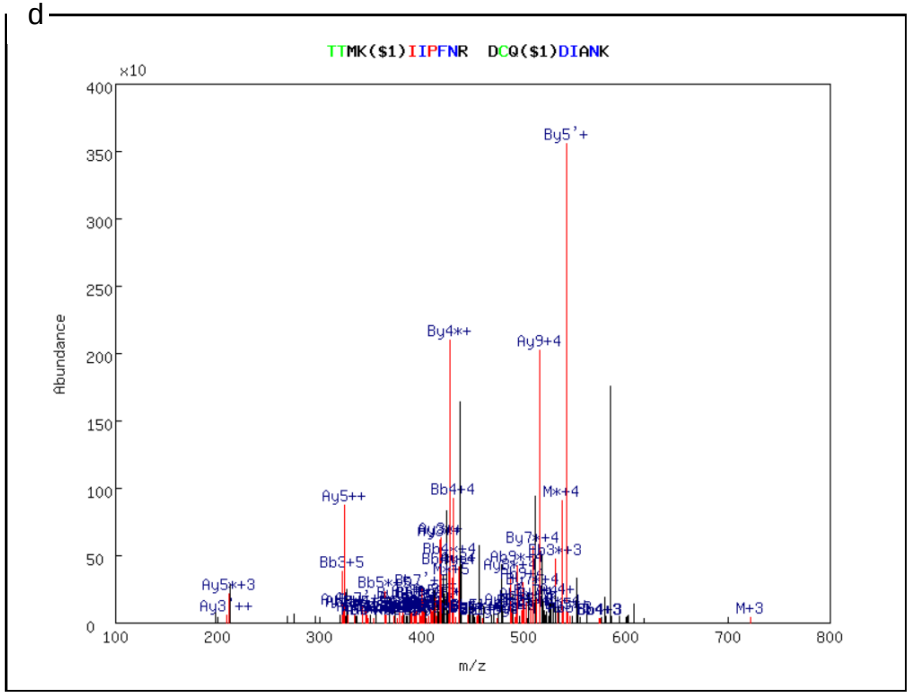
b



c


GAKQ($1)SGLYFIKPLK VGPEADK($1)YR
×100
Abundance
m/z
By4+5
Ab10*+5
Ay13+5
Ab5*+5
Ab13*+4
Ab5++
By4*+3
Bb6'+3

d


TTMK($1)IIPFNR DCQ($1)DIANK
×10
Abundance
m/z
By5'+
By4*+
Ay9+4
Ay5++
Bb4+4
M*+4
Bb3+5
Ay5*+3
Ay3++
M+3

e

| charge | score | pp | $pp_2$ | $pp_{tag}$ | m/z | MW(obs) | MW | delta | sequence |
|---|---|---|---|---|---|---|---|---|---|
| +6 | 11 | 8.8 | 3.7 | 2.2 | 428.5750 | 2566.4133 | 2566.4133 | 0.0105 | **GAKQ($1)SGLYFIKPLK VGPEADK($1)YR** |
| +5 | 16 | 7.0 | 5.8 | 2.4 | 434.2223 | 2167.0822 | 2167.0716 | 0.0106 | ***TTMK($1)IIPFNR DCQ($1)DIANK** |

**Figure 3. 29 MassMatrix search result of the TG2 cross-linked fibrinogen TG2_2 dimers.** Two potential cross-links were found by this MassMatrix search. The positions of the cross-linked peptides in the fibrinogen gamma sequence are shown in red text (panel a) and the corresponding heat map is shown in panel b. The corresponding MS2 spectra of each peptide pair are displayed in panels c and d. The score details for each peptide hit are given in panel e. The sequence of the second identified peptide is marked with an asterisk, indicating that the precursor peptide ion is not monoisotopic and contains one $^{13}C$.

There is no published information on the position of fibrinogen cross-linking sites whose formation is catalysed by TG2, therefore the results from the TG2 experiment cannot be verified in this way. While fibrinogen cross-linking sites whose formation is catalysed by Factor XIIIa have been reported in previous studies, they can't be compared with our result due to important differences in experimental conditions. Purves et al. reported the formation of a cross-linking site within the C-termini of two gamma subunits, but cross-linking was performed only after inducing fibrinogen clotting with thrombin [75]. Clotting causes extensive polymerisation of fibrinogen, which probably provides a different juxtaposition and orientation of fibrinogen subunits relative to each other compared to free fibrinogen hexamers in solution, as well as possible changes in the folding of subunits. These differences would promote the formation cross-links at different sites.

In order to assess the MassMatrix scoring of a true positive cross-linked peptide, we tried to induce and identify the previously reported cross-linking site by clotting and cross-linking fibrinogen using thrombin and Factor XIIIa, respectively. For this experiment the reaction conditions were identical to the published protocol. The MassMatrix search results however did not identify the expected Q398-K406 cross-link. we did however find deamidation of Q398 that resulted from the activity of the enzyme. This deamidation was absent in the negative control sample where clotting was performed in the absence of the Factor XIIIa enzyme. The m/z values corresponding to the cross-linked peptide were not found in the list of MS2 precursor peaks, which suggests that the cross-linked peptide was not formed or too low in abundance to be fragmented.

### 3.3 Discussion

The first aim in this chapter was to demonstrate that naturally-occurring cross-linked peptide models can be synthesized using chemical and enzymatic techniques. These models could then be employed to develop methods for the mass spectrometric identification of peptides with the same type of cross-link. Since chemical synthesis did not depend on the peptide sequence, another application of these model peptides is to use peptide sequences matched to predicted tryptic peptides of target proteins as positive controls in identifying or confirming unmapped cross-linking sites.

The synthesis of HP cross-linked peptides provided difficulties caused by the complexity of the synthesis pathway. This involved several steps including two activation and two coupling reactions. As a result, several intermediate products were expected to be formed and purified from the excess reagents before the next reaction step could be performed. Every reaction and purification step entailed a partial product loss, which appears to have precluded obtaining an adequate yield of the final product.

HP cross-linked model peptides would have been useful for preliminary fragmentation experiments that would have built on the work of previous research groups that had shown ESI-MS/MS fragmentation of peptide backbone using peptides with this type of cross-link purified from tryptic digests [47]. There are however certain limitations of such model peptides. First, their synthesis would result in a cross-linked product containing three identical peptides, which would not be representative of naturally cross-linked peptides. This implies that these model peptides would not be suitable as positive controls as described above, and their use would instead be restricted to understanding fragmentation patterns. Nonetheless, this would still provide useful information for future method development. Second, there

is currently no available software that could automatically identify a peptide triplet cross-linked by a pyridinoline group. Third, we were able to make only limited progress with the synthesis of a model cross-linked peptide triplet, as we could only establish the appearance of the first intermediate using ESI-MS/MS. In future, approaches to synthesise this model peptide could incorporate HPLC purification of intermediates or use activating reagents whose excess can be quenched after the reaction is completed. Another important aspect is achieving a sufficient yield of each intermediate before the next reaction can occur. This is likely to require monitoring, optimising and scaling up each reaction step.

In contrast with the HP reaction, the synthesis of transglutaminase cross-linked peptides showed better results, because the two procedures - chemical and enzymatic - were much more straightforward. The enzymatic method is preferable for most cross-linking reactions, since it is much cleaner (no side products are formed), it can be done in one simple step, and it does not require the use of any toxic reagents. On the other hand, the advantage of a chemical synthesis of the isopeptide bond is that the reaction rate does not depend on peptide sequence, whereas TG2 shows a preference for certain types of amino acid residues adjacent to the glutamine and the lysine residues [129].

The success in the synthesis of peptides with this type of cross-linking allowed me to further investigate their fragmentation behaviour. The reasons underlying the poor data obtained from the fragmentation in the HCD mode remains unclear, but the absence of cross-linked fragments from the spectrum was the reason why we did not use HCD for MS2. Transglutaminase cross-linked peptides from tryptic digests have previously been successfully identified using HCD fragmentation by a Q-Exactive mass spectrometer [78]. Although the Velos and Q-Exactive both rely on Orbitrap

technology, it is likely that the newer Q-Exactive instrument produced a different fragmentation pattern from our Velos instrument. Specifically, closer inspection of the MS2 spectra in this publication reveals that the cross-linked fragments are typically either underrepresented compared to linear fragments, or completely absent. The poor intensity of cross-linked fragment ions corroborates our hypothesis that these species are characteristically less stable than linear fragments. We showed that this poor intensity occurs at different charge states of these fragments, and we suggest that these ions undergo further fragmentations and rearrangements under HCD conditions. Accordingly, these processes would be expected give rise to the high-intensity peaks present in the MS2 spectrum. This hypothesis could be confirmed by a structure analysis of these species using an MS3 experiment.

A strategy for identifying cross-linked peptides via HCD that has been described in the literature is to use a combination of higher and lower energy collision spectra for each identification [138]. In this manner, less stable cross-linked fragments are identified from the low collision energy spectra, while higher energy collision spectra show linear fragment peaks with higher abundance. On appropriate instruments this appears to be an effective way to improve coverage of cross-linked peptide sequences. Exactly which criteria determine the viability of an instrument remain to be defined. However, as we did not detect any cross-linked fragments of our model peptide even at lower HCD energies, we did not adopt this strategy.

The poor fragmentation of the isopeptide bond itself can be rationalised by the mobile proton model [139]. In this model cleavage of a peptide bond occurs via proton transfer from the N-terminus to the site of fragmentation via the peptide backbone. Since the isopeptide bond is isolated from the polypeptide backbone by the hydrophobic hydrocarbon side chains of the glutamine and the lysine residues, proton transfer to this site would be

unlikely. This model is however likely to be at best incomplete as it does not explain the reported cleavability of disuccinimidyl dibutyric urea (DSBU) cross-links by gas collision in the HCD mode [140] (Figure 3.30). It is possible that the MS cleavable group of DSBU is less stable than the peptide bond and does not require a protonation to be fragmented.

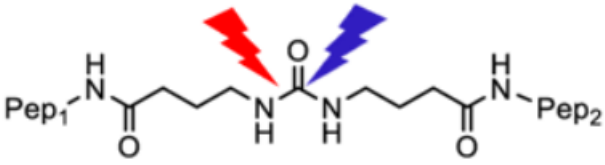



**Figure 3. 30 Illustration of two peptides cross-linked by DSBU.** The HCD cleavable sites are indicated with arrows. (Figure adapted from [140]).

In the experiments described in this chapter we demonstrated that better fragmentation-based coverage of the peptide sequence can be achieved on our instrument using CID than using HCD. This is likely to be due to the better stability of the cross-linked fragments when lower collision energies are applied. However, the lower mass accuracy of the measured m/z ratios and the loss of fragments with an m/z value less than one third of the precursor ion (one-third effect) [141] are inherent disadvantages in comparison with HCD. Hence, using CID fragmentation to identify transglutaminase cross-linked peptides affects the confidence with which the cross-linking position can be identified. In addition, CID provides limited information, as no internal fragments are formed. Thus, the detection of a reporter ion derived from the isopeptide bond is not possible with this method.

Considering the results from the CID analysis of our tryptic, cross-linked, model peptide, we concluded that an energy similar to that standardly used for linear peptides (30-35) in our proteomics experiments gives sufficient fragmentation to identify the sequences of the cross-linked peptides. This

observation seemed consistent across all the investigated precursor and fragment ion charge states. Fragmentation of the +4-charge state showed the formation of fragments with charge states from +1 to +4 with similarly high abundance. Thus, this energy is sufficient to provide good sequence coverage, at least for the model peptide. Considering the dependence of the normalised fragment intensity on the CID energy we inferred that the +2-charged precursor produced qualitatively similar results. However, while the total intensity reached a maximum at a CID energy of 35, it remained low compared to the intensity of the precursor peak. As lower precursor charge states are generally more stable, we can expect that they will be more difficult to fragment. Once fragmented, the stability of a daughter ion does not seem to correlate with the stability (charge state) of its precursor - it most likely depends on the charge of the fragment alone.

A critical question arising from our model peptide and protein studies concerned the scoring threshold we should adopt for subsequent studies for the identification of cross-linked peptides using MassMatrix and MeroX software. The results from the MassMatrix analysis of the model peptide ESI-MS/MS data showed an overall score of 19 could be achieved when the CID energy was set at 30 to achieve optimal fragmentation of transglutaminase cross-linked peptides. In this analysis the overall score for a false-positive identification was 10. Setting a conservative score threshold at 19 in our studies was supported by the results of the fibrinogen cross-linking experiments, as the identified cross-linked peptide from the Factor XIIIa cross-linking reaction produced an average score of 63 from a better fragmentation spectrum than the model peptide (Figure 3.28). In contrast, the samples from the TG2 cross-linking reaction, whose overall score was below this threshold, showed a slightly less convincing spectrum with several unidentified fragment ions. Good spectrum matches can therefore be

discriminated from poor ones by comparing their overall scores. The score in MassMatrix depends on the ratio between the abundance of matched product ions and the total abundance of all ions in the experimental spectrum. It therefore tends to increase with a decreasing number of unidentified fragments. This helps explain the lower score obtained for the TG2 cross-linked identifications in comparison with the Factor XIIIa and model peptide identifications.

In addition to the overall score, a high pp value was another important criterion that we investigated to determine the quality of a match, as it depends on the number of matching fragment ions. All MassMatrix search results obtained from the fibrinogen cross-linking experiment showed high pp values, higher in fact than the model peptide that reached a maximal value of 6.8 at CID 30 (Figures 3.28 and 3.29). However, a false-positive match found in the MassMatrix search results of the model peptide also possessed a considerably higher pp value (11.4) than the true-positive match (6.8) at CID 30.

Another property that can be compared is the pp2 value where again lower values were obtained for the TG2 cross-linked peptide identifications than the Factor XIIIa and model peptide identifications. However, comparing the pp2 values is less reliable, as it only relies on the total abundance of the identified fragment ions without considering their number. This limitation helps explain the high pp2 value of the model peptide as it was mainly determined by the very high intensity of the precursor peak.

In summary, although according to the MassMatrix software developer the pp value should be the main discriminator for quality of matches [123], additional characteristics including the overall score and the number of matching MS2 scans must be considered for the selection of reliable identifications.

Although only a few peptides were studied, the MeroX search results suggest that the score from this program is very sensitive to the charge state of the peptide as the +8-charged cross-linked peptide from the Factor XIIIa cross-linking experiment led to a sudden decrease of the score in comparison with the results obtained from the TXLP. These results suggest that the MeroX search within this thesis should be limited to cross-linked peptides with a charge state of +4 or lower, as only the TXLP provided the expected results for a true-positive match. For the MeroX score analysis of higher-charged ions other peptide models will need to be studied.

In conclusion, the result of the analysis of the Factor XIIIa cross-linking site in fibrinogen is a high-confidence identification on the gamma subunit based on the criteria selected from the model peptide assay, while the results for the TG2 cross-linking site identifications in fibrinogen provided borderline evidence for a different site also located on the gamma subunit. The methodology developed in this chapter was adopted for subsequent cross-linking analysis of more complex biological samples.

# Chapter 4

# Characterisation of ECM cross-linking in pro-fibrotic cell and mouse models

## 4.1 Introduction

Biological models with lower complexity and higher experimental reproducibility than their real counterpart are useful tools for revealing crucial pathways causing disease progression and for discovering new drugs. Due to its simplicity relative to its real biological counterpart, a model can provide particularly useful information in the early stages of research, although any such findings need to be validated using the biological counterpart. A biological model may enable us to directly study a pathway of interest by varying one factor with little interference from other processes, and this can make the results easier to interpret.

Using a cell culture model was beneficial for the early stages of our research into developing a method to explore molecular differences between control and fibrotic ECM as it would be more easily available than biological tissue and would also allow me to study the effects of stimulation with transforming growth factor beta (TGFβ), which is a widely important factor in fibrotic progression. Although useful for method development and identification of molecular changes to the ECM, it remained important to compare results obtained using ex vivo models with those obtained from tissue samples.

Before being able to analyse molecular changes occurring in the ECM, we needed to find an optimal method for ECM extraction from cell cultures and skin tissue. While there was an established ECM proteomics protocol for skin tissue analysis in our group, we were lacking a method for cell culture ECM. Therefore, the first aim of the work described in this chapter was to find an optimal proteomics workflow for ECM analysis from cell cultures. The second

objective was to find a suitable cell culture model for comparison of pro-fibrotic cross-linking changes. There are different cell culture models that resemble different processes involved in SSc progression and our task was to select the ones that were likely to show increased ECM cross-linking. To achieve this, we searched for evidence for an increased expression and activity of different cross-linking enzymes in the ECM. Following the achievements described in the previous chapter, we concentrated our analysis mainly on TG2, since we had previously succeeded in developing a method to identify cross-linking sites arising from the activity of this enzyme on purified proteins. A previous study has demonstrated the proteomic upregulation and the increased activity of TG2 in dermal fibroblast cell cultures as a response to TGFβ-1 stimulation [73]. However, as expression cannot be equated to its extracellular secretion or to increased cross-linking of any specific proteins, no conclusion could be drawn about the cross-linking activity of this enzyme in the ECM. We therefore decided to further study the dermal fibroblast cell culture model as it had shown promising results in terms of TG2 upregulation. We also decided to use two genetic mouse models - tight skin 2 (TSK2) and DNR. The TSK2 mouse model carries a genetic mutation of the *Col3a1* gene that triggers different pro-fibrotic mechanisms in the skin, such as an increased collagen production and TGFβ-1 upregulation [95]. TSK2 mice develop a skin condition that phenotypically resembles skin fibrosis [95]. A fibrotic condition in the DNR mouse model is caused by the selective expression of the kinase-deficient human type II TGFβ receptor (TβRIIΔk) that was expected to lead to a disruption of the TGFβ signalling [98]. However, the reverse effect was observed, as the mouse developed skin and pulmonary fibrosis caused by hyperactivation of this signalling pathway. TGFβ has is a major pro-fibrotic cytokine that triggers numerous fibrotic mechanisms including the overexpression of different cross-linking enzymes [73, 142, 143]. Therefore,

we hypothesized that any fibrotic mouse model with an activated TGFβ signalling cascade can potentially show an increased ECM cross-linking.

## 4.2 Results

### 4.2.1 ECM extraction

#### *4.2.1.1 ECM extraction from dermal fibroblast cell cultures*

To assess the efficiency of different methods designed to prepare ECM proteins from cell cultures, the proteomic results obtained using either trypsinisation or decellularisation with detergent to remove cellular proteins were compared. Initially, this was achieved by comparing the total number of different ECM proteins identified using each method. Isolation of ECM proteins after trypsin decellularisation resulted in 447 protein identifications, of which 84 (19%) were ECM proteins, while the experiment using detergent decellularisation resulted in 711 protein identifications including 63 (9%) ECM proteins. The reproducibility of both methods was first assessed by calculating the number of ECM proteins found in both technical replicates as a percentage of the total number of ECM proteins found in either biological replicate. For the trypsin method the reproducibility was approximately 88%, while for the detergent method the reproducibility was approximately 83%. Reproducibility was also assessed by comparing the iBAQ values of biological replicates. Figure 4.1 shows the scatter plots of the iBAQ values for biological replicates together with the Pearson correlation coefficient for each plot. Pearson's correlation coefficient (r) gives the degree and nature (positive or negative) of a linear relationship between two variables. A value of 1 would imply that all the data on the scatter plot fall on a straight line with

a positive slope [144]. Thus, in this experiment, the closer the value of Pearson's correlation coefficient is to 1, the higher the similarity of the protein iBAQ values between biological replicates. Comparing the Pearson's correlation coefficients from both plots, it can be inferred that isolation of ECM proteins via the detergent decellularisation method shows a significantly higher value (0.858, in comparison with 0.776 for the trypsin method) and is therefore the more reproducible technique for protein quantitation. The data points obtained for the different collagen proteins (marked in red) also show a smaller variation of iBAQ values for replicates generated via the detergent decellularisation method (Pearson's correlation coefficient: 0.825 in comparison with 0.0492 for the trypsin method) suggesting that this technique would provide a better level of reproducibility for the proteins of particular interest to this study.

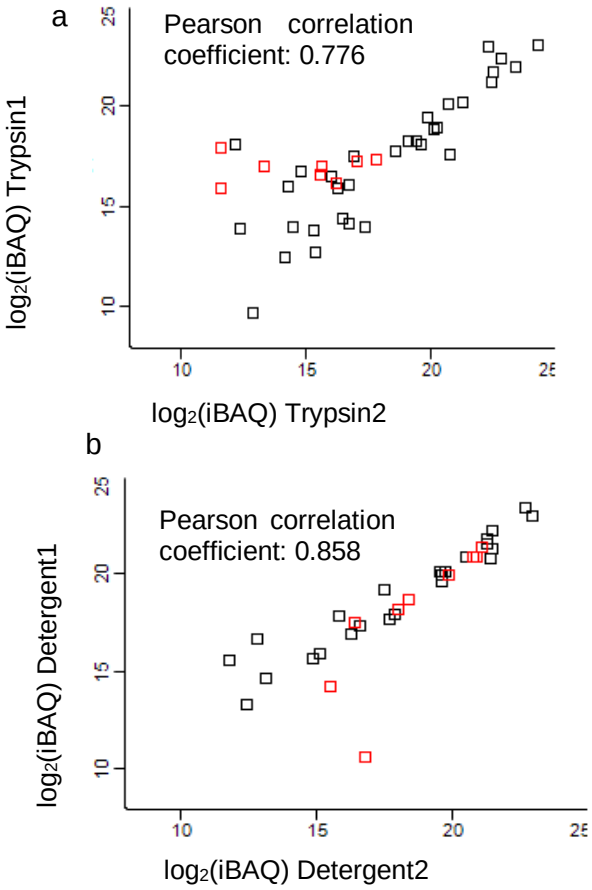


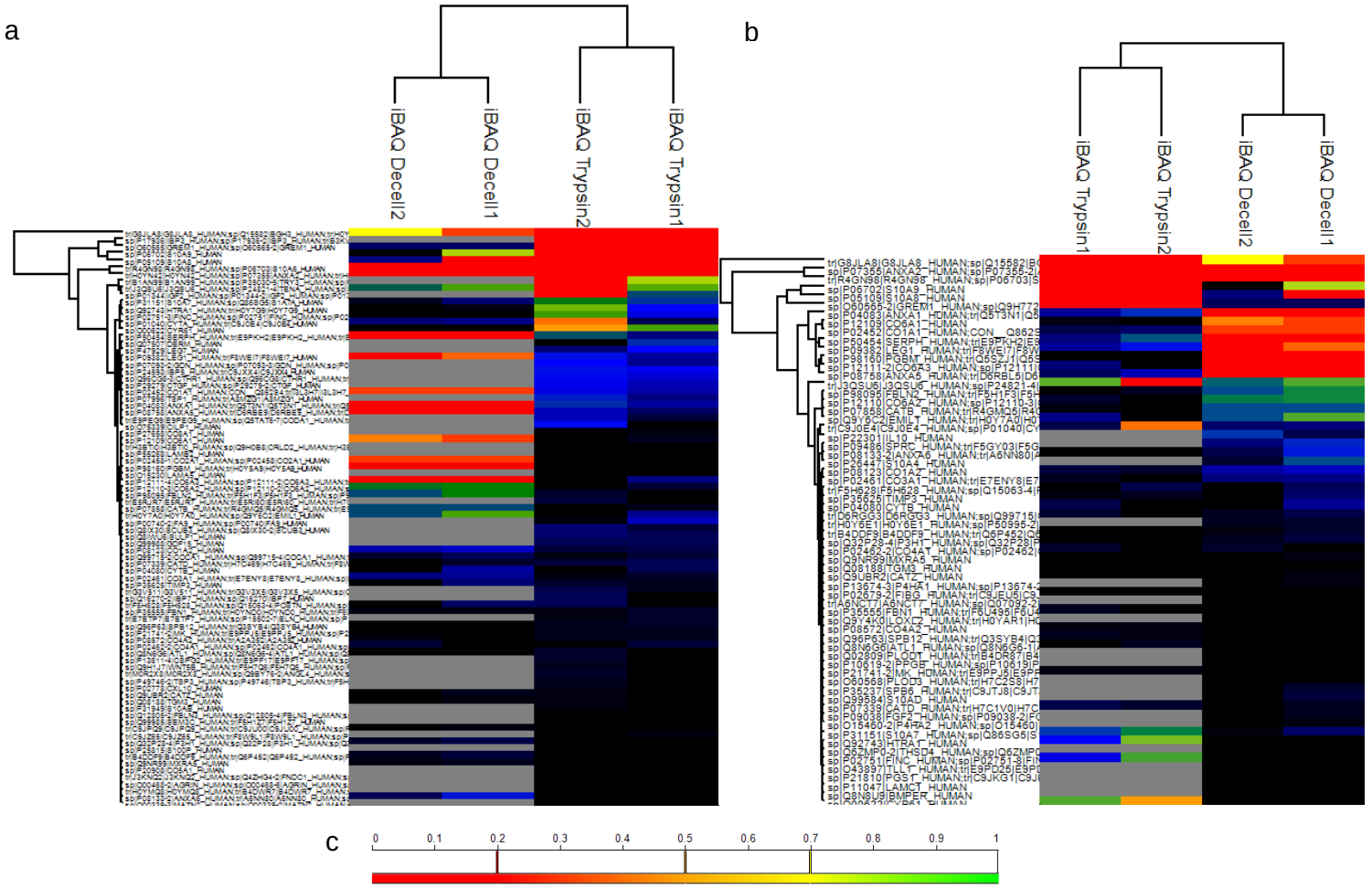

**Figure 4. 1 Scatter plots of the log$_2$(iBAQ) values of the identified ECM proteins.** The scatter plots show only data belonging to proteins that were identified using both methods. In each scatter plot, the values obtained from biological replicates are compared. (a) Trypsin method, (b) Detergent method. The data points belonging to collagens are marked in red.

Figure 4.2 shows the hierarchical clustering of samples and the iBAQ heatmap of all the ECM proteins found in the extracts prepared using the detergent and trypsin methods. ECM proteins identified using each method that were not identified in the ECM extracts using the other method are coloured grey.

These data showed that a large proportion of proteins identified using the trypsin method were not identified using the detergent method (39 proteins, equivalent to 46% of all ECM proteins found using the trypsin method). Among the trypsin-method specific identifications one can find proteins that

are potential substrates for cross-linking enzymes and are therefore of interest to this project, such as collagens V, VIII and XIII and elastin. In addition, trypsin-method-specific proteins included several at relatively high levels of abundance (more than 2 identified peptides).

Analysis of the proteins identified using the detergent method also showed that some ECM proteins were missing compared to the trypsin extracts, including several proteins that might provide important information for the analysis of cell culture ECM cross-linking, such as collagen XVI and lysyl oxidase homolog 2. However, the number of missing proteins is lower (19 proteins, 30% of all protein identifications identified using the detergent method). Several proteins found with a high abundance in the detergent experiment were present in the trypsin method with considerably lower and/or with highly variable iBAQ values (Figure 4.2b). Among these proteins was collagen type I, that is one of the most abundant ECM proteins and a well-known target for cross-linking enzymes [48]. This observation suggests that the detergent method is more effective for the isolation and cross-linking analysis of this protein. Interestingly, no proteins specifically identified using the detergent method fell in the highest abundance range (iBAQ values in Figure 4.2e). From these results it can be concluded that the dermal fibroblast ECM prepared using tryptic cell removal contains a more comprehensive set of ECM proteins compared to the ECM prepared using detergent method, while the detergent method recovers a greater amount of several high-abundance ECM proteins.


a
iBAQ Decell2
iBAQ Decell1
iBAQ Trypsin2
iBAQ Trypsin1
b
iBAQ Trypsin1
iBAQ Trypsin2
iBAQ Decell2
iBAQ Decell1
c
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

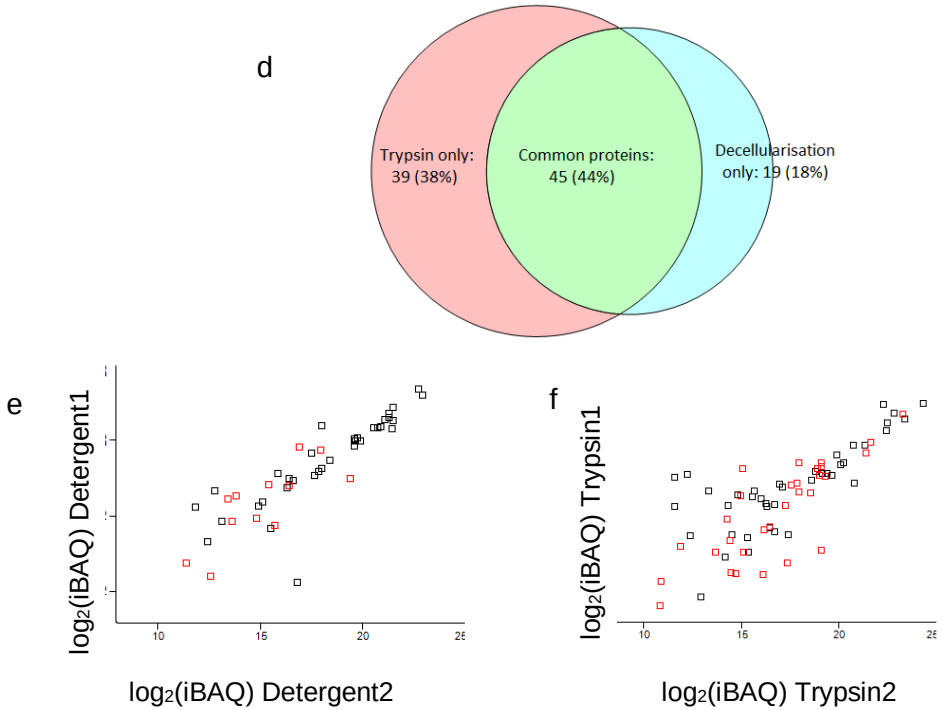



**Figure 4. 2 Cluster analysis and iBAQ heatmap of the ECM proteins identified using the trypsin method (a) and the detergent method (b).** In each panel the intensity of each identified protein is compared with the corresponding intensity obtained using the other method. Grey areas indicate no corresponding protein identification. (c) Colour gradient key for the iBAQ values in panels a and b, (d) Venn diagram showing the overlap between the ECM proteins identifications using both methods. Percentage values represent the proportion of the total number of ECM proteins identified using both methods. (e, f) Scatter plots showing the $\log_2$(iBAQ) values of the ECM proteins that were found using replicate preparations. Proteins identified by only one of the methods are outlined in red.

As the aim of this thesis is to analyse the cross-linking patterns in ECM proteins, a more focused assessment of the results was made for known cross-linking enzyme substrates, such as different types of collagen and elastin. Table 4.1 shows a list of the collagen proteins and their corresponding sequence coverages obtained from replicate preparations resulting from each method. In terms of collagen identifications and coverage it can be noted that the most abundant fibrillar collagens in skin, types we and III [145], were successfully extracted using both methods. The other relatively abundant collagen types that are also present in the ECM of the dermis are the network collagen types IV and VI [145]. Like collagens I (α1 and α2 subunits) and II, collagens IV (α1 and α2 subunits) and VI (α1, α2 and α3 subunits) were present in all analysed ECM extracts. Comparing both techniques in terms of sequence coverage across all of these abundant collagens, it can be deduced that neither of the methods is consistently more efficient than the other for collagen extraction.

A higher number of collagen protein types was identified using the trypsin method (9 types in comparison to 7 using the detergent method). As previously mentioned, isolation of dermal fibroblast ECM via the trypsin method led to the identification of three collagen types (V, VIII and XIII) that were not identified using the detergent method. However, a closer look at the percentages of each protein sequence covered by the identified peptides (sequence coverage) reveals again the poor reproducibility between replicates obtained using the trypsin method: collagens types V and VIII were both not detected in one of the replicates. In addition to these collagens, elastin was also found only using the trypsin method, albeit with a large variation in sequence coverage between replicates (18.7 and 6.8%). The sequence coverage results coming from the detergent method also reveal an unsatisfying reproducibility that reflects in the high variation of the

sequence coverages between the replicates. A possible explanation why the use of trypsin led to a larger number of protein identifications is the improved protein solubilisation due to a partial proteolytic degradation. However, only a modest improvement in the elastin extraction was observed, as the number of identified peptides for the different replicates were 3 and 1 and the corresponding iBAQ values showed a low protein abundance (data not shown).

| ***Protein*** | Sequence coverage (%) Trypsin method | Sequence coverage (%) Detergent method |
|---|---|---|
| Collagen α-1 (I) | 36.4; 29.0 (32.7) | 13; 30.5 (21.8) |
| Collagen α-2 (I) | 23.1;19.9 (21.5) | 7.9; 16.3 (12.1) |
| Collagen α-1 (II) | 2.2; nd (1.1) | 13.3; 0.5 (21.8) |
| Collagen α-1 (III) | 19.9; 13.4 (16.7) | 5.3; 18.3 (11.8) |
| Collagen α-1 (IV) | 2.3; 3.4 (2.9) | 1.2; 3.0 (2.1) |
| Collagen α-2 (IV) | 6.2; 4.8 (5.5) | 1.5; 3.3 (2.4) |
| Collagen α-1 (V) | 1.7; nd (0.9) | nd; nd |
| Collagen α-1 (VI) | 1.8; 1.3 (1.6) | 26.4; 25.0 (25.7) |
| Collagen α-2 (VI) | 4.5; 0.0 (2.3) | 15.4; 14.4 (14.9) |
| Collagen α-3 (VI) | 5.0; 7.0 (2.9) | 31.8; 28.1 (30.0) |
| Collagen α-1 (VIII) | 5.1; nd (2.6) | nd; nd |
| Collagen α-1 (XII) | 3.5; 2.7 (3.1) | 9.8; 5.7 (7.8) |
| Collagen α-1 (XIII) | 2.2; 4.3 (3.3) | nd; nd |
| Collagen α-1 (XVI) | nd; nd | 2.0; 2.0 (2.0) |
| Elastin | 18.7; 6.8 (12.8) | nd; nd |

**Table 4. 1 Sequence coverages of the identified collagen proteins.** The values in each cell are the sequence coverage results from the replicates of each method and the average value (in brackets). For each protein subunit, the higher coverage results are shown in red text. (nd, not detected).

While sample preparation via decellularisation with detergent resulted in the identification of one collagen type, collagen XVI, that was not identified using the trypsin method, only one (unique) peptide was found in the corresponding samples (data not shown). This identification is therefore uncertain and may have been a false positive. The collagen proteins from Table 4.1 that were not identified in the detergent method were present in the data obtained from the trypsin method with poor reproducibility. This was particularly clear for

collagen V and VIII that were present in only one replicate with 2 and 3 identified unique peptides. Figure 4.3 compares the logarithms of the average iBAQ values for collagen subunits identified using either method. In ten of the thirteen polypeptides, the average value obtained using the detergent method are higher, which implies that it provides better yields of most collagen proteins.

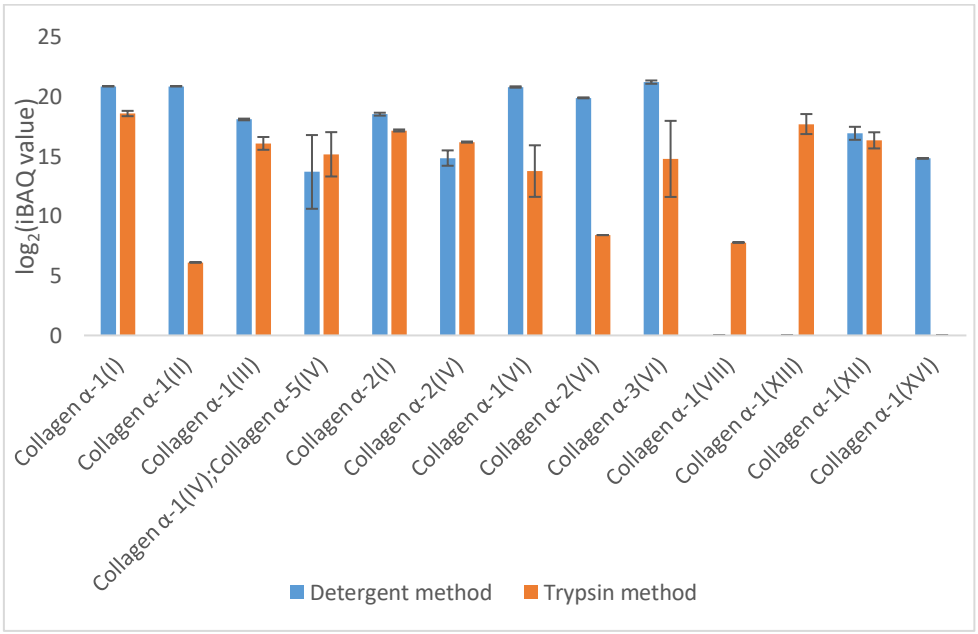



**Figure 4. 3 Comparison of the iBAQ values of all identified collagens between the detergent and trypsin method.** The full range was not given for the data presented for collagen α-1(II and VIII) and collagen α-2(VI) belonging to the trypsin method, and collagen α-1(XVI) for the detergent method, since these polypeptides were identified in only one replicate.

In conclusion, as the detergent method provided higher and more reproducible amounts of collagen (iBAQ values) it was selected as the more suitable method for a quantitative comparison of cross-linked peptides between control fibroblasts and fibrotic myofibroblasts.

#### *4.2.1.2 Optimization of the extraction of ECM from dermal fibroblasts via decellularisation using detergent*

Despite the above presented evidence that the ECM extraction using a detergent gives more reproducible results, the limited number of identified ECM proteins in comparison with the data obtained from the trypsin method remains a major disadvantage. The amount of identified proteins is indicative for the ECM extraction efficiency that is crucial for the extensive comparison in the cross-linking patterns between control and fibrotic ECM. Alongside with the increased sample complexity caused by the presence of cellular proteins, other factors such as the presence of residual detergent suppressing the peptide MS signal, interactions between ECM proteins and the cell culture flask and proteolytic degradation could have a negative effect on the ECM extraction efficiency. To increase the protein yield and thereby obtain better sequence coverage of the main proteins of interest (specifically, all collagen types detected by this method) as well as to identify more ECM proteins, several changes were made to the detergent method. A second aim at this stage was to efficiently remove as much as possible of the detergent after decellularisation, as this appeared as a contaminant in the LC-MS chromatograms of the previous experiments (data not shown).

In order to achieve the first aim three changes were made. First, dermal fibroblasts were cultured in larger flasks that allowed approximately 3-fold more cells to be grown than in previous experiments. Second, protein degradation by endogenous proteases was reduced by the addition of a protease inhibitor cocktail and an increased EDTA concentration of 20mM in the decellularisation buffer. Third, after the cell lysis had completed, scraping the ECM from the cell culture flask in decellularisation buffer followed by centrifugal preparation of a protein pellet was replaced by aspiration of the decellularisation buffer followed by dissolution of ECM proteins attached to the flask surface using 8M urea.

8M urea can act as a solubilising agent of protein aggregates by denaturing them [146]. Therefore, we reasoned that treatment with urea could improve the extraction of the insoluble ECM. In addition to the improved solubilisation, a disruption of the protein interactions with the polymer of the cell culture flask was desirable. To verify whether the use of 8M urea leads indeed to an increased ECM extraction, the results were compared with control samples where this step was omitted.

Table 4.2 and Figure 4.4 summarize the number of peptides and the corresponding iBAQ values for each collagen protein. Overall, based on the lower peptide numbers and iBAQ values for most collagens, it can be concluded that introducing an 8M urea extraction in the sample preparation caused an unexpected loss of these proteins.

| ***Protein*** | ***Peptides with urea*** | | ***Peptides without urea*** | |
|---|---|---|---|---|
| | *Preparation 1* | *Preparation 2* | *Preparation 1* | *Preparation 2* |
| *Collagen alpha-1(I)* | 16 | 10 | 15 | 29 |
| *Collagen alpha-1(III)* | 2 | 5 | 10 | 16 |
| *Collagen alpha-1(IV)* | 2 | 3 | 1 | 2 |
| *Collagen alpha-2(I)* | 4 | 7 | 11 | 16 |
| *Collagen alpha-2(IV)* | 2 | 4 | 4 | 6 |
| *Collagen alpha-1(VI)* | 0 | 1 | 17 | 17 |
| *Collagen alpha-3(VI)* | 7 | 9 | 81 | 71 |
| *Collagen alpha-1(XII)* | 0 | 1 | 26 | 18 |
| *Collagen alpha-2(VIII)* | 1 | 0 | 1 | 0 |

**Table 4. 2 Results of the peptide numbers found in each replicate prepared with and without 8M urea extraction.** The results showing a higher average peptide number for a given polypeptide are shown in red text.

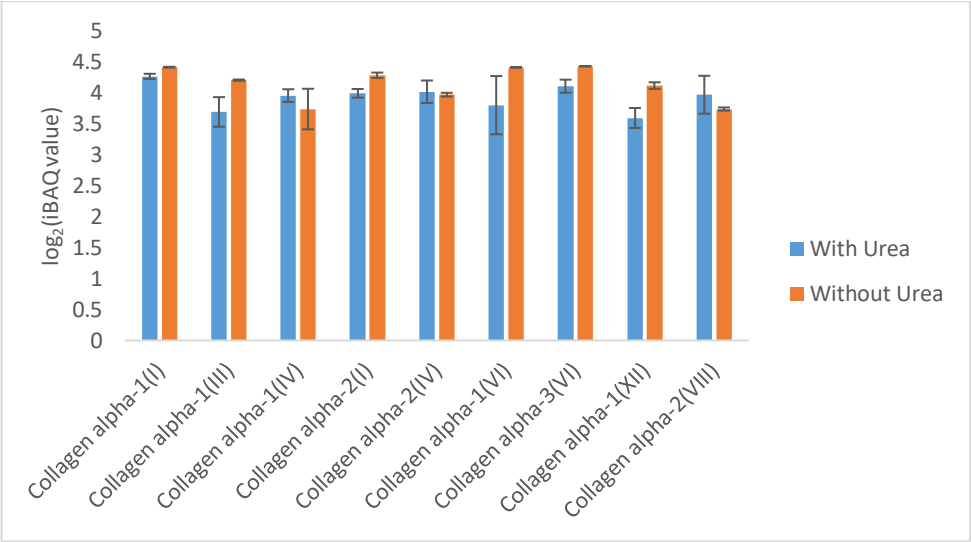


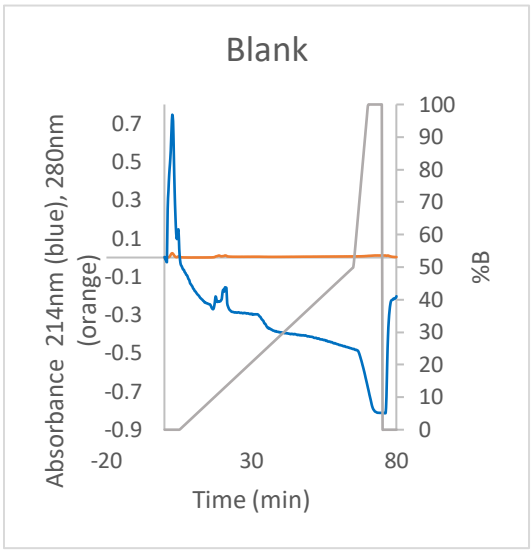

**Figure 4. 4 Comparison of the iBAQ values for the identified collagen subunits obtained using ECM extraction with and without urea.** Since only two replicates were measured for each condition, error bars show the full range of the obtained value.

Surprisingly, the peak intensities for each collagen subunit varied little with and without urea. Moreover, the LC-MS chromatograms revealed a much higher amount of detergent in extracts obtained with 8M urea than those obtained without urea (data not shown). Consequently, we hypothesized that the relatively poor results obtained with urea resulted from signal suppression by unremoved detergent.

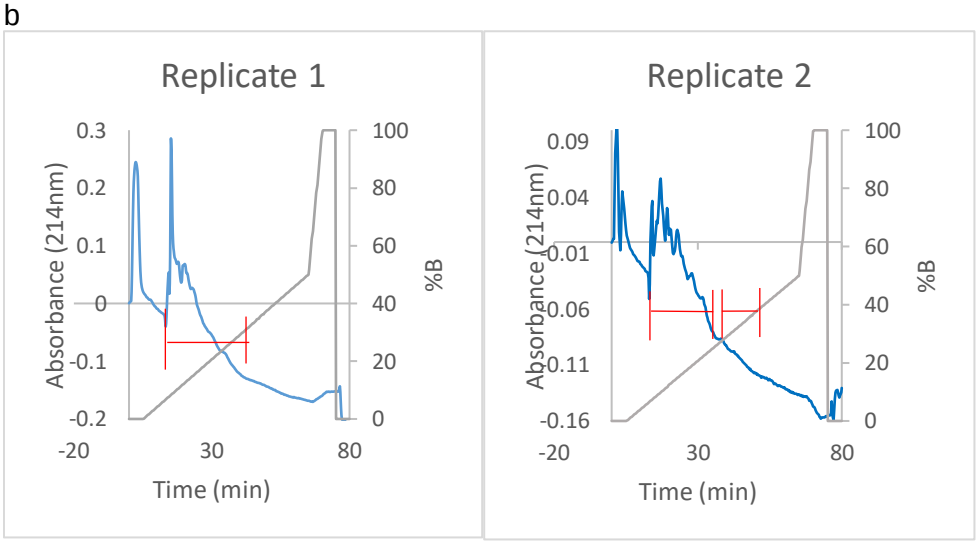
In order to remove more of the contaminating detergent from the samples prior to mass spectrometry, we introduced a purification step that employed strong cation exchange (SCX) chromatography of the urea extract. At mildly acidic pH values tryptic peptides were expected to bind the SCX column, while the uncharged detergent was expected to be found in the flow through. Figure 4.5 shows the chromatograms obtained for each replicate sample. The uncharged detergent was not expected to interact with the stationary phase and therefore to elute before the salt gradient. This was monitored by the absorbance at 214nm where the detergent possesses absorbance (absorbance maximum: 230nm when dissolved in solvent A). As expected, a peak in the flow through was seen in all the chromatograms that

presumably came from unbound detergent. However, detergent was still detected in the LC-MS chromatogram of the peptide-containing fractions that had been collected well after the solvent gradient had started. This suggested that the initial aim of this separation step - to separate the detergent from the ECM peptides - was only partially successful.

a




b




c

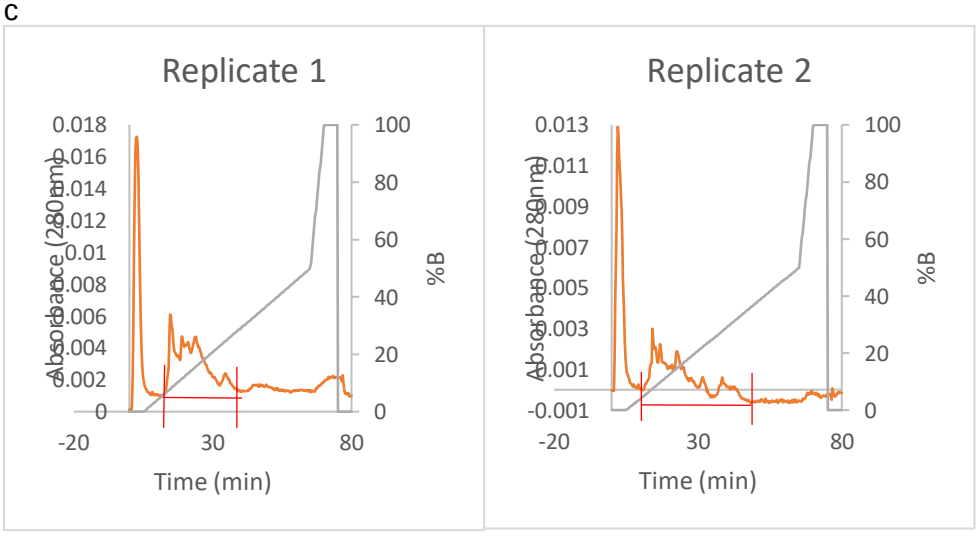



**Figure 4. 5 Strong cation exchange chromatograms of the ECM extracts from control fibroblasts.** The red areas indicate the positions of the fractions that were collected for MS analysis.

Another commonly used method for separation of proteins from a detergent is via precipitation of proteins at -20°C in an organic solvent [147]. Under these conditions, only the detergent is soluble while the proteins form a precipitate and can be isolated via phase partitioning. One clear advantage of this technique is that it can be easily performed. In order to test whether this technique is efficient for detergent removal, another ECM proteomics experiment was performed with an acetone precipitation step introduced immediately after the ECM was separated from the detergent buffer. The collagen extraction efficiency was assessed in both methods - acetone precipitation and urea extraction followed by SCX (Table 4.3 and Figure 4.6). Considering the results presented below, it can be deduced that overall, sample preparation via acetone precipitation leads to a higher number of collagen identifications and a higher peptide number per protein. This is due to the more efficient detergent removal, as a clear drop of the corresponding signal was observed in the LC-MS chromatograms (data not shown). Hence, this preparation technique was used for further experiments.

| Protein | Peptides identified with acetone precipitation | | Peptides identified with SCX+8M urea | |
|---|---|---|---|---|
| | *Preparation 1* | *Preparation 2* | *Preparation 1* | *Preparation 2* |
| *Collagen alpha-1(XII)* | 23 | 28 | 1 | 4 |
| *Collagen alpha-1(XVI)* | 4 | 3 | 0 | 0 |
| *Collagen alpha-1(I)* | 19 | 20 | 27 | 35 |
| *Collagen alpha-2(I)* | 10 | 18 | 21 | 37 |
| *Collagen alpha-1(II)* | 2 | 2 | 2 | 3 |
| *Collagen alpha-1(III)* | 6 | 11 | 10 | 13 |
| *Collagen alpha-5(IV); Collagen alpha-1(IV)* | 1 | 1 | 0 | 0 |
| *Collagen alpha-1(VI)* | 19 | 18 | 5 | 5 |
| *Collagen alpha-2(VI); Collagen alpha-2(VI)* | 15 | 16 | 2 | 3 |
| *Collagen alpha-3(VI)* | 75 | 82 | 8 | 8 |

**Table 4. 3 Peptide numbers found for each protein identified from fibroblast ECM prepared with acetone precipitation and with 8M urea + SCX chromatography.** The results showing a higher average peptide number for a given polypeptide are shown in red text.

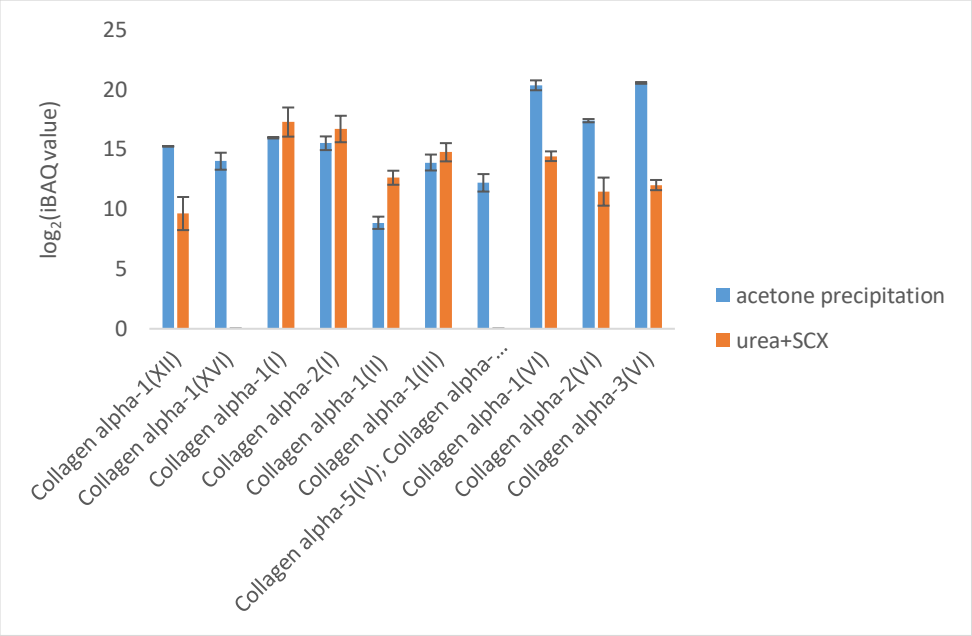



**Figure 4. 6 Comparison of the iBAQ values for the identified collagen proteins obtained using ECM extraction with acetone precipitation and with 8M urea and SCX.** Since only two replicates were measured in each

condition that does not allow statistical tests, the error bars show the full range of the obtained values.

The likely explanation why the ECM preparation strategy using 8M urea and SCX gave poorer results can be found in the presence of detergent the in the urea extracts prior to ion exchange chromatography. Inefficient detergent removal could limit peptide interactions with the stationary phase of the SCX system and the stage tip resulting in sample loss.

The SCX chromatograms also contain material with absorption at 214 and 280nm in the flow through fraction. These peaks might have arisen from peptides that interacted weakly with the stationary phase. Loss of weakly-bound peptides was investigated as a possible reason for the overall reduction in protein identifications. One of the SCX fractions in this region of the chromatogram (fraction 3) was analysed using LC-MS. To prevent a potential overload with detergent and/or tryptic peptides, 10% of the fraction volume from each replicate was used and purified with a stage tip. The MaxQuant results revealed that no peptides derived from ECM proteins were present in either replicate fraction, which suggested that sample loss did not result from insufficient peptide binding to the stationary phase, although sample loss during stage tip purification caused by the presence of detergent cannot be excluded.

Further evidence that indicated that the SCX preparation step was not responsible for the observed sample loss was that no back-pressure increase was observed during chromatography (a constant pressure of around 4MPa was recorded throughout the chromatography). Thus, we found no evidence for peptide insolubility at increasing salt concentrations and that the peptides attached to the column could not be effectively eluted with the applied salt gradient.

To confirm that SCX chromatography did not cause protein loss we measured the absorbance at 280nm of a specific amount of a tryptic rat-tail collagen digest and compared it with the absorbance of the SCX fractions in the same total volume. This experiment indicated no decrease in the amount of peptide after chromatography, and we therefore concluded that the partial loss of ECM peptides was not caused by the chromatography system.

As platform access was limited, we also attempted to reduce the MS analysis time for each sample by reducing the number of Offgel isoelectric focussing (IEF) fractions from 12 to 6. Table 4 shows the data obtained after pooling pairs of neighbouring fractions. This led to a significant deterioration of the MaxQuant results that indicates that the increased sample complexity (number of peptides per fraction) caused an inefficient separation on the LC-MS. Thus, the 12-fraction separation was used in subsequent experiments.

| *Protein* | $log_2$(iBAQ) 12-fraction sample | $log_2$(iBAQ) 6-fraction sample | peptide number 12-fraction sample | peptide number 6-fraction sample |
|---|---|---|---|---|
| *Collagen alpha-6(VI)* | 12.19 | nd | 3 | 0 |
| *Collagen alpha-1(I)* | 25.03 | 23.90 | 32 | 20 |
| *Collagen alpha-1(II)* | 14.95 | nd | 2 | 0 |
| *Collagen alpha-1(III)* | 22.56 | 20.60 | 18 | 4 |
| *Collagen alpha-2(V)* | 15.67 | nd | 2 | 0 |
| *Collagen alpha-2(I)* | 23.91 | 22.71 | 26 | 13 |
| *Collagen alpha-2(IV)* | 19.82 | 17.93 | 2 | 3 |
| *Collagen alpha-1(XI)* | 14.75 | 13.58 | 1 | 2 |
| *Collagen alpha-1(VI)* | 24.24 | 23.48 | 16 | 11 |
| *Collagen alpha-2(VI)* | 23.23 | 22.09 | 13 | 11 |
| *Collagen alpha-3(VI)* | 25.02 | 23.67 | 65 | 47 |
| *Collagen alpha-1(V)* | nd | nd | 4 | 1 |
| *Collagen alpha-3(V)* | nd | nd | 2 | 1 |
| *Collagen alpha-1(XVIII)* | 13.92 | nd | 1 | 1 |
| *Collagen alpha-4(IV)* | 19.20 | 17.93 | 1 | 1 |
| *Collagen alpha-1(XIV)* | 17.86 | 12.91 | 6 | 0 |
| *Collagen alpha-1(XVI)* | 18.51 | 17.62 | 5 | 4 |
| *Collagen alpha-1(XXIV)* | 17.59 | nd | 1 | 1 |
| *Collagen alpha-1(XXVIII)* | 18.18 | 16.89 | 0 | 3 |
| *Collagen alpha-1(XIII)* | 12.31 | nd | 1 | 0 |
| *Collagen alpha-1(XXVII)* | nd | 17.57 | 2 | 1 |
| *Collagen alpha-1(XII)* | 20.67 | nd | 44 | 31 |
| *Collagen alpha-1(XXV)* | 11.64 | 12.57 | 1 | 0 |
| *Fibronectin* | 22.96 | 21.70 | 63 | 56 |

**Table 4. 4 MaxQuant search results from the samples prepared using 6 and 12 Offgel IEF fractions.** The result for each polypeptide showing a higher peptide number or iBAQ value is shown in red text. (nd: not detected)

#### *4.2.1.3 ECM extraction from mouse skin tissue*

At the beginning of this project isolation of ECM proteins from mouse skin tissues was performed using a protocol that led to the identification of up to 78 ECM proteins (1% FDR). This number is significantly lower in comparison with the results from other protocols used for ECM extraction from tissues that led to the identification of several hundred ECM proteins [147, 148]. Thus, our preparation procedure needed improvement. The limitations of this technique were mainly caused by insufficient removal of the detergent from the protein pellet after decellularisation and the insufficient separation of peptides in each IEF fraction by the inline HPLC column prior to MS detection.

For reasons of space these improvements are summarised here. To efficiently remove residual detergent from the ECM pellet a precipitation step in 80% acetone was introduced before collagenase digestion. In this method, the separation principle relies on solubility differences – while the ECM proteins remain as a precipitate the detergent is solubilized. In addition, peptide separation using HPLC was improved by extending the elution gradient from one to two hours allowing me to identify more tryptic peptides in each fraction. As a result of these changes, the number of ECM protein identifications was raised to 150 and the sequence coverages were improved. For instance, the sequence coverage for collagen I was increased from 46% to 63%. A more comprehensive study of the collagen identifications revealed that the number of identified collagens had doubled in comparison with the results from the previous method (18 different collagen polypeptides in comparison with 9). Moreover, the signals arising from the detergent on the LC-MS chromatogram were considerably lower, suggesting that detergent removal had been successful.

Regarding the reproducibility of the data, approximately 95% of the proteins identified with high confidence (proteins with ≥2 unique peptides at an FDR of 1%) were found in both technical replicates of each biological sample. A scatter plot of the iBAQ values calculated for two technical replicates from the same biological sample (Figure 4.7) showed that the method had a high level of reproducibility (Pearson correlation coefficient: 0.925). 17 types of collagens were identified from this experiment (collagen I-XII, XIV-XVII, XXVI), including the most common collagen types in skin (I, III and V) as well as other collagen types that are known to be essential for normal skin functioning and are therefore expected to be present in the ECM extract (IV, VII and XVII) [145]. A comparison of the dataset with collagen mRNA data from human skin tissue (Protein Atlas, see Appendix 1) led to the conclusion that we had successfully identified the majority of collagen protein types in skin. In addition to these potential targets of cross-linking, several cross-linking enzymes such as different transglutaminase isoforms (Tgm1, 2, 3 and 6) and lysyl oxidase homolog 1 (Loxl1) were also identified in skin ECM.

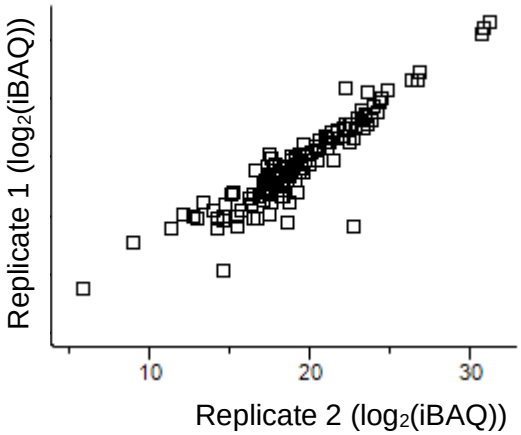



**Figure 4. 7 Scatter plot of the $log_2$(iBAQ) values of the ECM proteins identified from mouse skin samples after changes were introduced to the sample preparation protocol.**

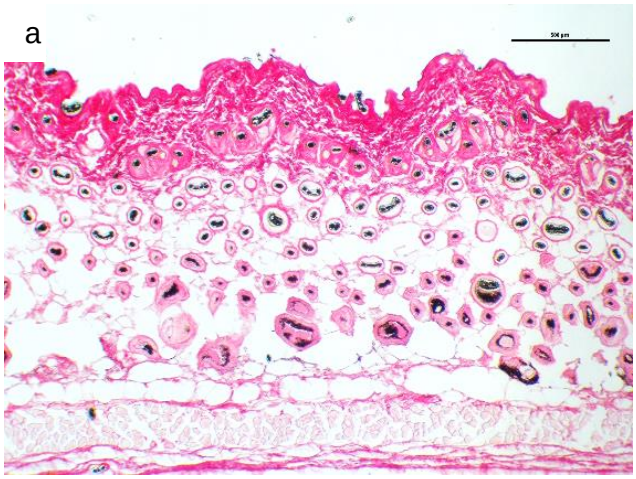
### 4.2.2 Characterization of pro-fibrotic mouse models of scleroderma

As described in the introduction (Section 1.7), there are several animal models for systemic sclerosis available that mimic different aspects of the disease: vasculopathy, inflammation and fibrosis. Since there is no established model expressing the whole range of pathways leading to disease progression, an appropriate animal model must be selected for the study of each molecular mechanism of interest. Fibrotic development is a complex pathologic event whose progression and irreversibility relies on numerous processes leading to ECM overproduction and impaired catabolism. Due to the simplicity of the animal models, showing a fibrotic phenotype is not a sufficient evidence for the presence of a certain fibrotic mechanism. Thus, prior to the study of changes in the ECM cross-linking pattern in a particular mouse model, preliminary evidence about the increased activity of cross-linking enzymes in the ECM is required. This can be achieved by analysing their expression in the ECM on a protein level and/or by comparing the amount of a cross-link product with an appropriate control [49]. Since there is no current evidence on the increased transglutaminase cross-linking in the available SSc models, the aim of this section was to analyse the transglutaminase expression in the ECM in the skin of two fibrotic SSc models. An increase of the enzyme levels in the ECM of the mouse models could serve as an indication for the hyper activation of the corresponding cross-linking pathway. The TSK2 and DNR mouse models that we used for our investigation were selected based on their accessibility. Furthermore, evidence about the increased expression of TGFβ in this models that can be a cause for and a consequence of the increased ECM cross-linking [93, 98] led to the hypothesis that they can be useful for the aims of this project.

#### *4.2.2.1 TSK2 mouse*

For this study, we obtained control and TSK2 mouse skin samples from 10- and 23-week-old mice. To confirm that these TSK2 mouse skin samples had fibrosis, we first stained sections with haematoxylin and eosin and with Picrosirius Red. The images showed a slight thickening of the dermal layer in TSK2 relative to control skin after 23 weeks that is characteristic for this mouse model [91, 95]. However, an ImageJ analysis of the Picrosirius Red stained areas did not show an increased collagen I synthesis. A previous study on 10-days old TSK2 male mice however showed a much more pronounced fibrosis from the clear increase of the dermal thickness accompanied with excessive accumulation of dense collagen fibres [91], suggesting that TSK2 male mice are a better fibrotic model. Increased density of the collagen fibres and the higher number of nucleated cells in the dermis of the TSK2 skin are consistent with activated fibroblast proliferation and increased production of collagen fibres that are typical features of fibrotic tissue.

a



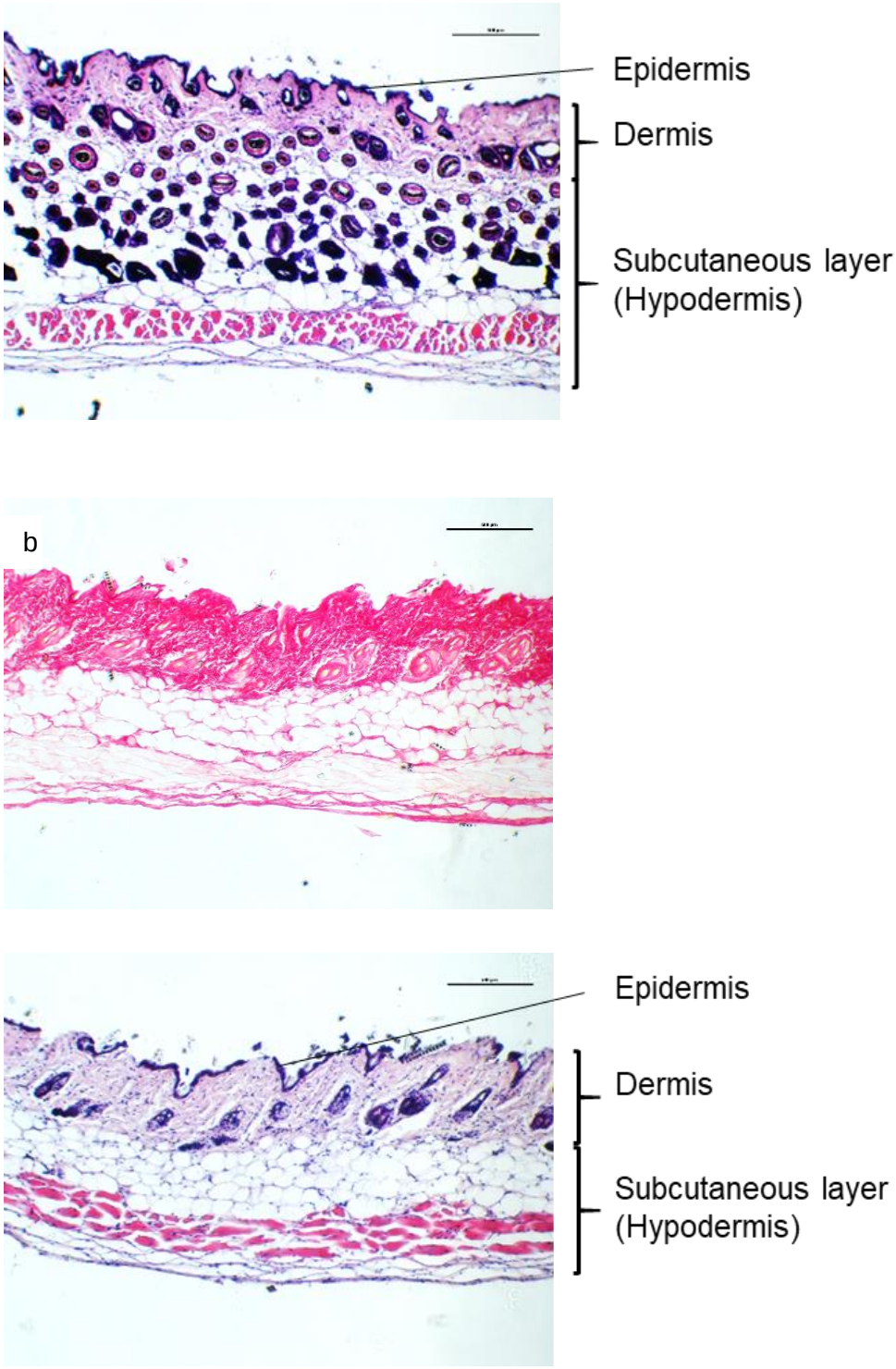



**Figure 4. 8 Histochemical staining of 23-week-old TSK2 mouse skin tissue.** Typical control (a) and TSK (b) mouse skin sections are shown after staining with Picrosirius red (top) or haematoxylin and eosin (bottom).

To compare the levels of cross-linking enzymes between the control and TSK2 samples, we looked at the unnormalised and normalised protein levels. The unnormalised values were the total iBAQ values calculated by the

MaxQuant software that were representative of the total amount of each protein. Normalised values were obtained after transformation of these values across the whole sample dataset, so that the lowest iBAQ value was normalised to 0 and the highest value to 1. Thus, the normalised values provided information about the amount of a protein relative to the total amount of all the other identified ECM proteins. This fractional value was independent of the total amount of protein used for the measurement. Thus, while an increase in the unnormalised protein value could have been simply a consequence of an increase in total ECM protein, comparing the normalised values of individual proteins allowed me to draw conclusions about the enrichment of specific ECM proteins.

Data obtained from the proteomics experiments with skin samples from 10- and 23-week-old TSK2 mice revealed identification of the cross-linking transglutaminase (TG) 1, 2 and 3, but no lysyl oxidases or hydroxylases were identified. The results revealed no consistent increase across 10- and 23-week samples relative to control mice for any of the identified transglutaminases (Figures 4.9 and 4.10). While the TG2 iBAQ levels in the 10-week-old mice did show a modest increase (statistically insignificant), no such change was found in the 23-week-old mice. These data might be explained by a problem in the sample preparation - the 23-week-old mice experiment showed very poor results in one of the replicates of the TSK2 sample, where a smaller number of identified proteins and generally lower iBAQ values were observed.

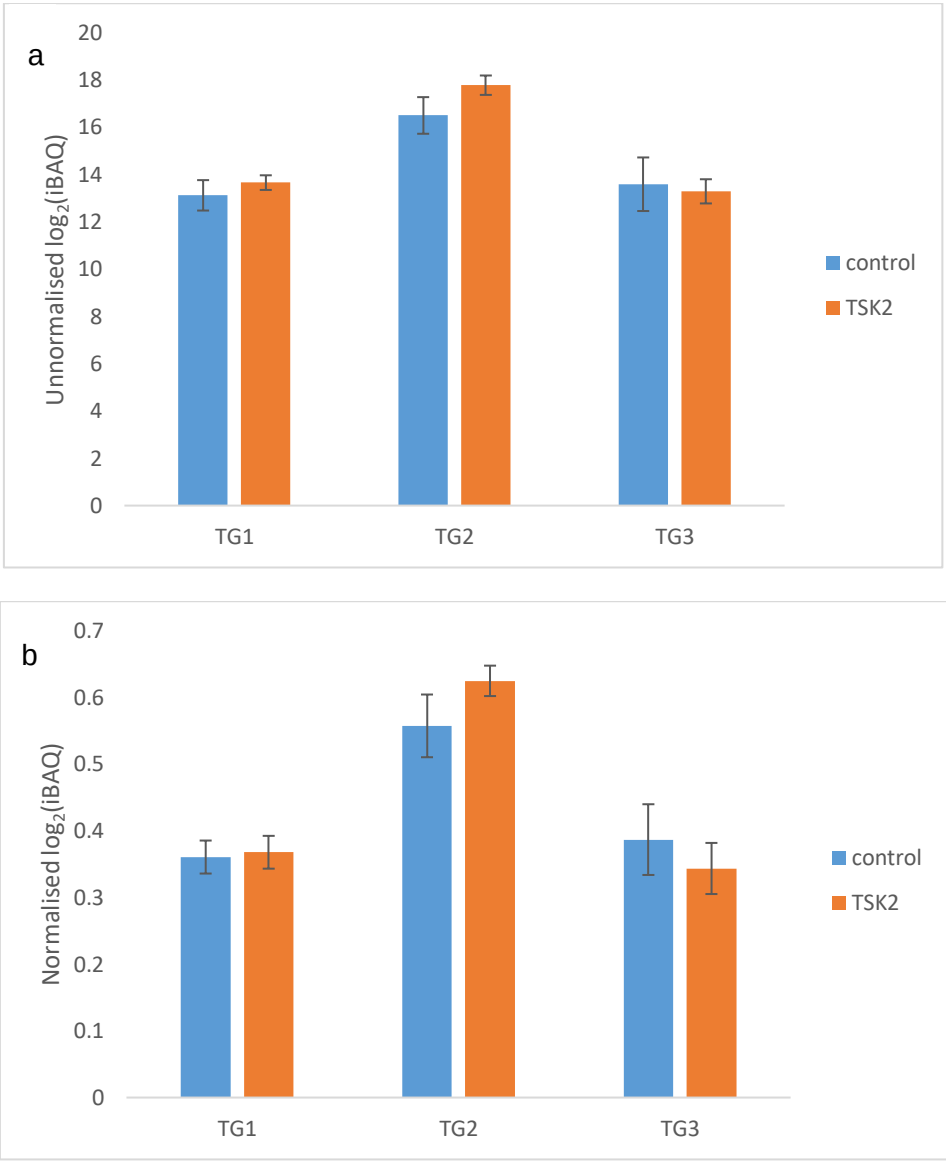



**Figure 4. 9 Transglutaminase proteomics results of the 10-week-old TSK2 mouse skin sample.** (n=4) a: normalised values; b: unnormalised values

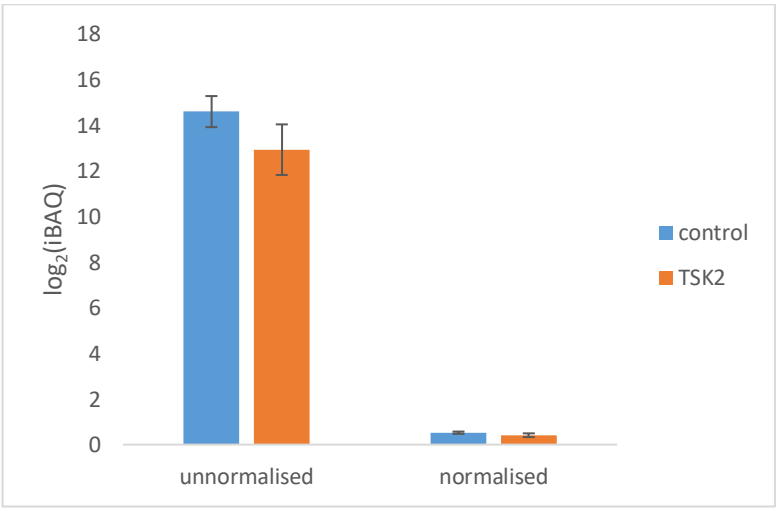



**Figure 4. 10 TG2 proteomics results of the 23-week-old TSK2 mouse skin samples.** (n=4)

The results were validated using Western blotting. Although PLOD2 is typically found in the cytoplasm, it is also secreted to the extracellular space [149]; thus, we expected to find bands in the ECM preparation (pellet) and the cell lysate (supernatant). However, while the supernatants showed signals for PLOD2 (Figure 4.11) and TG2 (Figure 4.12), the PLOD2 and TG2 levels in the ECM seemed to be lower than the detection limit. Moreover, we could not accurately quantify PLOD2 levels due to sporadic variation in the efficiency of protein transfer and detection, evident for example in Figure 4.11 in the first two lanes (from left to right). An additional issue was that the main detected band had an apparent molecular weight (approximately 55kDa) lower than expected (85kDa). A likely explanation for this was protein degradation. The presence of several protease-sensitive regions of this lysyl hydroxylase has been reported in previous literature. In particular, the removal of a 30kDa fragment as a result of proteolytic degradation [150] corresponds to the mass difference between the theoretical and the experimental molecular weight of our Western blot experiment. The sample was prepared according to an established sample preparation procedure from cell cultures that limits protein degradation by including protein inhibitors in the decellularisation buffer. However, we could not inhibit protein degradation occurring within excised skin tissue before the samples were received. Oddly, the TG2 blots showed a band with a molecular weight (130kDa) higher than expected (77kDa). This observation might be explained either by the low specificity of the primary antibody causing a failure to detect the protein of interest, or by a covalent TG2 interaction with another cell component that was not cleaved. As previous experiments with this antibody had shown a high level of specificity, the second hypothesis appears more likely. In this regard, TG2 has a propensity to homodimerize and form intermolecular cross-links in the presence of calcium [151]. The apparent size of this dimer on SDS-PAGE in this study is very similar to the

130kDa band we detected in our experiments. Lysyl hydroxylase and TG2 are apparently more abundant within cells than in the ECM, as they were detected only in the supernatant fractions. Comparing the signals from the Western blots, we found that the signal intensities of both enzymes in the TSK2 samples were lower than those in the corresponding controls. We deduced that there was no evidence for upregulation of either enzyme, although the results remained insufficient to draw definite conclusions. We could not detect any GAPDH signal with this blot, therefore the only normalisation was the use of an equal starting weight of skin tissue for each sample. In addition, the absence of a 77kDa-TG2 signal from the ECM extract did not allow me to orthogonally confirm our findings from the proteomic data.

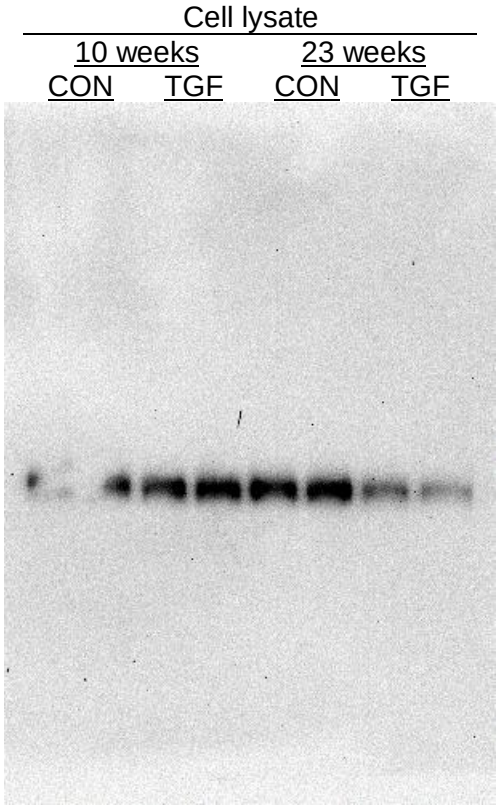



**Figure 4. 11 PLOD2 Western blot of TSK2 mouse skin samples. Each condition was analysed using biological duplicates.** The predicted molecular weight of PLOD2 is 85kDa.

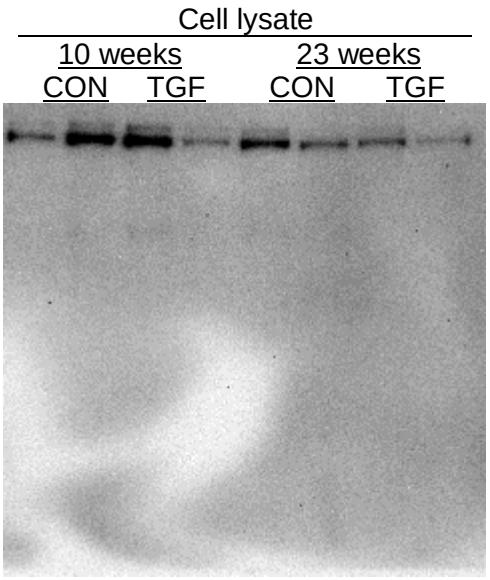
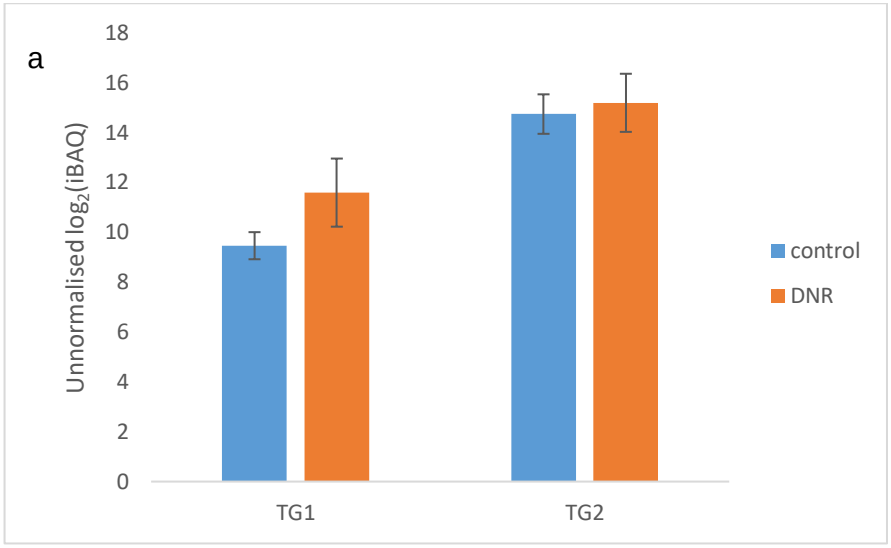
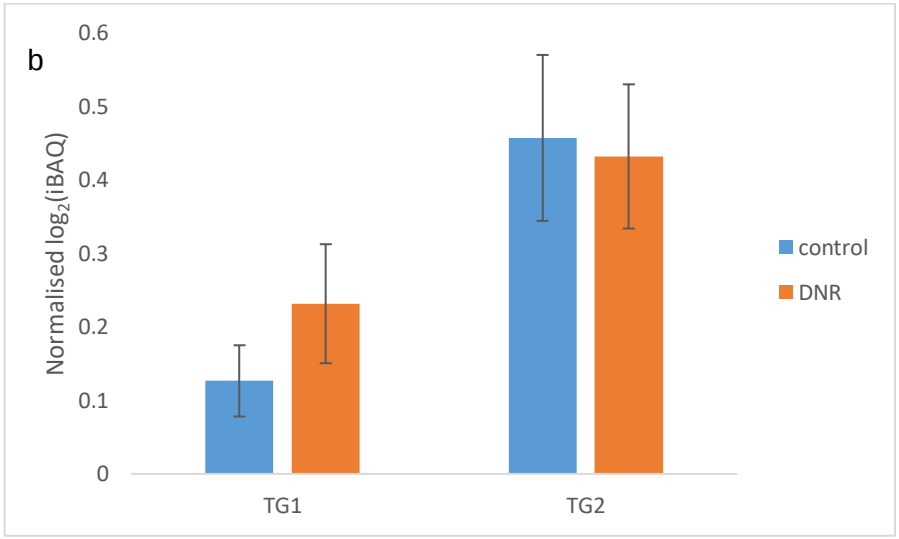



**Figure 4. 12 TG2 Western blot of TSK2 mouse skin samples. Each condition was analysed using biological duplicates.** The predicted molecular weight of TG2 is 77kDa.

#### *4.2.2.2 DNR mouse*

The analysis described in the previous section was also performed using samples of 23-week-old female DNR and control mouse skin. A skin histology analysis of this transgenic mouse line in a previous study revealed thickening of the dermis from the age of 12 weeks regardless of sex [98]. This observation was accompanied by an increase in the content of non-cross-linked fibrillar collagen in 14-week-old male mice. Therefore, we inferred that fibrosis was already established in the obtained samples. Similar to the TSK2 skin ECM proteomics results, the DNR ECM data revealed identification of the cross-linking TG1, 2 and 3, while no lysyl oxidases or hydroxylases were identified. Quantitative comparison of DNR and control data indicated an increase in TG1 and 3 in DNR skin (Figure 4.13). However, the statistical significance of these findings could not be determined, since limited platform time restricted their measurement to duplicates (one technical replicate for each biological replicate). In terms of TG2 expression, while the protein level seemed to increase (Figure 4.13a), no enrichment in the ECM was observed (Figure 4.13b).



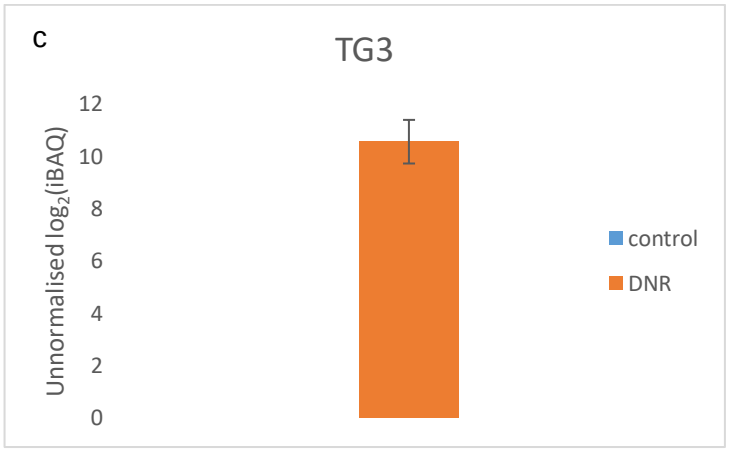




**Figure 4. 13 Transglutaminase proteomics result from DNR mouse skin samples.** (n=2) a: unnormalized values; b: normalized values; c: TG3 results showing that the protein was absent from the control and appeared in the fibrotic replicates.

Western blot analysis of the DNR ECM samples showed no detectable TG2 or GAPDH. PLOD2 was detected, but no difference in abundance was apparent in either the ECM or the cell lysate relative to control samples (Figure 4.14). The apparent size of the PLOD2 (60kDa) was again lower than expected and similar to the 55kDa band detected in TSK2 mice (Figure 4.11), consistent with proteolytic degradation.

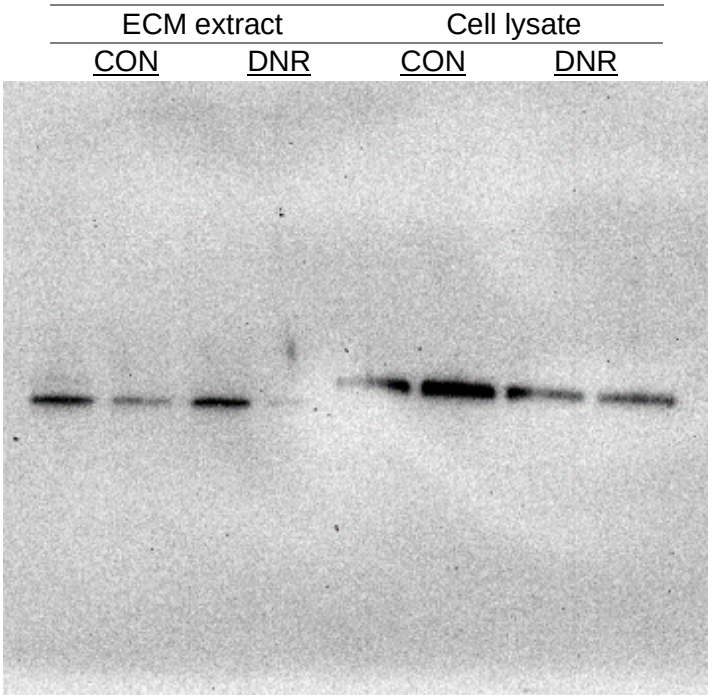



**Figure 4. 14 PLOD2 Western blot of DNR mouse skin samples. Each type of sample was analysed in biological duplicates.** ECM extracts and the corresponding lysates are shown for control and DNR skin. The expected molecular weight of PLOD2 is 85kDa.

Considering the proteomic and Western blot data obtained from the two mouse models, we found no evidence for upregulation of any cross-linking enzymes in the TSK2 mouse model. Hence, we deduced that there was unlikely to be any increase in skin ECM cross-linking in this model. In contrast, TG1 and 3, but not TG2, appeared to be more abundant as well as enriched in the DNR skin ECM, although the upregulation was not statistically testable. A possible reason for the limitation to TG1 and 3 in one model might be that that the measured number of biological replicates was insufficient for detecting a significant difference in TG2 and PLOD2 in DNR samples, and in

all enzymes in the TSK2 model. Providing a large number of replicates per condition is particularly important for the detection of small protein changes. Alternatively, the fibrotic stage of the analysed samples might be too early to observe a significant hyperactivation of the ECM cross-linking pathway. Although there is considerable information about the role of TG2 as an ECM cross-linking enzyme, it is known that TG1 and 3 are also present in the ECM [152]. Therefore, the potential ECM cross-linking activity of these forms deserves further study.

### 4.2.3 Dermal fibroblast cell culture model

For the cell culture model, dermal fibroblasts were isolated from a healthy 41-year-old female donor. The reason for the selection of this donor age is that it matches the typical age of SSc patients. Indeed, this donor age is close to the average age of SSc diagnosis [153].

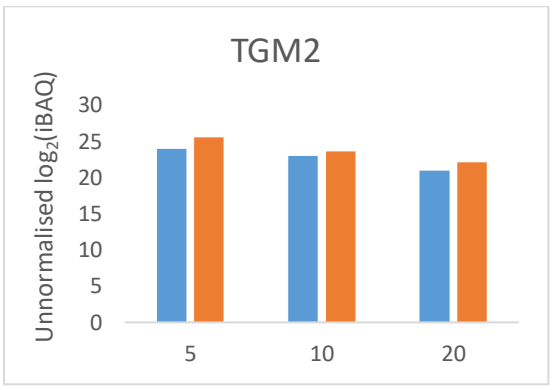
TG2 upregulation in dermal fibroblasts as a response to TGFβ-1 stimulation had previously been demonstrated in whole cell lysates [73]. Based on this evidence, we hypothesized that the observed upregulation could be present in the ECM as this isoform is a well characterised ECM cross-linking enzyme [152]. In addition, since the TGFβ pathway is also associated with the increase of other types of cross-linking enzymes [142, 154], it was possible that an accumulation of these enzymes and their corresponding products were also present in this model.

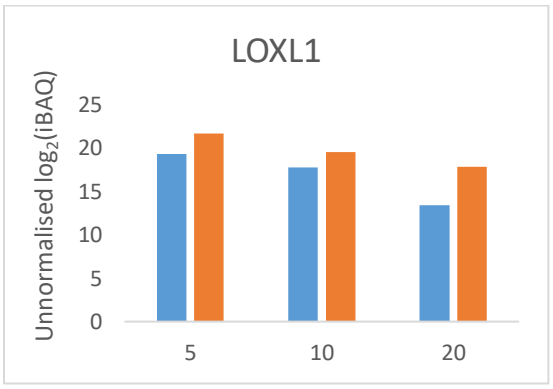
Unlike the genetic mouse models where the fibrotic process takes place over several months, in cell culture experiments time is much more limited, as very long incubation times raise the risk of infection and can trigger apoptosis. On the other hand, sufficient exposure time to a pro-fibrotic agent was needed to allow any overexpression of cross-linking enzymes, their

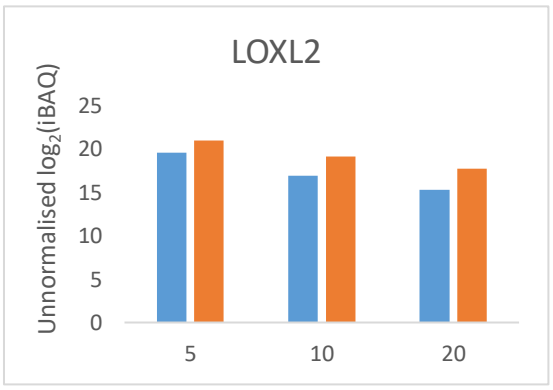
translocation to the ECM, and the resulting accumulation of cross-links. In other words, we might have been unable to find any differences in the cross-linking pattern, if we did not allow sufficient incubation times with TGFβ-1. Hence, the first aim in this study was to investigate how much time was required to detect the hypothesised upregulation of cross-linking enzymes in the ECM in response to sustained TGFβ-1 treatment.

#### *4.2.3.1 Pilot time course*

To address this question, we conducted a pilot time-course proteomics experiment to identify which cross-linking enzymes were present in the cell culture ECM preparations, and to monitor any changes in the level of these ECM enzymes in response to sustained exposure to exogenous TGFβ-1 for 15 days. Inspection of the cell culture ECM protein identifications revealed that only one transglutaminase, TG2, was identified, while two lysyl oxidases, LOXL1 and 2, and two lysyl hydroxylases PLOD1 and 2, were identified and quantified. Figures 4.15 and 4.16 show the unnormalised and normalised levels of these enzymes after 5, 10 and 20 days of incubation with or without addition of TGFβ-1. The cultured dermal fibroblasts remained 100% confluent for both conditions during the 20-day assay period. Interestingly, while the absolute iBAQ levels of each protein were consistently higher in the presence of exogenous TGFβ-1, even at the earliest time point (Figure 4.15), normalisation of these data (Figure 4.16) indicated that enrichment of all five enzymes occurred only after 20 days of incubation, albeit for n=1 data.


TGM2
Unnormalised log2(iBAQ)
30
25
20
15
10
5
0
5
10
20


LOXL1
Unnormalised log2(iBAQ)
25
20
15
10
5
0
5
10
20


LOXL2
Unnormalised log2(iBAQ)
25
20
15
10
5
0
5
10
20

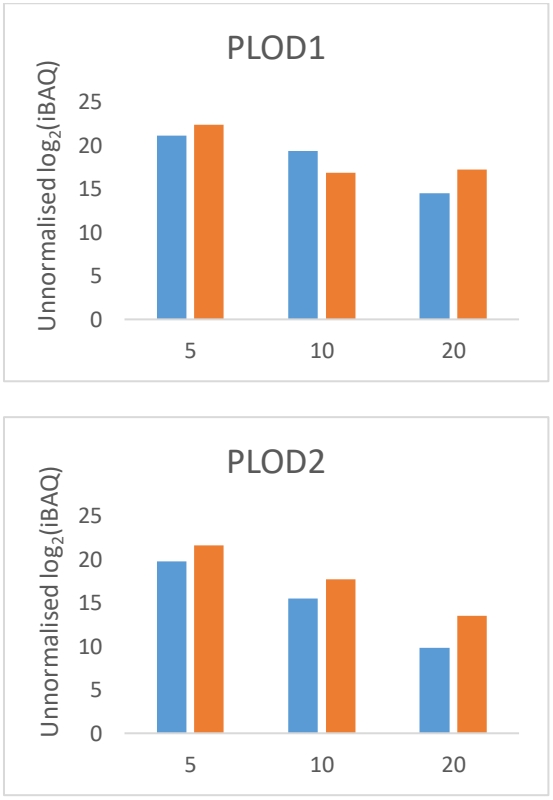
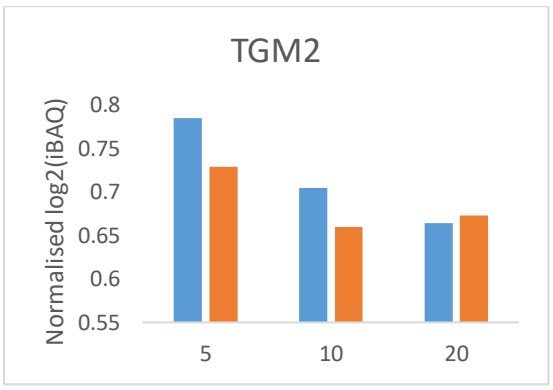
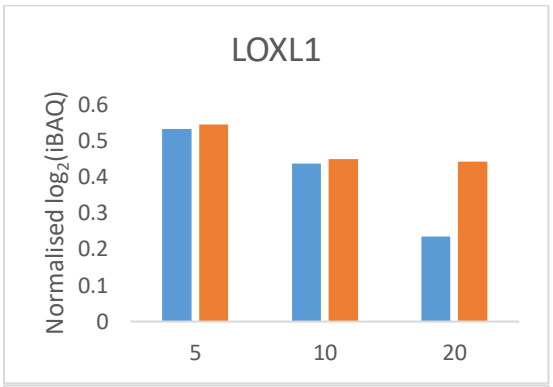
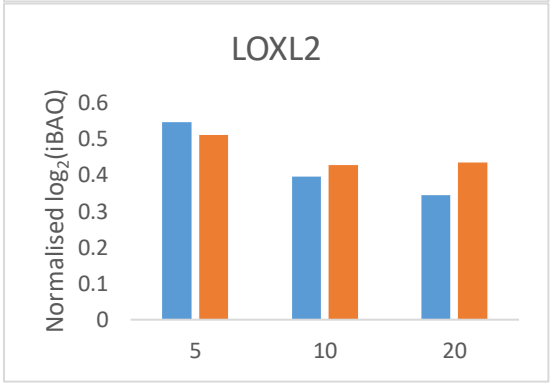



**Figure 4. 15 Time course of unnormalised cross-linking enzyme levels in the ECM during sustained stimulation of dermal fibroblasts with TGFβ-1.** The selected protein levels were obtained from a pilot time-course experiment that analysed the ECM proteome during sustained TGFβ-1 stimulation of dermal fibroblasts (n=1). The blue and the orange bars represent the protein amounts in the control and in the TGFβ-treated samples respectively.


TGM2
Normalised log2(iBAQ)
0.8
0.75
0.7
0.65
0.6
0.55
5
10
20


LOXL1
Normalised log2(iBAQ)
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
20


LOXL2
Normalised log2(iBAQ)
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
20

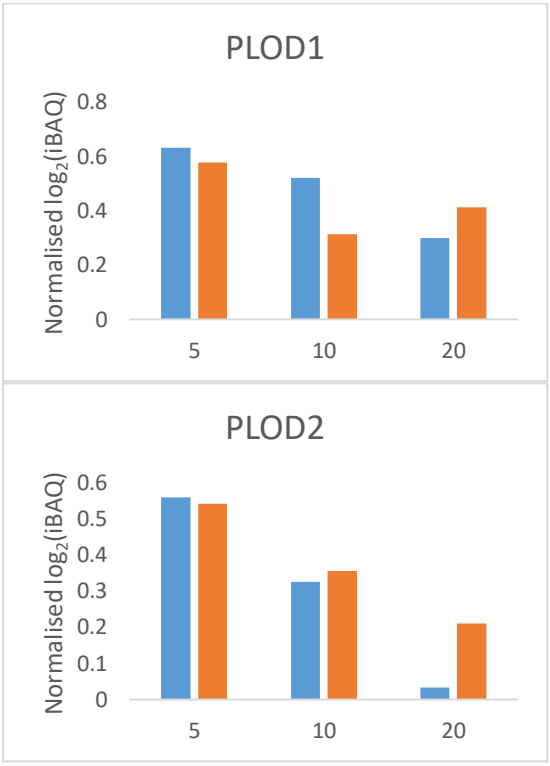


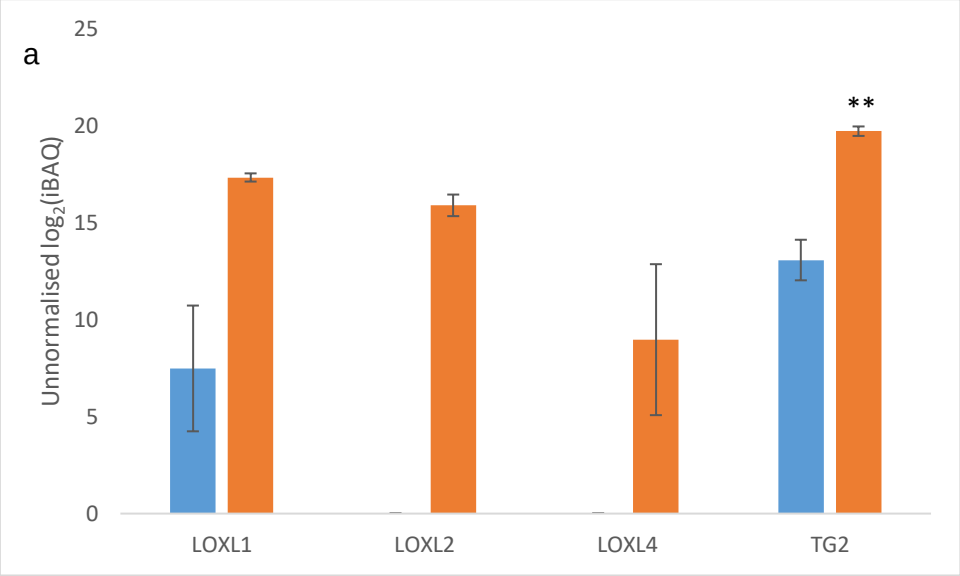

**Figure 4. 16 Normalized protein levels obtained from the time course proteomics experiment with the TGFβ-1 stimulation of dermal fibroblasts.** (n=1) The blue and the orange bars represent the protein amounts in the control and in the TGFβ treated samples respectively.

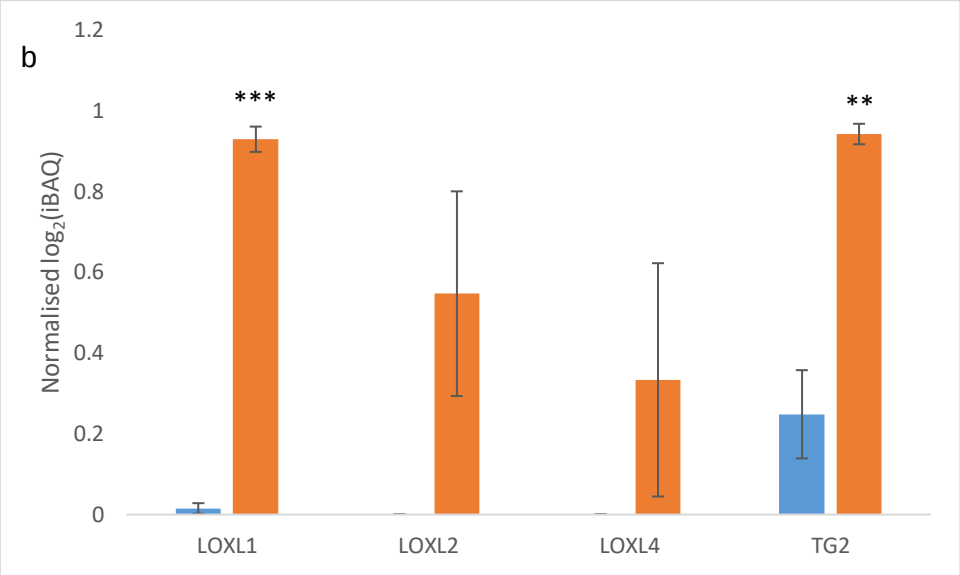
#### *4.2.3.2 TGFβ treatment for 20 days*

In order to assess the statistical significance of the apparent upregulation of five crosslinking enzymes at 20 days in this model, we prepared three biological replicates of samples that had been treated with and without TGFβ-1 for 20 days. Figure 4.17 summarizes the outcome of these experiments. The cross-linking enzymes TG2, and LOXL1, 2 and 4 were identified, but the previously identified PLOD enzymes (Section 4.2.3.1) disappeared from the list of protein IDs. This could have been caused by variations in sample preparation that experience had shown to have a large impact on the number of identified proteins.

Similar to the 20-day pilot data, there was significant upregulation and enrichment of TG2 as a result of sustained TGFβ-1 stimulation (Fig 4.17a,

b). TG2 upregulation by TGFβ-1 was also demonstrated by the higher number of tryptic peptides identified for this enzyme under pro-fibrotic compared to control conditions (Fig 4.17c). Another enzyme that appeared to be statistically upregulated was LOXL1. The absence of the enzyme from one of the control replicates led to a very high variation in this dataset, which explains why statistical significance was not observed for the unnormalised iBAQ values. This effect was diminished after normalisation due to the overall low abundance of the enzyme compared to the other ECM proteins. The other two cross-linking enzymes from the LOXL family (LOXL2 and 4) also showed an increase in level as they were only found in the TGFβ-1-treated replicates (Figure 4.17c). Statistical analysis of these proteins was not possible. The absence of LOXL4 from one TGFβ-1-treated sample (TGF1 in Figure 4.17c) was consistent with its absence from the pilot data.




c

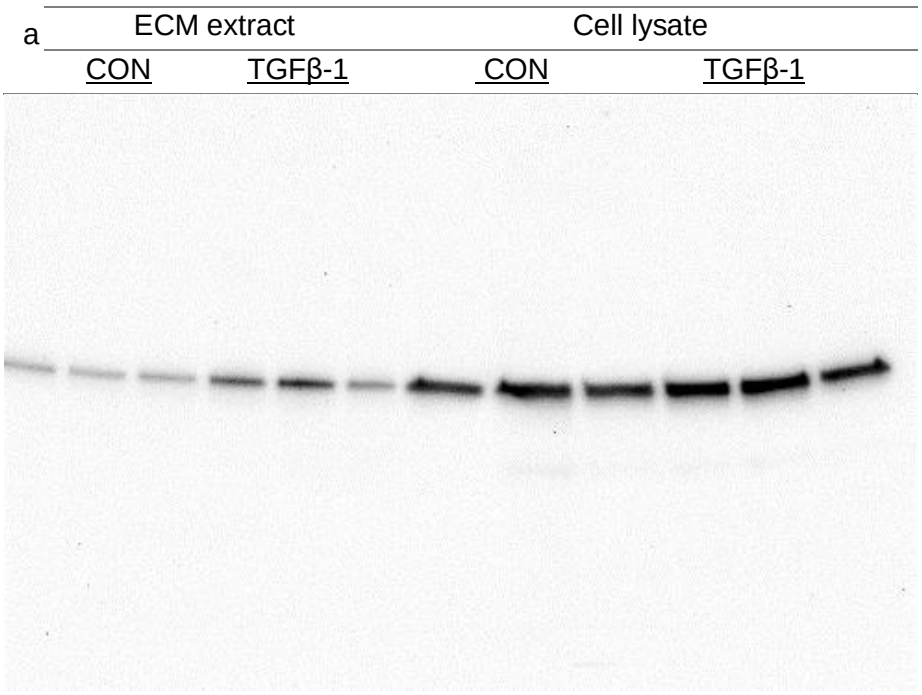
| *Gene name* | CON1 | CON2 | CON3 | TGF1 | TGF2 | TGF3 |
|---|---|---|---|---|---|---|
| *TGM2* | 3 | 5 | 3 | 16 | 15 | 17 |
| *LOXL1* | 1 | 1 | 0 | 4 | 6 | 5 |
| *LOXL2* | 0 | 0 | 0 | 4 | 9 | 9 |
| *LOXL4* | 0 | 0 | 0 | 0 | 2 | 4 |

**Figure 4. 17 Proteomics results for the cross-linking proteins identified after 20 days of TGFβ-1 stimulation.** A one-tailed, two-sample t-test was used to analyse the statistical significances in protein abundance between the control and the TGFβ-1 treated cells (n=3). The blue and the orange bars represent protein amounts in the control and in the TGFβ treated samples, respectively. a: unnormalised iBAQ values (TG2: p=0.003); b: normalised TG2 iBAQ values (LOXL1: $p=8.64x10^{-5}$; TG2: p=0.003); c: Number of peptides found for each cross-linking enzyme in the replicate control (CON1-3) and TGFβ-1-treated (TGF1-3) ECM preparations.

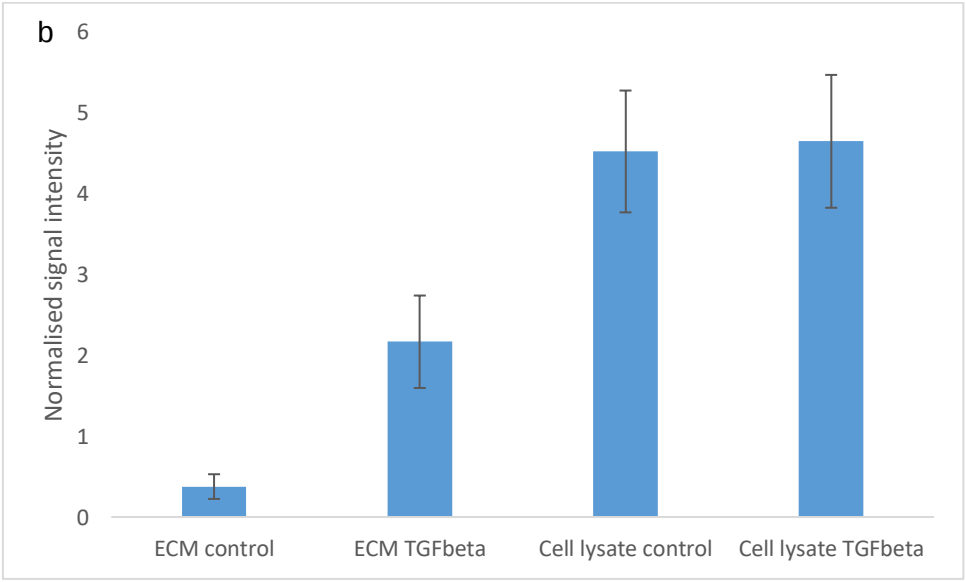
Due to the small number of replicates, there was no evidence that the data followed a normal distribution. Nevertheless, we chose a t-test for our analysis, since it is more sensitive compared with non-parametric tests. We tried to confirm the resulting significance of the changes in LOXL1 (normalised data) and TG2 (normalised and unnormalised data) expression with a non-parametric Wilcoxon test. This was chosen based on the assumption that both datasets contained identical subjects, as all replicates originated from the same donor. Hence, the compared samples were related. We did not find any significant TG2 or LOXL1 upregulation using this test (for both variables, $p = 0.109$).

Similar to our analysis of the mouse models, the proteomics results from the cell culture model were tested orthogonally using Western blotting. The blot shown in Figure 4.18a appeared to confirm the upregulation of TG2 in the ECM preparation, but this marginally failed to reach statistical significance ($p=0.06$) after normalising to the GAPDH levels in the cell lysate. The TG2 band in the set of ECM preparations appeared to be perturbed by a relatively weak signal produced by the third TGFβ-1 replicate (lane 6) compared to the other replicates (lanes 4 and 5), which was not apparent in the cell lysate GAPDH level. No change was detected in cell lysate TG2 ($p=0.5$) or lysyl hydroxylase 2 (PLOD2) levels (Figure 4.18b, $p = 0.13$). In contrast, significant upregulation of PLOD2 was present in the ECM extract ($p = 0.04$). The presence of additional bands was probably due to nonspecific binding of the primary antibody.

a


ECM extract
Cell lysate
CON
TGFβ-1
CON
TGFβ-1

b


Normalised signal intensity
0
1
2
3
4
5
6
ECM control
ECM TGFbeta
Cell lysate control
Cell lysate TGFbeta

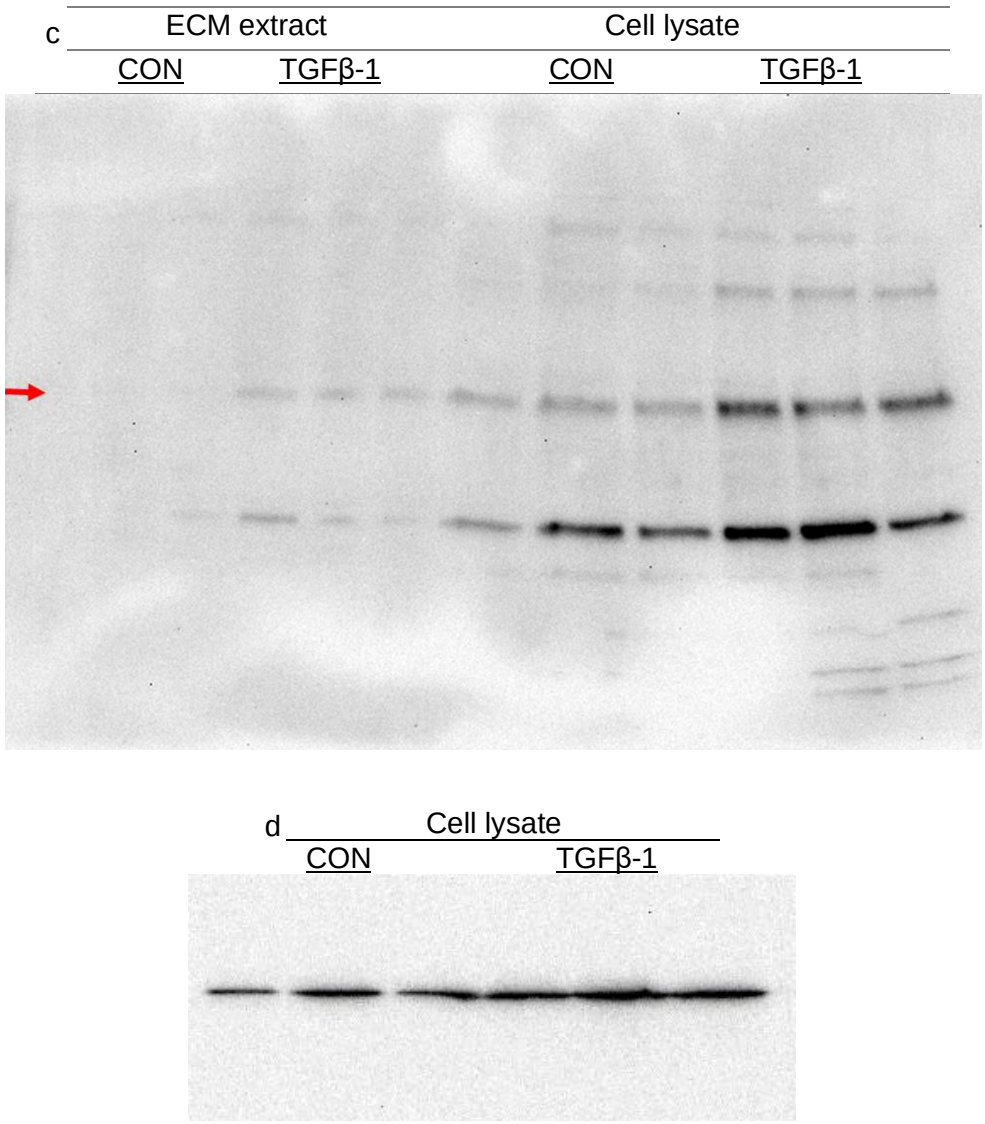
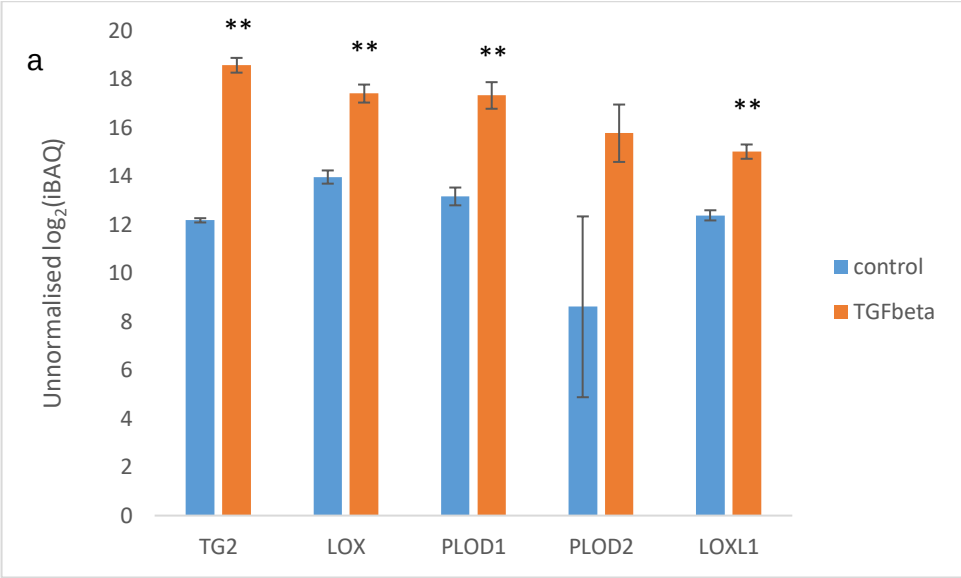


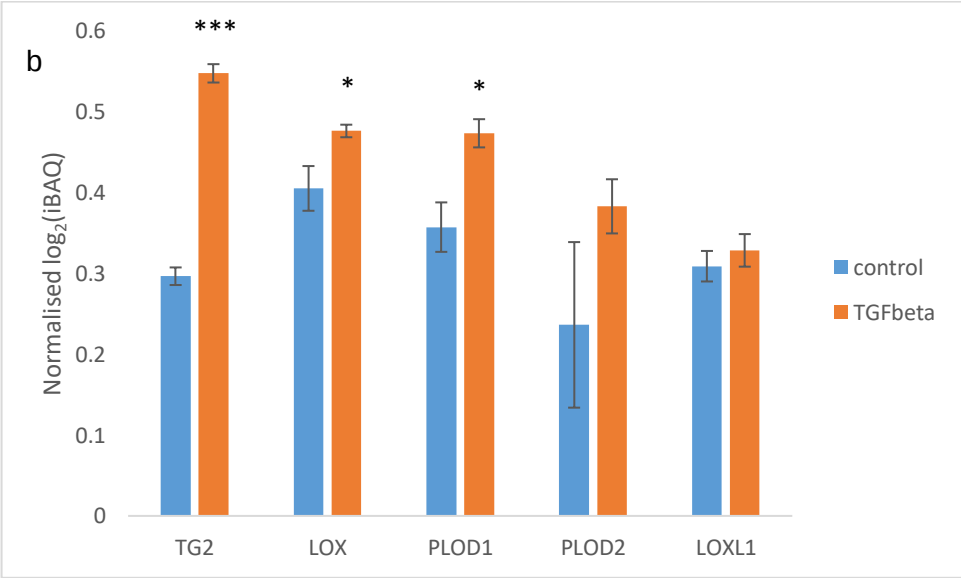

**Figure 4. 18 PLOD2 Western blot results from the 20-day stimulation of cultured dermal fibroblasts with TGFβ-1.** ECM preparations from control or TGFβ-1-treated cultures, and cell lysate preparations from control and TGFβ-1-treated cultures were immunoblotted for TG2 (panel a); PLOD2 (panel c), where the bands matching the expected PLOD2 molecular weight (85kDa) are indicated with an arrow; and GAPDH (panel d); (n=3). A bar chart showing the results of the normalised TG2 signal intensities for each condition is displayed in panel b. The higher variation of the dataset obtained from the TGFβ-treated ECM extracts in comparison with the control was caused by the low value of one of the data points.

#### *4.2.3.3 TGF treatment for 30 days*

In addition to the 20-day incubation, a further experiment with an extended treatment to 30 days was performed, to assess whether exposure times longer than 20 days would produce a larger upregulation of cross-linking enzymes. However, after 30 days the cell number had visibly dropped from

100% to approximately 70% confluence. This was not a result of the higher level of TGFβ-1 signalling as there was no apparent difference in confluence between the control and TGFβ-1 replicates. Figures 4.19 and 20 summarise the proteomics and Western blot results, respectively, of this experiment. Regarding the TG2 levels, an increase ($p = 0.002$ (unnormalised); $p = 7.97 \times 10^{-5}$ (normalised)) was found in the ECM fraction, similar to the 20-day experiment (Figures 4.16 and 4.17). The 30-day data also showed a significant ECM enrichment of LOX ($p = 0.002$ (unnormalised); $p = 0.05$ (normalised)), LOXL1 ($p = 0.002$ (unnormalised); $p = 0.3$ (normalised)) and PLOD1 ($p = 0.006$ (unnormalised); $p = 0.02$ (normalised)) cross-linking enzymes, again suggesting the accumulation of allysine- and hydroxyallysine-based cross-links in parallel with transglutaminase cross-links. Significant upregulation of PLOD2 was not detected, which might be due to the absence of the enzyme from one of the control replicates causing increased variance. Statistical comparison of LOXL2 and LOXL4 levels were not possible, since the corresponding iBAQ values were missing in the control group. The TGFβ-1-dependent increase in the level of TG2 found in the proteomics data was supported by Western blot analysis, which showed that TG2 was detected only in the fibrotic samples (Figure 4.20a). This could not be corroborated with a statistical test as the control levels were 0. No difference between the levels of TG2 in TGFβ-1-treated and control cell lysates was detected after normalization to GAPDH levels. Although the PLOD2 iBAQ values in the proteomic data (Figure 4.19) were consistent with an increase in this enzyme in response to 30 days of TGFβ treatment, we did not find a statistically significant increase in this enzyme in the Western blot analysis after GAPDH normalization ($p = 0.29$).



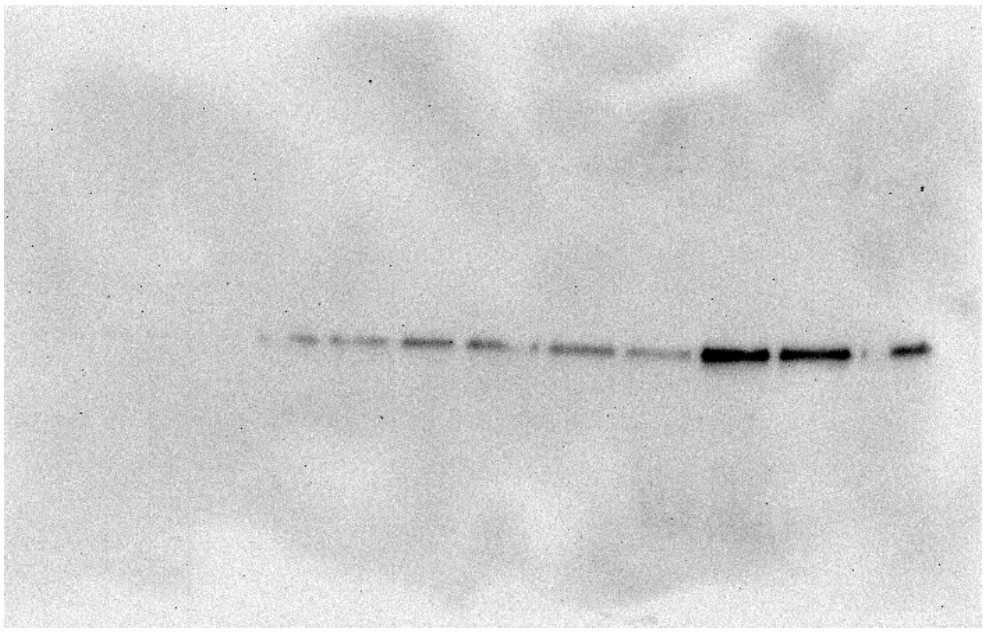

c

| ***Gene name*** | **CON1** | **CON2** | **CON3** | **TGF1** | **TGF2** | **TGF3** |
|---|---|---|---|---|---|---|
| *TGM2* | 2 | 1 | 3 | 5 | 8 | 11 |
| *LOX* | 3 | 4 | 1 | 4 | 6 | 7 |
| *LOXL1* | 1 | 1 | 1 | 2 | 2 | 2 |
| *LOXL2* | 0 | 1 | 0 | 1 | 1 | 3 |
| *LOXL4* | 0 | 1 | 0 | 1 | 2 | 5 |
| *PLOD1* | 4 | 4 | 2 | 4 | 10 | 11 |
| *PLOD2* | 1 | 3 | 0 | 1 | 7 | 9 |

**Figure 4. 19 Proteomics results for cross-linking enzymes after 30 days of TGFβ-1 stimulation.** a. Unnormalised values; b. Normalised values; c. Number of peptides found for each cross-linking enzyme in the replicate control (CON1-3) and TGFβ-1-treated (TGF1-3) ECM preparations.

a

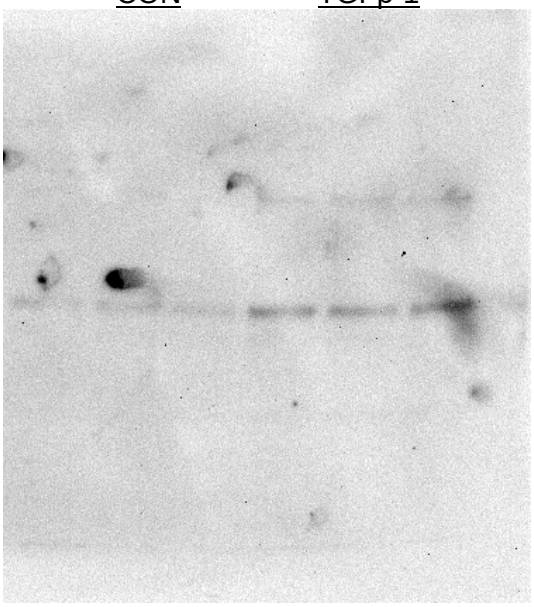
| ECM extract | | Cell lysate | |
|---|---|---|---|
| CON | TGFβ-1 | CON | TGFβ-1 |



b

Cell lysate

CON TGFβ-1



c

Cell lysate

CON TGFβ-1

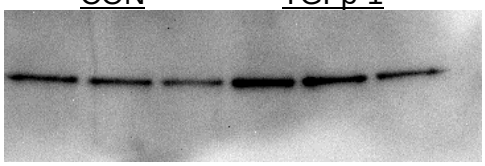


**Figure 4. 20 Western blot results from the 30-day stimulation of dermal fibroblasts with TGFβ-1.** (n=3) a, TG2 detection; b, PLOD2 detection (no signal was detected in the ECM extracts); c, GAPDH detection

#### *4.2.3.4 Effect of sustained TGF treatment on TG2 enzyme activity*

Despite consistent results showing enrichment of TG2 in the cell culture ECM, this was not direct evidence for an increase in the cross-linking activity of this enzyme. We therefore assayed transglutaminase activity using a protocol similar to that published by Wang et al [155]. In this assay, we measured the incorporation of the TG2 substrate biotin-cadaverine into cell-culture ECM after 15 days of culture with or without TGF-β1. This incubation time was selected as it was the longest time point before we started observing cell death. Detection of the product was based on attachment of streptavidin-horseradish peroxidase (HRP). As we were only interested in the TG2 activity within the ECM, we measured HRP levels after decellularization and careful removal of the cell lysate in order to minimise disturbance of the ECM on the culture surface. We quantified HRP levels by measuring the chemiluminescence emitted by a luminol substrate, which is a more sensitive method than measuring the absorbance using the widely employed reagents. This was necessary to detect signals that proved to be too weak to be detected using absorbance. Alongside a comparison of control and TGFβ-1-treated cultures, we included two additional controls for TG2 specificity. One (blank control) lacked the TG2 substrate, and the other (negative control) included EDTA in order to inhibit TG2 activity. The aim of this assay was to subtract nonspecific signals coming from background binding of the substrate to the ECM, or from any HRP binding that did not result from TG2 activity, from the specific signal arising from bona fide TG2 activity in the ECM. The obtained results showed a high variation of the signal intensities within the replicates of the same condition (Figure 4.21). This effect could have been caused by the presence of contaminants in some of the wells that lead to side reactions with the luminol reagent. The high variation remained clearly evident in the blank controls, which contained no

biotin, and yet the chemiluminescence intensity varied from 0 to 174 relative luminescent units (RLU). After comparing the mean signal intensity produced by control ECM with the negative control via a 2-tailed, unpaired, heteroscedastic t-test, we concluded that there was no significant difference between both groups (p = 0.09). This suggests that any TG2 activity in control cultures was too low to be detected and quantified. Nonetheless, despite the high level of replicate variation, after excluding the outliers from the datasets we discovered more than a ten-fold higher signal in the TGFβ-1-treated cell culture ECM compared to the control cell culture ECM (p = 0.001).

a

| ***Replicate*** | **blank control** | **negative control** | **control** | **TGFβ** |
|---|---|---|---|---|
| *1* | 62 | 65 | 26 | 104 |
| *2* | 34 | 54 | 17 | 135 |
| *3* | 373 | 87 | 3 | 310 |
| *4* | 174 | 12 | 25 | 224 |
| *5* | 0 | 0 | 15 | 920 |
| *6* | 0 | 66 | 246 | 228 |
| *Mean* | 107 | 47 | 55 | 320 |
| *Mean ex OL* | 54 | 39 | 17 | 200 |

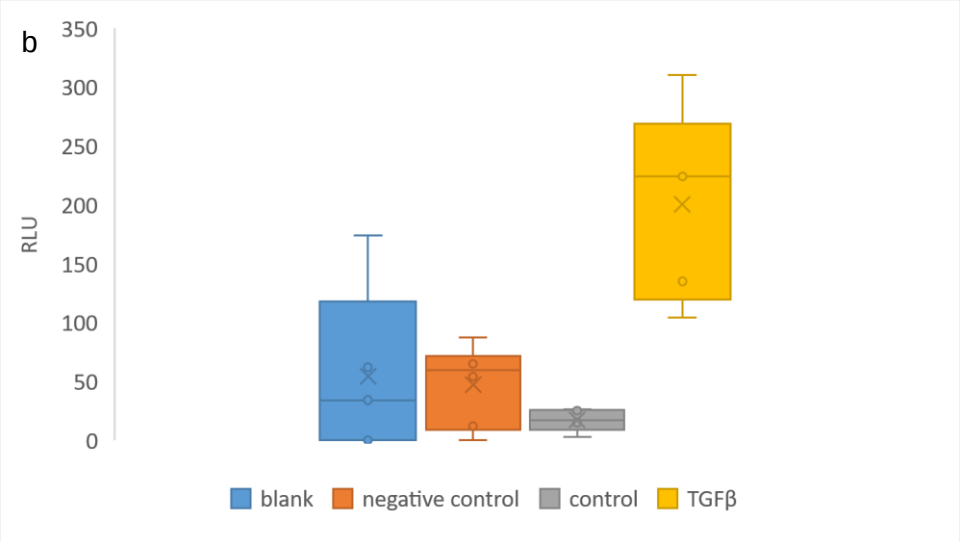



**Figure 4. 21 TG2 activity of decellularised control and TGFβ-1-treated cell cultures.** Data are presented in relative luminescent units (RLU). blank control: decellularised cell cultures incubated without biotin-cadaverine; negative control: decellularised cell cultures incubated with 100 µM biotin-cadaverine + 10mM EDTA; control and TGFβ-1: decellularised cell cultures incubated with 100 µM biotin-cadaverine. A table of all data values is given in panel a (*ex OL*, excluding outliers; outliers are shown in red text), while panel b depicts the datasets (excluding the outliers) in a box and whisker plot.

In addition to the activity assay, we performed a Western blot to check whether ECM TG2 protein levels were also increased at 15 days. The results showed that there was a clear increase in TG2 levels in the lysates prepared from TGFβ-1-treated cells ($p = 0.02$) (Figure 4.22) as well as in the ECM extracts ($p=0.003$). Therefore, we conclude that the increased cross-linking activity in the TGFβ-1 treated cell cultures was confirmed with the TG2 accumulation in the ECM. It is also possible that the increased activity also originates from other forms of transglutaminases that play an essential role in fibrotic progression, but have not yet been characterised. mRNA analysis of various transglutaminase isoforms (TG1,2,3,4 and 5) in skin biopsies from scleroderma patients and matching controls revealed a significant increase only in the TG2 concentration (data not shown). In addition, there is no current literature about the role of the other transglutaminase isoforms in any fibrotic malignancy. Another type of transglutaminase that had previously been found to have a cross-linking activity in the ECM is Factor XIIIa [156]. However, there is no current evidence about changes in the activity of this enzyme in relation to fibrosis. Increased TG2 cross-linking activity in the TGFβ-1 treated cells might also originate from other factors in the in the ECM environment.

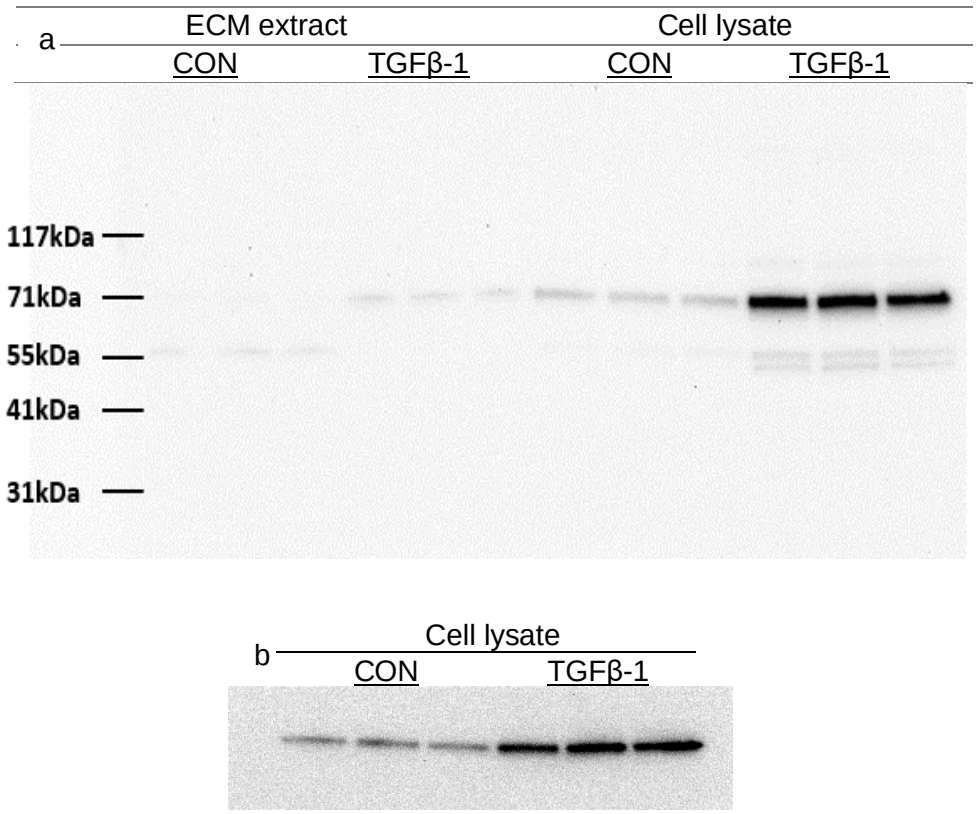



**Figure 4. 22 TG2 Western blot of dermal fibroblast ECM and lysate preparations after 15 days of stimulation with TGFβ-1.** ECM and cell lysates prepared from control and TGFβ-1-treated cultures (n=3) were immunoblotted with antibodies to TG2 (a) and GAPDH (b).

## 4.3 Discussion

### 4.3.1 Isolation of ECM proteins from human dermal fibroblasts

#### *4.3.1.1 Comparison of trypsin and detergent methods*

The suitability of the methods that we developed for the isolation of ECM from dermal fibroblast cell cultures was assessed according to different criteria that were required to achieve the objectives of this project. One of the main goals during method development was to achieve the highest possible

sequence coverage for the proteins of interest. This was an important general criterion, as higher sequence coverage for any given protein was thought to allow a more complete mapping of its cross-linking positions. Another criterion that was considered to be important in the choice of a method was the reproducibility of protein quantitation between biological replicates. In other words, a method providing minimal variation in the iBAQ values for a given protein across biological replicates was essential for the detection of a maximum number of significant differences between the cross-linking patterns of the control and the fibrotic cellular states. A third criterion that was beneficial for the aims of this project was to identify the highest possible number of ECM proteins. In this study, proteomic data analysis was focused on the extracted levels of known substrates of cross-linking enzymes, such as specific collagen types and elastin. Alongside collagen I and III, which are the most abundant collagen types in skin [40], other collagen types that had previously been reported to be produced by dermal fibroblast cell cultures, such as type V [157] and type VII [158], were expected to be identified. Moreover, a sample preparation technique that provided a higher total number of identified ECM proteins was expected to be beneficial for the discovery of new cross-linking targets that might also play a crucial role in skin fibrosis during SSc progression.

Two different methods for the isolation of ECM proteins from dermal fibroblasts were performed and their suitability for the goals of this project was assessed. In the first method, cells were removed from the cell culture flask via tryptic digestion, which allowed their detachment from the ECM. During this procedure, tryptic digestion of ECM proteins was minimised by incubating at room temperature and by using the minimal incubation time that we found to be necessary for full removal of cells (3 min). The advantage of this procedure was that cells could be removed intact, thereby preventing the

release of cellular components and resulting in higher purity of the ECM extracts. Although the presence of cellular proteins in the ECM extract is unavoidable, minimising their amount is particularly important for the identification of a high number of ECM proteins and for the achievement of higher sequence coverages. This can be explained by the decrease in the sample complexity that enables a higher fractional presence of ECM-derived peptides that can be crucial for their MS detection. After the removal of fibroblasts, ECM proteins that remained on the surface of the culture flask were dissolved in 8M urea. Urea is known for its chaotropic properties and is able to break hydrogen bonds and disrupt hydrophobic interactions both between and within proteins. Thus, the use of 8M urea increased the solubility of ECM proteins, including highly insoluble collagen fibres, whose triple-helical structures are held together by numerous hydrogen bonds. We therefore hypothesized, that this strategy would be appropriate for the extraction of the insoluble ECM. Subsequent reduction of the sample volume via centrifugation through a 3kDa filter was performed to increase the protein concentration, which was necessary for the efficiency of the subsequent treatments with enzymes and other reagents. Sample dilution to reduce the urea concentration to 2M was then necessary for the activity of the enzymes used to prepare samples for mass spectrometry.

The second method that was used to isolate ECM from dermal fibroblasts was a modification of the method developed by our group to isolate ECM proteins from tissue. As detachment of cells from tissue had been found not to be optimal using extensive homogenisation and consequent cell lysis, or using non-specific protease treatment and consequent loss of ECM proteins, the preferred method employed mild solubilisation using the ionic detergent sodium deoxycholate and the non-ionic detergent Tween-20, in which ECM components remained insoluble. To obtain the cell culture ECM, cell lysis

was followed by mechanical removal of the combined cellular and ECM components from the cell culture flask (scraping). Isolation of ECM proteins was achieved by centrifugation to pellet insoluble proteins (including ECM proteins), while the majority of cytoplasmic proteins remained in solution.

By comparing the results obtained using each method we deduced that the trypsin method provided a larger number of ECM protein identifications (84 proteins) compared to the detergent method (63 proteins). The significantly lower number of ECM proteins identified using the detergent method compared to the trypsin method was likely to have been at least partially due to the higher number of cellular proteins that were released during detergent solubilisation and co-precipitated with the ECM proteins during phase partitioning (648 compared to 363 proteins for the trypsin method). As a result of the lower purity of the ECM extract, the total load of the ECM on the LC-MS was reduced and this was likely to have impeded the detection of low-abundance ECM peptides. Another possible cause of the difference in the number of protein identifications was the use of 8M urea after cell removal with trypsin. Solubilisation of ECM proteins in urea was likely to have been more efficient at extracting ECM proteins from the cultures flask than scraping and pelleting. Finally, while the used detergents had previously been reported to effectively decellularise tissue [159], it may have also solubilised a subset of ECM components.

As already mentioned, among the proteins identified using the trypsin method were ECM proteins that were potential targets for cross-linking analysis. However, further analysis of the reproducibility of these proteins between the biological replicates revealed that collagen type V, which had previously been reported to be expressed by dermal fibroblasts [157], was identified in only one of the biological replicates. The same irreproducibility was also observed for collagen type VIII, for which no previous information

about its presence in the ECM of dermal fibroblasts could be found. As the number of identified (unique) peptides in the second replicate were 2 and 3 for collagen V and VIII respectively, it is more likely that the result was due to the limited reproducibility of the technique rather than false-positive identifications. Therefore, although the trypsin method gave a larger number of ECM proteins, including elastin that was not identified using the detergent method, it seemed to be unsuitable for quantitative comparisons of different cellular states due to its poor reproducibility.

Despite the lower number of ECM protein identifications obtained using the detergent method, a significant number of proteins identified using this technique were not found in the ECM extracts prepared using the trypsin method. Interestingly, most of these unique proteins appeared to be low in abundance. This could have been partially due to the presence of false-positive identifications, since some of the proteins were identified based on only one peptide (data not shown). Finally, the results of the trypsin method may have been compromised by the presence of urea during protein digestion; however, this is rather unlikely, as there was no evidence for decreased activity of the selected enzymes in 2M urea.

Although the reproducibility of ECM protein identifications in biological replicates exceeded 80% for both methods, quantitation (iBAQ values) was more reproducible using the detergent method, indicating that this technique would be able to provide more reliable information when applied to cross-linking analysis. A possible explanation for the higher variation of the biological replicate iBAQ values obtained using the trypsin method could be partial digestion of the ECM by trypsin during cell removal or partial deglycosylation by deglycosidases in 2M urea. These factors could also cause the slightly lower iBAQ values obtained using the trypsin method for the abundant collagen proteins.

In summary, considering the number and reproductivity of identifications as well as the reproducibility of quantification by each method, we concluded that the detergent method was more suitable for the aims of this project. Although the trypsin method identified a higher number of ECM proteins it was less useful for ECM cross-linking analysis in cell culture and skin samples.

#### *4.3.1.2 Development of the detergent method*

In order to improve protein sequence coverage, the detergent method was further developed by inhibiting unwanted protein digestion during cell lysis. To achieve this a protease inhibitor cocktail and 20mM EDTA were added to the detergent buffer to inhibit the activity of matrix metalloproteinases (MMPs) and serine proteinases, such as cathepsins, that show collagenolytic and elastinolytic activity [160]. The activity of $Zn^{2+}$-dependant MMPs was inhibited by EDTA, while the activity of serine proteases was inhibited by two cocktail components (AEBSF and aprotinin).

In addition to the improvement of the ECM extraction yield, it was necessary to develop a method to remove the detergent used during decellularisation, as we found that the MS signal suppression by residual detergent had a negative effect on the proteomics results (data not shown). We hypothesized that a separation of the tryptic digest from the detergent would be achievable via SCX chromatography. The separation principle relied on the difference in charge between peptides and detergent in the mildly acidic buffer (pH = 2). The uncharged and protonated detergents ($pK_a$ of cholate = 6.58) were not expected to interact with the strongly anionic ($-CH_2-SO_3^-$) groups of the SCX stationary phase ($pK_a \approx -2$) and were therefore expected to elute in the flow through fraction. In contrast, the positively-charged peptides ($pK_a$ of Asp and

Glu = 3.65 and 4.35, respectively) were expected to bind to the stationary phase and be eluted by the salt gradient. In addition, ion-exchange chromatography is commonly used for the separation of linear peptides from cross-linked peptides. This strategy relied on the difference in their charge states. At acidic conditions (pH = 2) most non-cross-linked tryptic peptides have a charge state of +2, whereas cross-linked peptides carry higher charges [161]. This implies, that cross-linked peptides elute at higher salt concentrations, which would allow their isolation form the rest of the digest. This separation step was particularly important, as cross-linked peptides can be very low in abundance, hindering their detection [77]. In support of this approach, SCX is commonly used for the analysis of cross-linked peptides from protein digests in chemical cross-linking experiments [162]. Therefore, at this stage of the method development, enrichment of cross-linked peptides was another reason for using SCX prior to LC-MS analysis.

Fractions collected during the salt concentration gradient were expected to contain tryptic peptides from the ECM digestion and no detergent. However, the results from this experiment showed only a partial separation of the detergent from the peptides. The resulting chromatograms did not show insufficient separation of the species eluting before and after the gradient was started, which suggested that this was not due to insufficient peak resolution. A possible cause for the incomplete removal of detergent is a hydrophobic interaction between detergent and lipophilic regions of the proteins that could have taken place after protein denaturation by 8M urea. These interactions may not only have reduced the efficiency of detergent removal, but they may also have inhibited interactions with the SCX stationary phase and with the C18 stationary phase during stage-tip purification resulting in sample loss. This is also a possible explanation for the similarly poor results obtained from an experiment in which a sample

prepared using detergent was combined with a second sample prepared using 8M urea, followed by the rest of the detergent-based preparation procedure. Although it offered no enrichment of cross-linked peptides, a more efficient and much less laborious technique for detergent removal from ECM extracts proved to be precipitation with 80% acetone prior to enzyme treatment and this was used for subsequent sample preparation.

In conclusion, only some of the planned changes were used for subsequent sample preparation, as only this subset led to an improvement of the data. These changes were (i) culturing a larger number of cells for ECM extraction, (ii) inclusion of protease inhibitors during decellularization, and (iii) detergent removal via acetone precipitation.

### 4.3.2 Isolation of ECM proteins from mouse skin samples

As the final goal of this project was to analyse the cross-linking changes in SSc skin samples, another important part was to find an efficient method for the isolation of ECM proteins from skin tissue. The success of the experiment performed for ECM extraction from skin tissue was assessed using the same criteria as used for the experiments with dermal fibroblasts, including the number of identified proteins; more importantly, the number and the amount of identified collagens; and the qualitative and quantitative reproducibility of the results. An advantage of this sample type in comparison with dermal fibroblast cultures was the higher amount of ECM that led to an increased extraction yield, increased load onto the mass spectrometer, and better sequence coverage of the identified ECM proteins. In addition, one would have expected that higher ECM extraction yields would also have improved the level of qualitative reproducibility, particularly of low abundance proteins, since an increased amount of a protein in a sample was expected to raise

the probability that it would be identified. This explains the better qualitative reproducibility of the results obtained from skin tissue in comparison with the results obtained from cell culture ECM extracts. Among the identifications, a significant number of proteins involved in ECM cross-linking, either as substrates or enzymes, were found. Several transglutaminases were successfully detected that might play a role in fibrotic development in the TSK2 mouse model. Most importantly, we reproducibly identified and quantified transglutaminase 2 (TG2), which is the main cross-linking enzyme of interest for this project (Section 3.3).

In conclusion, the methods we developed to isolate and analyse ECM proteins from skin tissue and dermal fibroblast cultures provided useful information about ECM proteomic changes and were used in subsequent experiments to characterize changes of the cross-linking pattern in pro-fibrotic models.

### 4.3.3 Characterization of ECM cross-linking enzymes in different SSc models

#### *4.3.3.1 Mouse models*

In this project, two different mouse models, TSK2 and DNR, were analysed for the expression of ECM cross-linking enzymes. We hypothesized first that these models would have elevated expression of a cross-linking enzyme relative to controls; second, that elevated expression would cause a higher level of cross-links to be formed; and third, that skin tissue from these models could be used to develop a method to analyse altered, site-specific cross-linking of ECM proteins.

Existing evidence in the literature led me to the hypothesis that the TSK2 model would be suitable for this type of study. First, the phenotypic appearance of the TSK2 mouse model showed thicker and less elastic skin [95], and it had previously been established that ECM stiffness directly correlates with the extent of ECM cross-linking [154]. Thus, the hypothesised increase in cross-linking was consistent with the altered physical properties of skin in this model. Second, it had previously been demonstrated, that increased ECM cross-linking also activates pro-fibrotic mechanisms, such as increased TGFβ-1 signalling [55], meaning that diverse types of tissue fibrosis might possess similar increases in these types of modification. In addition, compared to the cultured fibroblast model, using a pro-fibrotic genetic mouse model was expected to have advantages for the comparison of cross-linking between control and fibrotic states, because the much longer exposure to pro-fibrotic pathways in the mouse models was expected to allow greater accumulation of cross-links and increase the likeliness of finding more significant changes. For similar reasons it was also clear that neither genetic model was expected to represent the full change in the cross-linking pattern in SSc, as at the time of diagnosis, as skin fibrosis in patients is prolonged over several years and is therefore much more advanced. This difference is important for the discovery of new targets for effective anti-fibrotic agents, as fibrotic irreversibility is caused by the sufficient accumulation of cross-links over a longer period of time. Due to the limitations in the fibrotic stage achievable with animal models, it would not be clear whether any cross-linking changes identified in these models would be directly relevant to fibrotic irreversibility in SSc. This would require studies of human patient samples. Therefore, although fibrotic mouse samples were more easily accessible and therefore suitable for method development, they could be used to identify only candidate cross-linking changes that might lead to the pathologic condition in man.

The 23-week-old TSK2 skin samples we received showed the expected histological features, including thickening of the dermal layer accompanied by excessive accumulation of fibrillar collagen [91]. As the literature confirmed a significant increase of the hydroxyproline content and the accumulation of fibrillar collagen in TSK2 mice at 10 and 23 weeks [95], we was surprised to find an insignificant increase in these proteins in the proteomic data at both time points (data not shown). One possible explanation for this result is that the increase in our samples was small and required a higher number of biological replicates to establish statistical significance. This suggestion is consistent with the histological assessment of these samples (Section 4.2.2.1).

The detected levels of most cross-linking enzymes in TSK2 relative to control skin were not increased in line with our hypothesis at 10 weeks, and not at all at 23-weeks. More specifically, the proteomic results from the 10-week-old mice showed a modest increase only in TG2, although this could not be validated using Western blotting due to an inadequate level of sensitivity. A possible reason might be insufficient TG2 extraction from the mouse skin ECM pellet that was generally bigger than the pellet obtained from cell cultures. In this regard, more efficient resuspension (e.g. via sonication) in SDS-PAGE buffer prior to Western blot analysis proved to be necessary for the tissue ECM, but not the cell culture ECM, in order to isolate a sufficient amount of TG2 for detection. We expected that the levels of this enzyme would be even higher after 23 weeks, however this was not the case and the levels detected in these samples seemed lower than the control, although the variation in the 23-week data was relatively high.

Finally, it is clearly possible that, apart from the elevation of TG2 at 10 weeks, this model did not reveal any other increase of cross-linking enzymes. Indeed, our Western blot analysis could not detect any upregulation of lysyl

hydroxylase 2 at either time point based on normalised amounts of tissue used for sample preparation (as we couldn't normalise these data to GAPDH).

The DNR mouse model is another genetic model of dermal and pulmonary fibrosis. Fibrosis in this model occurs as a result of activation of TGFβ signalling pathways [98], that potentially also activate other fibrotic mechanisms such as increased ECM cross-linking [73]. our proteomics results suggested that TG2 and TG3 were upregulated, although TG2 was not enriched, in the ECM. The above mentioned hypothesis that the fibrotic stage could be too early in the skin samples to detect a significant upregulation applies to our analysis of this model as well.

As already mentioned, there is no current evidence in the literature about increased ECM cross-linking activity in these mouse models. Although the increase of TG2 found in the TSK2 mouse model was statistically insignificant, it is possible that the difference in the protein levels between the control and the fibrotic condition was too small to be detected with statistical significance using only two biological replicates per condition. Hence, another experiment with a higher number of biological replicates could provide evidence for significant TG2 upregulation.

Higher levels of TG2 in both genetic models provided indirect evidence for the increase of transglutaminase cross-links in the ECM, as TG2 is the most well-known transglutaminase isoform with ECM cross-linking activity [152]. However, protein enrichment is not direct evidence for increased cross-linking, since enzyme activation is another crucial factor. Analysis of TG2 activity in the ECM of skin samples was not possible, as multiple freezing and thawing combined with centrifugation during the isolation of the ECM had probably affected the activity of the enzyme. In contrast with the TSK2

data, the results from the DNR samples were less clear in terms of TG2 upregulation, especially as no enrichment was found after normalisation.

Interestingly, TG3 and TG1 levels were raised in the DNR mouse model. TG1 can be active in the ECM [152] and several ECM proteins were identified as targets for this transglutaminase isoform including fibrinogen β and α and collagen type XII [163]. Thus, it is not excluded that an increased TG1 expression can lead to increased transglutaminase cross-linking of the ECM. Evidence about the presence of TG3 in the ECM has not yet been published.

Considering the PLOD2 Western blot results, it can be concluded that this enzyme is not upregulated in any of the models, thus no enrichment of the corresponding allysine-based and hydroxyallysine-based cross-links can be expected. In conclusion, we could not find sufficient evidence that the studied mouse models are suitable for the analysis of ECM cross-linking changes in skin fibrosis.

#### *4.3.3.2 Cell culture model*

Similar to the analysis of the mouse models, the main focus of this study was to find a pro-fibrotic skin model showing significant TG2 upregulation. It is known that this enzyme is mainly active in the ECM [163]. As already mentioned, there is existing evidence for the upregulation of TG2 in TGFβ-1 stimulated dermal fibroblasts at the transcriptomic and proteomic level, while other cytokines that play a role during inflammatory processes in SSc, such as IL-1β and IL-2, do not trigger TG2 upregulation [73]. These results suggested that the dermal fibroblast model could be suitable for this project, however the published data on increased TG2 expression were insufficient for several reasons. Firstly, the proteomic upregulation had been demonstrated only using whole cell lysates. In this project we were interested

in the proteomic changes in the ECM, which can differ from changes in the rest of the cell. For instance, while cellular TG2 levels are increased after 24h of stimulation with TGFβ-1 [73], the secretion of this protein and its incorporation into the ECM may require more time. Therefore, it was essential for this project to study changes in the level of TG2 in the ECM fraction. Secondly, transglutaminases are multifunctional enzymes involved in several catalytic and non-catalytic processes [61]. Therefore, an increase in TG2 level in the ECM is not direct evidence for an increase in its ECM cross-linking activity. Therefore, we assayed exogenous amine incorporation into the ECM by transglutaminase activity.

The results presented in this study show that significant enrichment of the TG2 protein level in the ECM occurred at and after 20 days' incubation with TGFβ-1. Although the unnormalized TG2 levels increased after 5 and 10 days' incubation, the lack of enrichment before 20 days was most likely due to a similar overall increase in the abundance of other ECM proteins at early time points. The later increase in normalized TG2 values indicated a higher overall rate of synthesis, secretion and incorporation of this enzyme relative to most ECM proteins. Although we could not confirm this proteomics result using Western blotting, the fact that (i) the TG2 levels in the ECM extract were much closer to significance than the TG2 levels in the cell lysate, and (ii) Western blotting of 30-day samples showed increased TG2 in the ECM relative to untreated cells (see below), together indicate that the TG2 change detected in the ECM fraction arose from extracellular TG2 rather than from contamination by the much higher level of TG2 in the cell lysate.

Even though a statistical test for significance could not be performed with the Western blot results after 30 days of TGFβ-1 treatment, these results provided additional evidence for an increase and an enrichment of TG2 in the ECM. The increase in TG2 levels at this time point was accompanied by

an increase in other cross-linking enzymes from the LOXL and PLOD family, suggesting that this model can be also used for analysis of other types of cross-links. These findings are in line with previous studies that demonstrated their upregulation in osteoarthritic human synovial fibroblasts (hSF) as a result of TGFβ-1 treatment [142]. However, their overexpression in dermal fibroblast cultures had not been previously demonstrated. As collagen modification by PLOD2 typically occurs in the endoplasmic reticulum prior to secretion into the extracellular matrix [164], a significant increase in the amount of this enzyme in the cell lysate served as an evidence for an increase of the corresponding type of ECM cross-links. Finally, we also demonstrated that this model can elevate amine incorporation into the ECM using an established assay of the cross-linking activity of different transglutaminases [155].

To summarize the data presented in this chapter, it can be concluded that the sustained treatment of human dermal fibroblasts with TGFβ-1 for 20 days was a good model for this project. This model could not only serve as a tool with which to develop a method for cross-linking analysis in complex biological systems, but could also provide valuable information about changes in ECM cross-linking in response to stimulation by the major pro-fibrotic agent TGFβ-1, that could eventually be tested in human disease tissue.

# Chapter 5

# Positional analysis of transglutaminase cross-links in the ECM of pro-fibrotic models

## 5.1 Introduction

Following the results from the previous chapters, namely the identification of transglutaminase cross-linked sequences using mass spectrometry and the discovery of a pro-fibrotic cell culture model that showed increased transglutaminase cross-linking activity in the ECM, we used the methods we had developed to compare differences in the cross-linking patterns between control and pro-fibrotic ECM. The aim of this chapter was to detect the appearance of new transglutaminase cross-linked sites in the ECM in the early stages of fibrotic development. This was achieved by combining the methods for preparation of ECM digests from cell cultures and mouse skin with the mass spectrometry based identification of transglutaminase cross-linked peptides. As previously stated, TGFβs, and in particular TGFβ-1, are pro-fibrotic cytokines that induce the upregulation of several cross-linking enzymes including transglutaminases in various pro-fibrotic models [73, 143, 165]. Based on this evidence, models showing TGFβ-1 hyperactivity were appropriate for our analysis of ECM cross-linking changes that occur during fibrotic progression. In addition, our study of changes in ECM cross-linking pattern as a result of the action of a single pro-fibrotic agent would help identify the agonist triggering differences in cross-linking. As shown in the previous chapter, our pro-fibrotic cell culture model provided further evidence for increased TG2 activity in the ECM as a result of TGFβ-1 stimulation. Consequently, we expected to identify new transglutaminase cross-linking sites in this model that could serve as potential targets for new anti-fibrotic therapies. In order to find new TG2 substrates relevant to skin fibrosis, we applied the workflow we had developed for ECM proteomics with subsequent

software searches for transglutaminase cross-linked peptides. Searches for transglutaminase cross-linked peptides were also performed using the ECM digests of the 10-weeks-old TSK2 samples. The lack of statistical significance of the increase in TG2 levels in our proteomics results could have been caused by a relatively high level of variation arising from the experimental procedure. Therefore, we could not exclude the possibility of a presence of new cross-linking sites in this model. For the selection of the most reliable cross-link identifications from the MassMatrix searches, the criteria defined in the Chapter 3 – namely the availability of more than one fragmentation spectrum per peptide ID and a score and pp value that are equal or higher than the results for the TXLP - were applied in this chapter. In addition to the scoring criteria, the quality of the MS2 spectra were manually inspected. The selection criteria for a good MS2 spectrum match were adopted from the guidelines to avoid misassignments in cross-linking mass spectrometry proposed by Iacobucci and Sinz [166]. These are that the spectrum must show fragment ions from both peptides in a cross-linked peptide pair and that the majority of the high intensity peaks must be assigned to fragment ions.

## 5.2 Results

### 5.2.1 Identification of transglutaminase cross-linked sites in TGFβ-1 stimulated dermal fibroblast ECM

For the identification of transglutaminase cross-linked sites in cell culture ECM, digests of the ECM fraction prepared from dermal fibroblast cultures treated with TGFβ-1 for 30 days were compared with the corresponding controls with no added agonist. The longest treatment time point was chosen

for this comparative study between the control and pro-fibrotic condition, as these samples were most likely to show significant differences in their cross-linking pattern. Moreover, as shown in the previous chapter, a definite increase in TG2 level was demonstrated only after 30 days of TGFβ-1 stimulation. Any cross-links that could be found in at least two of the three replicates prepared for each condition would be considered to be significant hits.

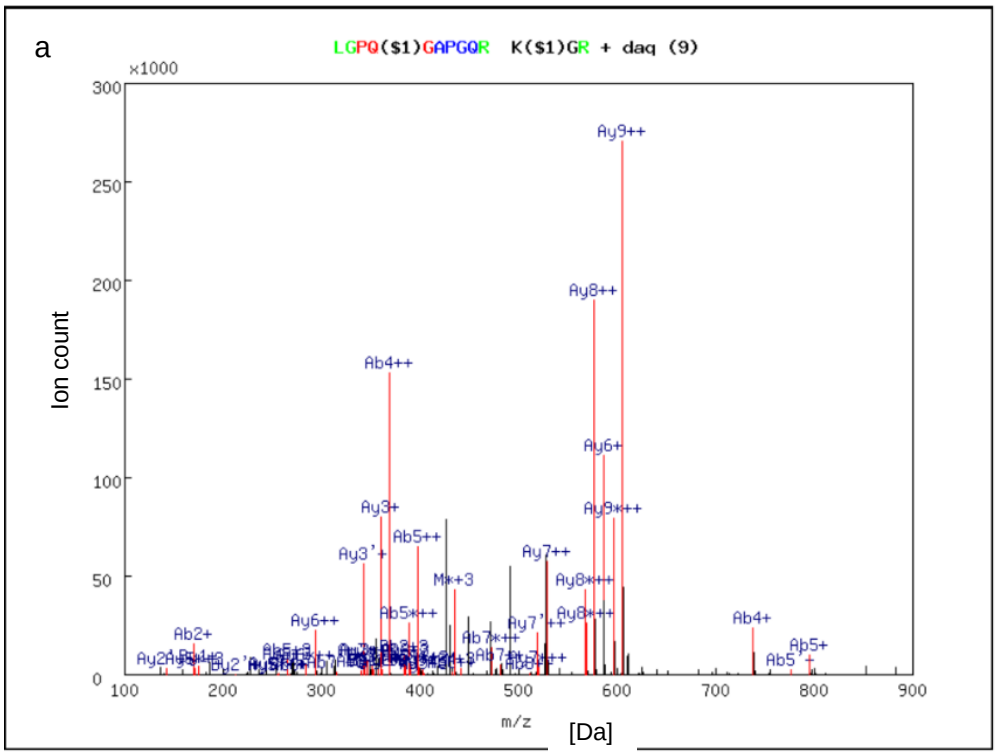
The MS2 spectra of the significant transglutaminase cross-linked peptide hits identified using MassMatrix are depicted in Figure 5.1 and the corresponding scoring details are given in Table 5.1. The positions of the cross-links in the protein sequences are shown in Figure 5.2. In general, because two peptides are involved in a cross-link, the search space increases with the square of the number of candidate sequences in the database [167]. Thus, the search of cross-links in large data sets such as the ECM proteome or the entire human proteome is particularly challenging as sophisticated scoring functions are required to discriminate the true positive from the false positive identifications. This problem has been addressed by a number of software that have adopted different strategies for the search of cross-linked peptides against large data sets [162, 168-170] although they still provide certain limitations [167]. An attempt to develop a strategy for the analysis of cross-linked peptides from a whole proteome database has also been made by the developers of the MassMatrix software [134]. However, this feature has not yet been fully developed to enable a reliable identification of cross-linked peptides. For this reason, we restricted our search to a database containing only several ECM proteins that are known as typical TG2 targets including collagen type I, III and V [44, 171, 172], fibrinogen [173] and fibronectin [135]. In general, the scores and the pp values are similar to or higher than the maximum score for the TXLP obtained from LC-MS data generated using the

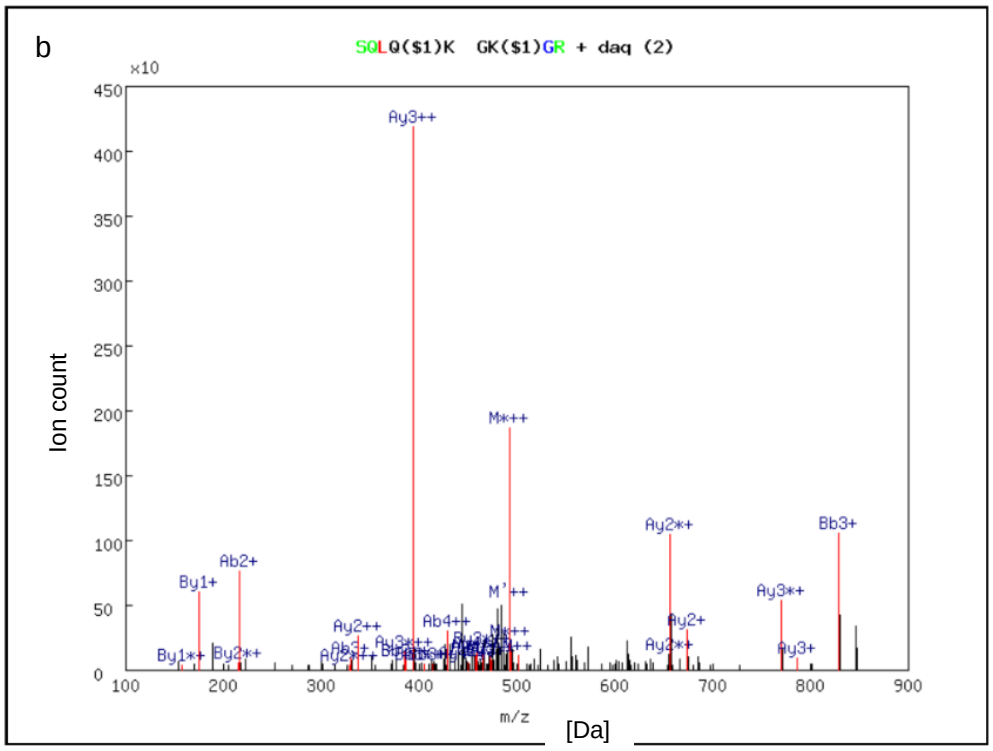
same fragmentation energy (CID 30). The MassMatrix score depends on the abundance of the matched fragments relative to the total ion abundance in the experimental spectrum [123]. As a consequence, slightly lower scores could be observed for spectra showing a minor presence of highly abundant unidentified product ions. In contrast, during the calculation of the pp value, only the number of identified fragments relative to the sequence length. For this reason, high pp values were prioritized in assessing the reliability of the peptide match. The identified cross-linked peptide pair between collagen alpha-1(I) and fibronectin shows slightly lower scores than the TXLP (12 and 13 in comparison with 19 for the model peptide). Nevertheless, this identification was considered as a significant match as it meets the rest of the criteria for a significant hit, such as a pp value that is higher than the one found for the model peptide and the high abundance of fragment ions. Two of the cross-linked peptides located on the collagen 1(I) protein have a hydroxyproline modification that could not be confirmed using the Uniprot database. However, the information provided by Uniprot is based on a similarity with another species (chicken) and not on any published study of the human protein or any mammalian orthologue. Hence, there was no reliable evidence that the identified hydroxyproline modifications were false. Another modification found in this search was the deamidation of glutamine to glutamic acid. According to previous literature, alongside their cross-linking activity, transglutaminases are also able to catalyse glutamine deamidation [174]. Thus, a deamidated glutamine in proximity to the position of cross-link provides further evidence for the increased catalytic activity of this enzyme within a specific region of a substrate protein.

To investigate whether the identified cross-linking positions were specific for TGFβ-1 treated cells, we also performed a search of the LC-MS data obtained from the control replicates (Table 5.2). The results showed that only

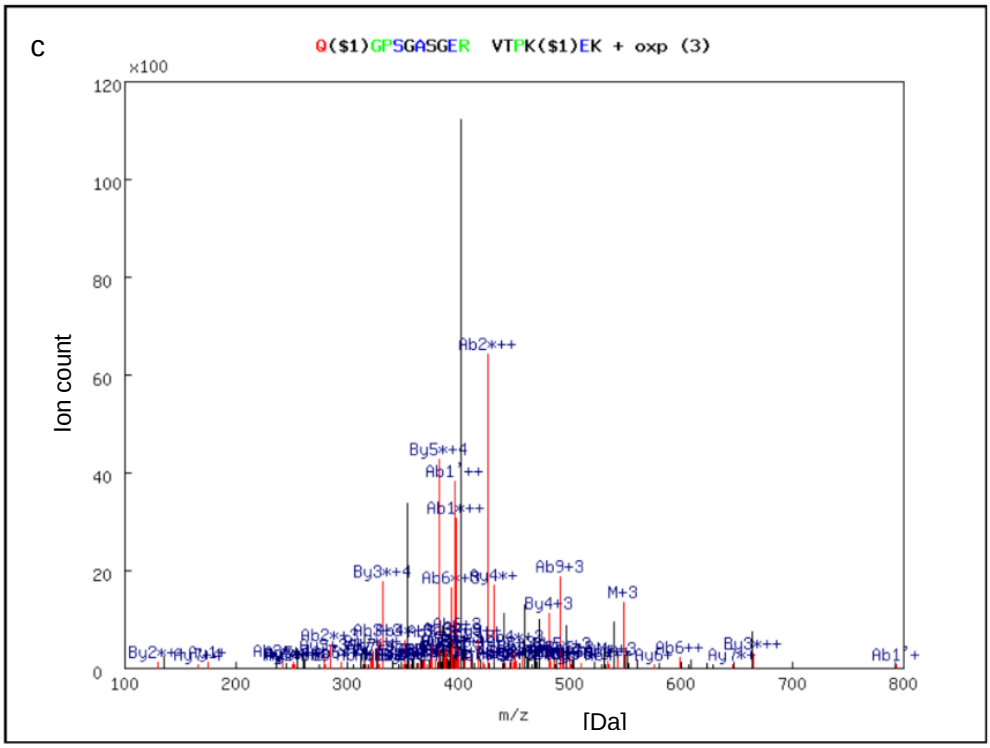
the cross-linked peptide pair between fibrinogen alpha and collagen alpha-3(V) did not appear in any of the control samples suggesting that either this cross-link appears as a result of TGFβ-1 stimulation or its abundance was too low to be identified in the control samples. All the other cross-linked peptides were found in two or more of the control replicates with scores that also match the set thresholds. These results imply that the rest of the identified cross-links did not form in response to the pro-fibrotic treatment; however, a potential difference in the extent of cross-linking on these sites as a result of the TGFβ-1 treatment cannot be excluded. This hypothesis could be tested by comparing the precursor ion intensities between the control and the TGFβ-1 treated samples. Prior to this signal comparison a normalisation of the signal intensities against an internal standard would be necessary to minimize variation between the MS signals. This experiment would require further method development that was not performed in the course of this project due to time limitations.

The MassMatrix peptide search results could not be confirmed with a MeroX search even after increasing the CID energy to 40, as no cross-linked peptides were identified.

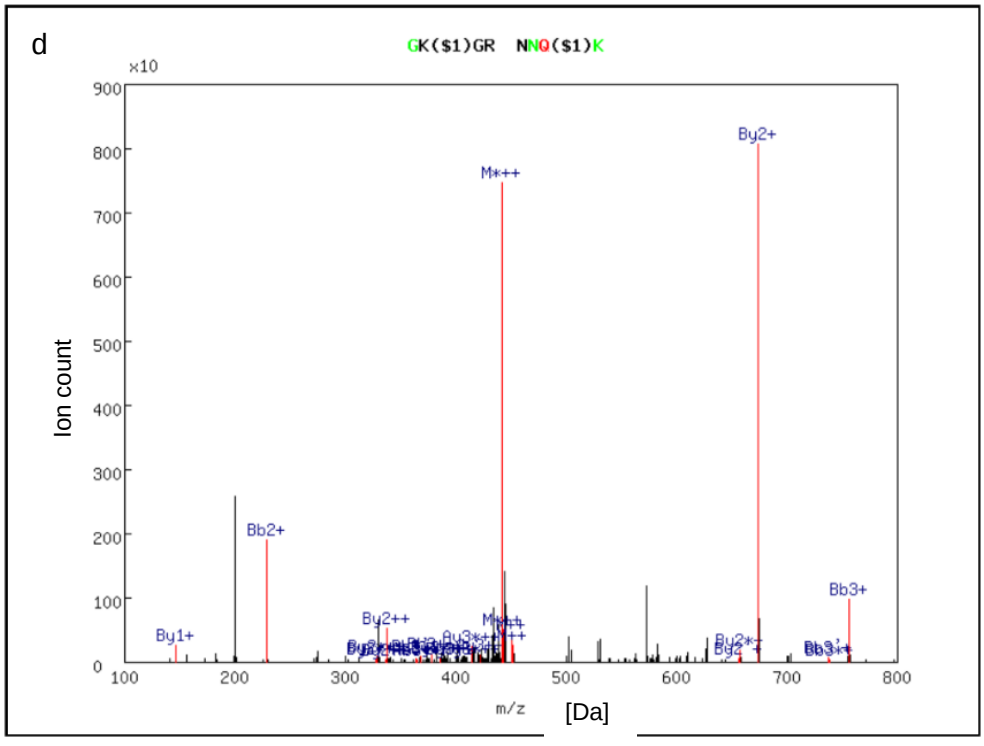

a
LGPQ($1)GAPGQR K($1)GR + daq (9)
×1000
300
250
200
150
100
50
0
Ion count
Ay9++
Ay8++
Ab4++
Ay6+
Ay3+
Ay9*++
Ab5++
Ay3'+
Ay7++
Mk+3
Ay8*++
Ab5*++
Ay6++
Ay7'++
Ay8*++
Ab4+
Ab2+
Ab7*++
Ab5+
Ab5+
100
200
300
400
500
600
700
800
900
m/z
[Da]


b
S0LQ($1)K GK($1)GR + daq (2)
×10
450
400
350
300
250
200
150
100
50
0
Ion count
Ay3++
Mk++
Ay2*+
Bb3+
Ab2+
By1+
M'++
Ay3*+
Ay2++
Ab4++
Ay2+
Ay2*+
By1*+
By2*+
Ay3+
100
200
300
400
500
600
700
800
900
m/z
[Da]

c


Q($1)GPSGASGER VTPK($1)EK + oxp (3)
x100
Ion count
m/z
[Da]
Ab2*++
By5*++
Ab1'++
Ab1*++
By3*++
Ab9+3
By4+3
M+3
By2*++
By1+
By6++
By3*++
Ab1'+

d


GK($1)GR NNQ($1)K
x10
Ion count
m/z
[Da]
By2+
M*++
Bb2+
Bb3+
By1+
By2++

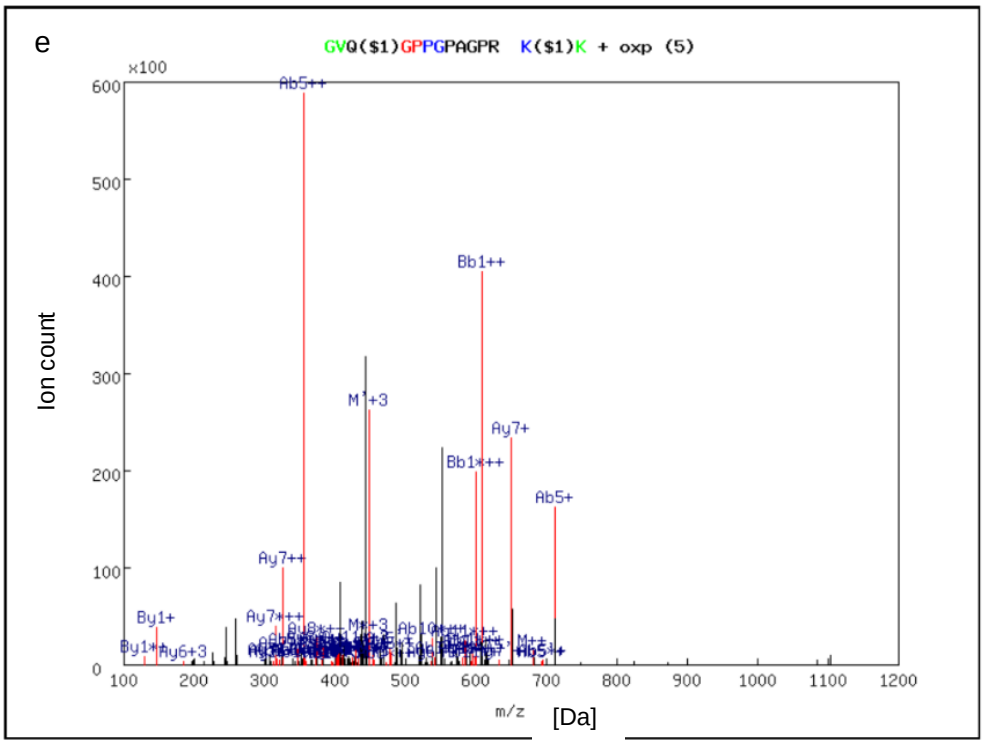



**Figure 5. 1 MS2 fragmentation spectra of the significant hits resulting from the MassMatrix search of transglutaminase cross-linked peptides in ECM isolated from TGFβ-1 treated dermal fibroblast cultures.** The found cross-links are between the following proteins: collagen alpha-2(V) and collagen alpha-3(V) (a), fibrinogen alpha and collagen alpha-3(V) (b), collagen alpha-1(I) and fibronectin (c), collagen alpha-3(V) and fibronectin (d), collagen alpha-1(I) and collagen alpha-3(V). Above each spectrum the cross-link position is marked within the identified peptide sequences with the symbol ($) after the glutamine and lysine residues. Inferred modifications are labelled as daq (glutamine deamidation) and oxp (hydroxyproline). The exact position of the modification is given in brackets that indicate the amino acid number within the peptide sequence. The signals are given in ion counts for a given m/z value.

| *Charge* | Score | pp | $pp_2$ | $pp_{tag}$ | m/z | MW(obs) | MW | Delta | Sequence + modifications | Cross-linked proteins |
|---|---|---|---|---|---|---|---|---|---|---|
| *+3* | 48 | 6.8 | 10.8 | 6.7 | 441.9103 | 1323.7163 | 1323.7128 | 0.0035 | **LGPQGAPGdaQR KGR** | Collagen alpha-2(V)+ Collagen alpha-3(V) |
| | 43 | 6.3 | 9.3 | 5.9 | 441.9102 | 1323.7159 | | 0.0031 | | |
| *+2* | 23 | 8.2 | 11.4 | 2.8 | 502.2798 | 1003.5524 | 1003.5531 | -0.0007 | **SdaQLQK GKGR** | Fibrinogen alpha + Collagen alpha-3(V) |
| | 19 | 10.8 | 10.3 | 2.4 | 502.2795 | 1003.5517 | | -0.0014 | | |
| *+4* | 13 | 13.1 | 4.6 | 2.4 | 411.9546 | 1644.7965 | 1644.8188 | -0.0222 | **QGoxPSGASGER VTPKEK** | Collagen alpha-1(I) + Fibronectin |
| | 12 | 11.7 | 3.9 | 1.5 | 411.9546 | 1644.7967 | | -0.0221 | | |
| *+2* | 21 | 6.9 | 8.2 | 2.3 | 451.7507 | 902.4941 | 902.4803 | 0.0138 | **GKGR NNQK** | Collagen alpha-3(V) + Fibronectin |
| | 22 | 10.2 | 8.0 | 2.3 | 451.7505 | 902.4937 | | 0.0134 | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *+3* | 12 | 6.9 | 4.2 | 2.2 | 454.9167 | 1362.7356 | 1362.7488 | -0.0132 | **GVQGoxPPGPAGPR KK** | Collagen alpha-1(I) + Collagen alpha-3(V) |
| | 20 | 7.8 | 5.7 | 2.2 | 454.9175 | 1362.7381 | | -0.0108 | | |

**Table 5. 1 List of the significant hits resulting from the MassMatrix search for transglutaminase cross-linked peptides in ECM isolated from dermal fibroblast cultures after 30 days of treatment with TGFβ-1 - TGFβ-1 treated samples.** For each peptide pair the scoring results from all the positive replicates are listed. The cross-linked glutamine (Q) and lysine (K) residues are shown in red text. Hydroxyproline and deamidated glutamine residues are labelled as oxP and daQ respectively. The second protein-binding partner of the last cross-linked pair could be any of the proteins listed in brackets, as they all possess the fragment KK in their sequences. All MW values are shown in Da.

| *Charge* | Score | pp | $pp_2$ | $pp_{tag}$ | m/z | MW(obs) | MW | Delta | Sequence + modifications | Cross-linked proteins |
|---|---|---|---|---|---|---|---|---|---|---|
| *+3* | 59 | 6.8 | 12.1 | 7.3 | 441.9091 | 1323.7129 | 1323.7128 | 0.0001 | **LGPQGAPGdaQR KGR** | Collagen alpha-2(V) + Collagen alpha-3(V) |
| | 44 | 8.3 | 12.5 | 4.5 | | | | | | |
| *+4* | 24 | 10.7 | 6.9 | 3.1 | 411.9546 | 1644.7967 | 1644.8188 | -0.0221 | **QGoxPSGASGER VTPKEK** | Collagen alpha-1(I) + Fibronectin |
| | 11 | 10.2 | 4.2 | 3.1 | 411.9547 | 1644.7969 | | -0.0219 | | |
| *+2* | 18 | 8.2 | 7.6 | 2.3 | 451.7508 | 902.4944 | 902.4803 | 0.0141 | **GKGR NNQK** | Collagen alpha-3(V) + Fibronectin |
| | 10 | 4.3 | 6.6 | 1.5 | 451.7509 | 902.4944 | | 0.0142 | | |
| | 15 | 6.4 | 7.6 | 2.3 | 451.7510 | 902.4947 | | 0.0145 | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *+3* | 24 | 7.4 | 6.6 | 2.2 | 454.9167 | 1362.7354 | 1362.7488 | -0.0134 | **GVQGoxPPGPAGPR KK** | Collagen alpha-1(I) + Collagen alpha-3(V) |
| | 18 | 4.3 | 5.2 | 1.8 | 454.9162 | 1362.7351 | | -0.0118 | | |

**Table 5. 2 List of the significant hits resulting from the MassMatrix search for transglutaminase cross-linked peptides in ECM isolated from dermal fibroblast cultures after 30 days of treatment with TGFβ-1 - control samples.** For each peptide pair the scoring results from all the positive replicates are listed. The cross-linked glutamine (Q) and lysine (K) residues are shown in red text. Hydroxyproline and deamidated glutamine residues are labelled as oxP and daQ respectively. The second protein-binding partner of the last cross-linked pair could be any of the proteins listed in the brackets, as they all possess the fragment KK in their sequences.

## Collagen alpha-2(V)

QEEDEDEGYGEEIACTQNGQMYLNRDIWKPAPCQICVCDNGAILCDKIECQDVLDCADPVTPPGECCPVCSQTPGGGNTNFGRGRKGQKGEPGLVPVVTGIRGRPGPAGPPGSQGPRGERGPKGRPGPRGPQ
GIDGEPGVPGQPGAPGPPGHPSHPGPDGLSRPFSAQMAGLDEKSGLGSQVGLMPGSVGPVGPRGPQGLQGQQGGAGPTGPPGEPGDPGPMGPIGSRGPEGPPGKPGEDGEPGRNGNPGEVGFAGSPGAR
GFPGAPGLPGLKGHRGHKGLEGPKGEVGAPGSKGEAGPTGPMGAMGPLGPRGMPGERGRLGPQGAPGQRGAHGMPGKPGPMGPLGIPGSSGFPGNPGMKGEAGPTGARGPEGPQGQRGETGPPGPVGS
PGLPGAIGTDGTPGAKGPTGSPGTSGPPGSAGPPGSPGPQGSTGPQGIRGQPGDPGVPGFKGEAGPKGEPGPHGIQGPIGPPGEEGKRGPRGDPGTVGPPGPVGERGAPGNRGFPGSDGLPGPKGAQGER
GPVGSSGPKGSQGDPGRPGEPGLPGARGLTGNPGVQGPEGKLGPLGAPGEDGRPGPPGSIGIRGQPGSMGLPGPKGSSGDPGKPGEAGNAGVPGQRGAPGKDGEVGPSGPVGPPGLAGERGEQGPPGPT
GFQGLPGPPGPPGEGGKPGDQGVPGDPGAVGPLGPRGERGNPGERGEPGITGLPGEKGMAGGHGPDGPKGSPGPSGTPGDTGPPGLQGMPGERGIAGTPGPKGDRGGIGEKGAEGTAGNDGARGLPGPL
GPPGPAGPTGEKGEPGPRGLVGPPGSRGNPGSRGENGPTGAVGFAGPQGPDGQPGVKGEPGEPGQKGDAGSPGPQGLAGSPGPHGPNGVPGLKGGRGTQGPPGATGFPGSAGRVGPPGPAGAPGPAGP
LGEPGKEGPPGLRGDPGSHGRVGDRGPAGPPGGPGDKGDPGEDGQPGPDGPPGPAGTTGQRGIVGMPGQRGERGMPGLPGPAGTPGKVGPTGATGDKGPPGPVGPPGSNGPVGEPGPEGPAGNDGTPG
RDGAVGERGDRGDPGPAGLPGSQGAPGTPGPVGAPGDAGQRGDPGSRGPIGPPGRAGKRGLPGPQGPRGDKGDHGDRGDRGQKGHRGFTGLQGLPGPPGPNGEQGSAGIPGPFGPRGPPGPVGPSGKE
GNPGPLGPIGPPGVRGSVGEAGPEGPPGEPGPPGPPGPPGPPGHL

## Collagen alpha-3(V)

DPVDVLKALGVQGGQAGVPEGPGFCPQRTPEGDRAFRIGQASTLGIPTWELFPEGHFPENFSLLITLRGQPANQSVLLSIYDERGARQLGLALGPALGLLGDPFRPLPQQVNLTDGRWHRVAVSIDGEMVTLVAD
CEAQPPVLGHGPRFISIAGLTVLGTQDLGEKTFEGDIQELLISPDPQAAFQACERYLPDCDNLAPAATVAPQGEPETPRPRRKGKGKGRKKGRGRKGKGRKKNKEIWTSSPPPDSAENQTSTDIPKTETPAPNLPP
TPTPLVVTSTVTTGLNATILERSLDPDSGTELGTLETKAAREDEEGDDSTMGPDFRAAEYPSRTQFQIFPGAGEKGAKGEPAVIEKGQQFEGPPGAPGPQGVVGPSGPPGPPGFPGDPGPPGPAGLPGIPGIDGIR
GPPGTVIMMPFQFAGGSFKGPPVSFQQAQAQAVLQQTQLSMKGPPGPVGLTGRPGPVGLPGHPGLKGEEGAEGPQGPRGLQGPHGPPGRVGKMGRPGADGARGLPGDTGPKGDRGFDGLPGLPGEKGQR
GDFGHVGQPGPPGEDGERGAEGPPGPTGQAGEPGPRGLLGPRGSPGPTGRPGVTGIDGAPGAKGNVGPPGEPGPPGQQGNHGSQGLPGPQGLIGTPGEKGPPGNPGIPGLPGSDGPLGHPGHEGPTGEKG
AQGPPGSAGPPGYPGPRGVKGTSGNRGLQGEKGEKGEDGFPGFKGDVGLKGDQGKPGAPGPRGEDGPEGPKGQAGQAGEEGPPGSAGEKGKLGVPGLPGYPGRPGPKGSIGFPGPLGPIGEKGKSGKTGQ
PGLEGERGPPGSRGERGQPGATGQPGPKGDVGQDGAPGIPGEKGLPGLQGPPGFPGPKGPPGPHQGKDGRPGHPGQRGELGFQGQTGPPGPAGVLGPQGKTGEVGPLGERGPPGPPGPPGPPGEQGLPGLEGR
EGAKGELGPPGPLGKEGPAGLRGFPGPKGGPGDPGPTGLKGDKGPPGPVGANGSPGERGPLGPAGGIGLPGQSGSEGPVGPAGKKGSRGERGPPGPTGKDGIPGPLGPLGPPGAAGPSGEEGDKGDVGAP
GHKGSKGDKGDAGPPGQPGIRGPAGHPGPPGADGAQGRRGPPGLFGQKGDDGVRGFVGVIGPPGLQGLPGPPGEKGEVGDVGSMGPHGAPGPRGPQGPTGSEGTPGLPGGVGQPGAVGEKGERGDAGD
PGPPGAPGIPGPKGDIGEKGDSGPSGAAGPPGKKGPPGEDGAKGSVGPTGLPGDLGPPGDPGVSGIDGSPGEKGDPGDVGGPGPPGASGEPGAPGPPGKRGPSGHMGREGREGEKGAKGEPGPDGPPGRT
GPMGARGPPGRVGPEGLRGIPGPVGEPGLLGAPGQMGPPGPLGPSGLPGLKGDTGPKGEKGHIGLIGLIGPPGEAGEKGDQGLPGVQGPPGPKGDPGPPGPIGSLGHPGPPGVAGPLGQKGSKGSPGSMGP
RGDTGPAGPPGPPGAPAELHGLRRRRRFVPVPLPVVEGGLEEVLASLTSLSLELEQLRRPPGTAERPGLVCHELHRNHPHLPDGEYWIDPNQGCARDSFRVFCNFTAGGETCLYPDKKFEIVKLASWSKEKPGG
WYSTFRRGKKFSYVDADGSPVNVVQLNFLKLLSATARQNFTYSCQNAAAWLDEATGDYSHSARFLGTNGEELSFNQTTAATVSVPQDGCRLRKGQTKTLFEFSSSRAGFLPLWDVAATDFGQTNQKFGFELGPV
CFSS

## Fibrinogen alpha

ADSGEGDFLAEGGGVRGPRVVERHQSACKDSDWPFCSDEDWNYKCPSGCRMKGLIDEVNQDFTNRINKLKNSLFEYQKNNKDSHSLTTNIMEILRGDFSSANNRDNTYNRVSEDLRSRIEVLKRKVIEKVQHIQL
LQKNVRAQLVDMKRLEVDIDIKIRSCRGSCSRALAREVDLKDYEDQQKQLEQVIAKDLLPSRDRQHLPLIKMKPVPDLVPGNFKSQLQKVPPEWKALTDMPQMRMELERPGGNEITRGGSTSYGTGSETESPRNP
SSAGSWNSGSSGPGSTGNRNPGSSGTGGTATWKPGSSGPGSTGSWNSGSSGTGSTGNQNPGSPRPGSTGTWNPGSSERGSAGHWTSESSVSGSTGQWHSESGSFRPDSPGSGNARPNNPDWGTFEEVS
GNVSPGTRREYHTEKLVTSKGDKELRTGKEKVTSGSTTTTRRSCSKTVTKTVIGPDGHKEVTKEVVTSEDGSDCPEAMDLGTLSGIGTLDGFRHRHPDEAAFFDTASTGKTFPGFFSPMLGEFVSETESRGSESGI
FTNTKESSSHHPGIAEFPSRGKSSSYSKQFTSSTSYNRGDSTFESKSYKMADEAGSEADHEGTHSTKRGHAKSRPVRDCDDVLQTHPSGTQSGIFNIKLPGSSKIFSVYCDQETSLGGWLLIQQRMDGSLNFNRT
WQDYKRGFGSLNDEGEGEFWLGNDYLHLLTQRGSVLRVELEDWAGNEAYAEYHFRVGSEAEGYALQVSSYEGTAGDALIEGSVEEGAEYTSHNNMQFSTFDRDADQWEENCAEVYGGGWWYNNCQAANLN
GIYYPGGSYDPRNNSPYEIENGVVWVSFRGADYSLRAVRMKIRPLVTQ

## Collagen alpha-1(I)

QLSYGYDEKSTGGISVPGPMGPSGPRGLPGPPGAPGPQGFQGPPGEPGEPGASGPMGPRGPPGPPGKNGDDGEAGKPGRPGERGPPGPQGARGLPGTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEP
GSPGENGAPGQMPRGLPGERGRPGAPGPAGARGNDGATGAAGPPGPTGPAGPPGFPGAVGAKGEAGPQGPRGSEGPQGVRGEGPPGPAGAAGPAGNPGADGQPGAKGANGAPGIAGAPGFPGARGPSG
PQGPGGPPGPKGNSGEPGAPGSKGDTGAKGEPGPVGVQGPPGPAGEEGKRGARGEPGPTGLPGPPGERGGPGSRGFPGADGVAGPKGPAGERGSPGPAGPKGSPGEAGRPGEAGLPGAKGLTGSPGSP
GPDGKTGPPGPAGQDGRPGPPGPPGARGQAGVMGFPGPKGAAGEPGKAGERGVPGPPGAVGPAGKDGEAGAQGPPGPAGPAGERGEQGPAGSPGFQGLPGPAGPPGEAGKPGEQGVPGDLGAPGPSGA
RGERGFPGERGVQGPPGPAGPRGANGAPGNDGAKGDAGAPGAPGSQGAPGLQGMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGPIGPPGPAGAPGDKGESGPSGPAGPTGARGAPGDRG
EPGPPGPAGFAGPPGADGQPGAKGEPGDAGAKGDAGPPGPAGPAGPPGPIGNVGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPSGNAGPPGPPGPAGKEGGKGPRGETGPAGRPGEVGPPGPPGPAG
EKGSGADGPAGAPGTPGPQGIAGQRGVVGLPGQRGERGFPGLPGPSGEPGKQGPSGASGERGPPGPMGPPGLAGPPGESGREGAPGAEGSPGRDGSPGAKGDRGETGPAGPPGAPGAPGAPGPVGPAG
KSGDRGETGPAGPTGPVGPVGARGPAGPQGPRGDKGETGEQGDRGIKGHRGFSGLQGPPGPPGSPGEQGPSGASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG
PPSAGFDFSFLPQPPQEKAHDGGRYYRA

Fibronectin

QAQQMVQPQSPVAVSQSKPGCYDNGKHYQINQQWERTYLGNALVCTCYGGSRGFNCESKPEAEETCFDKYTGNTYRVGDTYERPKDSMIWDCTCIGAGRGRISCTIANRCHEGGQSYKIGDTWRRPHETGGY
MLECVCLGNGKGEWTCKPIAEKCFDHAAGTSYVVGETWEKPYQGWMMVDCTCLGEGSGRITCTSRNRCNDQDTRTSYRIGDTWSKKDNRGNLLQCICTGNGRGEWKCERHTSVQTTSSGSGPFTDVRAAVY
QPQPHPQPPPYGHCVTDSGVVYSVGMQWLKTQGNKQMLCTCLGNGVSCQETAVTQTYGGNSNGEPCVLPFTYNGRTFYSCTTEGRQDGHLWCSTTSNYEQDQKYSFCTDHTVLVQTRGGNSNGALCHFPFL
YNNHNYTDCTSEGRRDNMKWCGTTQNYDADQKFGFCPMAAHEEICTTNEGVMYRIGDQWDKQHDMGHMMRCTCVGNGRGEWTCIAYSQLRDQCIVDDITYNVNDTFHKRHEEGHMLNCTCFGQGRGRWKC
DPVDQCQDSETGTFYQIGDSWEKYVHGVRYQCYCYGRGIGEWHCQPLQTYPSSSGPVEVFITETPSQPNSHPIQWNAPQPSHISKYILRWRPKNSVGRWKEATIPGHLNSYTIKGLKPGVVYEGQLISIQQYGHQ
EVTRFDFTTTSTSTPVTSNTVTGETTPFSPLVATSESVTEITASSFVVSWVSASDTVSGFRVEYELSEEGDEPQYLDLPSTATSVNIPDLLPGRKYIVNVYQISEDGEQSLILSTSQTTAPDAPPDTTVDQVDDTSIVVR
WSRPQAPITGYRIVYSPSVEGSSTELNLPETANSVTLSDLQPGVQYNITIYAVEENQESTPVVIQQETTGTPRSDTVPSPRDLQFVEVTDVKVTIMWTPPESAVTGYRVDVIPVNLPGEHGQRLPISRNTFAEVTGLS
PGVTYYFKVFAVSHGRESKPLTAQQTTKLDAPTNLQFVNETDSTVLVRWTPPRAQITGYRLTVGLTRRGQPRQYNVGPSVSKYPLRNLQPASEYTVSLVAIKGNQESPKATGVFTTLQPGSSIPPYNTEVTETTIVIT
WTPAPRIGFKLGVRPSQGGEAPREVTSDSGSIVVSGLTPGVEYVYTIQVLRDGQERDAPIVNKVVTPLSPPTNLHLEANPDTGVLTVSWERSTTPDITGYRITTTPTNGQQGNSLEEVVHADQSSCTFDNLSPGLEY
NVSVYTVKDDKESVPISDTIIPEVPQLTDLSFVDITDSSIGLRWTPLNSSTIIGYRITVVAAGEGIPIFEDFVDSSVGYYTVTGLEPGIDYDISVITLINGGESAPTTLTQQTAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLT
NFLVRYSPVKNEEDVAELSISPSDNAVVLTNLLPGTEYVVSVSSVYEQHESTPLRGRQKTGLDSPTGIDFSDITANSFTVHWIAPRATITGYRIRHHPEHFSGRPREDRVPHSRNSITLTNLTPGTEYVVSIVALNGRE
ESPLLIGQQSTVSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGSKSTATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEIDKPSQMQVTDVQDNSISVKWLPSSSPVTGYRV
TTTPKNGPGPTKTKTAGPDQTEMTIEGLQPTVEYVVSVYAQNPSGESQPLVQTAVTNIDRPKGLAFTDVDVDSIKIAWESPQGQVSRYRVTYSSPEDGIHELFPAPDGEEDTAELQGLRPGSEYTVSVVALHDDME
SQPLIGTQSTAIPAPTDLKFTQVTPTSLSAQWTPPNVQLTGYRVRVTPKEKTGPMKEINLAPDSSSVVVSGLMVATKYEVSVYALKDTLTSRPAQGVVTTLENVSPPRRARVTDATETTITISWRTKTETITGFQVDA
VPANGQTPIQRTIKPDVRSYTITGLQPGTDYKIYLYTLNDNARSSPVVIDASTAIDAPSNLRFLATTPNSLLVSWQPPRARITGYIIKYEKPGSPPREVVPRPRPGVTEATITGLEPGTEYTIYVIALKNNQKSEPLIGRKK
TDELPQLVTLPHPNLHGPEILDVPSTVQKTPFVTHPGYDTGNGIQLPGTSGQQPSVGQQMIFEEHGFRRTTPPTTATPIRHRPRPYPPNVGEEIQIGHIPREDVDYHLYPHGPGLNPNASTGQEALSQTTISWAPFQ
DTSEYIISCHPVGTDEEPLQFRVPGTSTSATLTGLTRGATYNVIVEALKDQQRHKVREEVVTVGNSVNEGLNQPTDDSCFDPYTVSHYAVGDEWERMSESGFKLLCQCLGFGSGHFRCDSSRWCHDNGVNYKIG
EKWDRQGENGQMMSCTCLGNGKGEFKCDPHEATCYDDGKTYHVGEQWQKEYLGAICSCTCFGGQRGWRCDNCRRPGGEPSPEGTTGQSYNQYSQRYHQRTNTNVNCPIECFMPLDVQADREDSRE

**Figure 5. 2 Transglutaminase cross-link positions in the sequences of the ECM proteins from the TGFβ-1 treated samples.** The cross-linked lysine and glutamine residues are highlighted in red. The nonhelical regions of the collagen proteins are highlighted in yellow. The cross-linked lysine residue in fibronectin is located in the type-III 14 domain whereas the cross-linked glutamine residue was found in the type-III 16 domain (highlighted in green). Both domains (highlighted in green) are a part of the heparin binding region (source: Uniprot).

Interestingly, among the five identified cross-linked peptides there are three different lysine-donor peptides that were all assigned to a short region of the collagen alpha-3(V) protein. These are the tryptic peptides with the sequences GKGR, KK and KGR (Figure 5.3). The peptide sequence (R/K)KK can however be found in many ECM proteins including collagen alpha-1(V) and fibronectin. Therefore, the identity of the protein that contained this cross-linking site cannot be determined. Based on these collagen alpha-3(V) sequence assignments, it is likely that the transglutaminase cross-linking activity is high in this region relative to other regions of this and other ECM proteins, at least as far as lysine residues are concerned. Other identified cross-liked peptides were also found in other ECM proteins, including the Latent transforming growth factor beta binding protein 2 (LTBP-2) variant and various laminin subunits. However, no direct evidence for the cross-linking of these proteins by TG2 has yet been reported. It was found the Latent transforming growth factor beta binding protein 1 (LTBP-1) is cross-linked by TG2 as a part of the TGFβ activation [69] and other isoforms such as LTBP-2 can potentially also be targets for TG2. The only peptide that was found to be unique to the protein database that we used for our search is the sequence SQLQK (Table 5.1).

Transglutaminases may have a preference for more highly exposed protein sites that can provide an explanation for the detected cross-linking in the non-helical region of the collagen alpha-3(V) protein. This result is consistent with another study on rhabdomyosarcoma cell cultures that detected TG2 cross-linking activity only in the non-helical region of collagen V [171]. Furthermore, a microbial transglutaminase activity assay using a peptide substrate library showed that neighbouring amino acid residues can play a crucial role in the rate of cross-link formation [129]. Hydrophobic and basic amino acid residues showed a particularly strong and positive influence on the rate of cross-link

formation. Although the selectivity of transglutaminases is not yet fully elucidated, and while it is therefore not possible to identify protein substrate, the high incidence of basic residues in the collagen alpha-3(V) non-helical region identified using MassMatrix is consistent with the current understanding of substrate exposure and sequence preference.

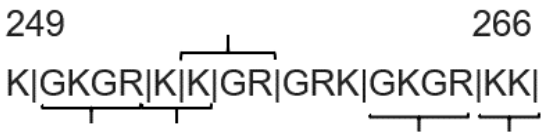



**Figure 5. 3 Amino acid sequence of the non-helical collagen alpha-3(V) region that contained multiple transglutaminase cross-links.** Deduced trypsin cleavage sites are marked with vertical lines, and the sequences identified as cross-linked tryptic peptides by MassMatrix searches are marked with horizontal brackets.

Type V collagen is a protein that shows some structural similarities with the other types of fibrilar collagen as well as several distinctive characteristics [175]. For instance, unlike collagen I and II it has a loosely twisted ribbon-like structure with a significantly smaller thickness of the collagen fibrils and little branching and merging [175]. The higher number of post translationally glycosylated hydroxylysine residues is another characteristic feature of collagen V [175]. Interestingly, a microscopic study suggests, that collagen V fibrils show a decreased binding with each other in comparison with collagen I [175]. Type V collagen is a typical target for cross-linking enzymes such as transglutaminases [171] and the lysyl oxidase [176].

The cross-linked fibronectin sites identified by MassMatrix do not belong to the collagen-binding region (residues 308 - 608; UniProtKB accession code P02751). Instead, they are located in type-III domains 13 (residues 1724 - 1817) and 15 (residues 1905 - 1995) that are located within heparin-biding site 2 (residues 1721 - 1991). However, the reliability of the latter domain function is questionable, as no reference to a publication is present in the

protein database. The collagen-binding domain of fibronectin was discovered based on studies of different fibronectin fragments that showed high affinity for gelatin-Sepharose matrices [177]. However, it has not yet been studied whether the heparin binding 2 domain has an affinity for collagen alpha-3(V). Hence, there is currently no evidence against the existence of transglutaminase cross-links in this region of fibronectin. Binding of fibronectin to the propeptide of collagen alpha-1(V) reported in the literature [178] suggests a possible interaction with this collagen type.

### 5.2.2 Identification of transglutaminase cross-linked sites in the TSK2 mouse model

The results from the search for transglutaminase cross-linked peptides in the 10-week-old TSK2 mouse model are presented in Figure 5.4 and Table 5.3. The positions of the cross-links in the protein sequences are shown in Figure 5.5. Although the search with the data generated from the 23-week-old mice experiment could have potentially led to a larger number of significant hits, these samples showed very high variation in terms of protein quantitation (data not shown). Thus, finding a significant peptide hit across the majority of the replicates within the fibrotic condition was less likely. For this reason, only data from the 10-week-old mice were analysed, as these samples showed lower variation. Cross-linked peptides identified in at least three of the four replicates per condition were considered to be significant.

A MassMatrix search resulted in the identification of two significant cross-links, both of which were between the alpha-1 chain of type V collagen and fibronectin. Both cross-links were also identified with similar scores in the control replicates (Table 5.4). As found in the cell culture model, these results also suggest that the identified cross-linking sites do not result from an

increase in the activity of transglutaminases in the ECM in response to pro-fibrotic stimulation. Nevertheless, as previously mentioned, quantitative differences in these cross-linking sites between control and TSK2 skin could still be present. To test this, further method development is required. Moreover, the search did not identify any of the cross-linking sites found in the cell culture model (Tables 5.1 and 5.2). This could have been an effect of the higher proteomic complexity of the skin tissue ECM in comparison with the cell culture ECM on the LC-MS separation efficiency. These limitations could also explain the overall lower amount of cross-linked peptides found in the skin tissue in comparison with the cell cultures. Furthermore, there might have been differences in the transglutaminase cross-linking pattern between different organisms and in different ECM settings. This could for example entail changes of cross-linking activity of different enzymes, or differences in the conformation of a substrate protein that affect the exposure of different residues.

Similar to the findings from the TGFβ-1 stimulation of dermal fibroblasts, cross-linking of the heparin 2 binding site of fibronectin was also identified in the TSK2 samples, as the EKTGPMK peptide is located in this region. The other cross-linked fibronectin peptide belongs to the cell-attachment site that is responsible for the interaction of the protein with integrin receptors on the cell surface [179].

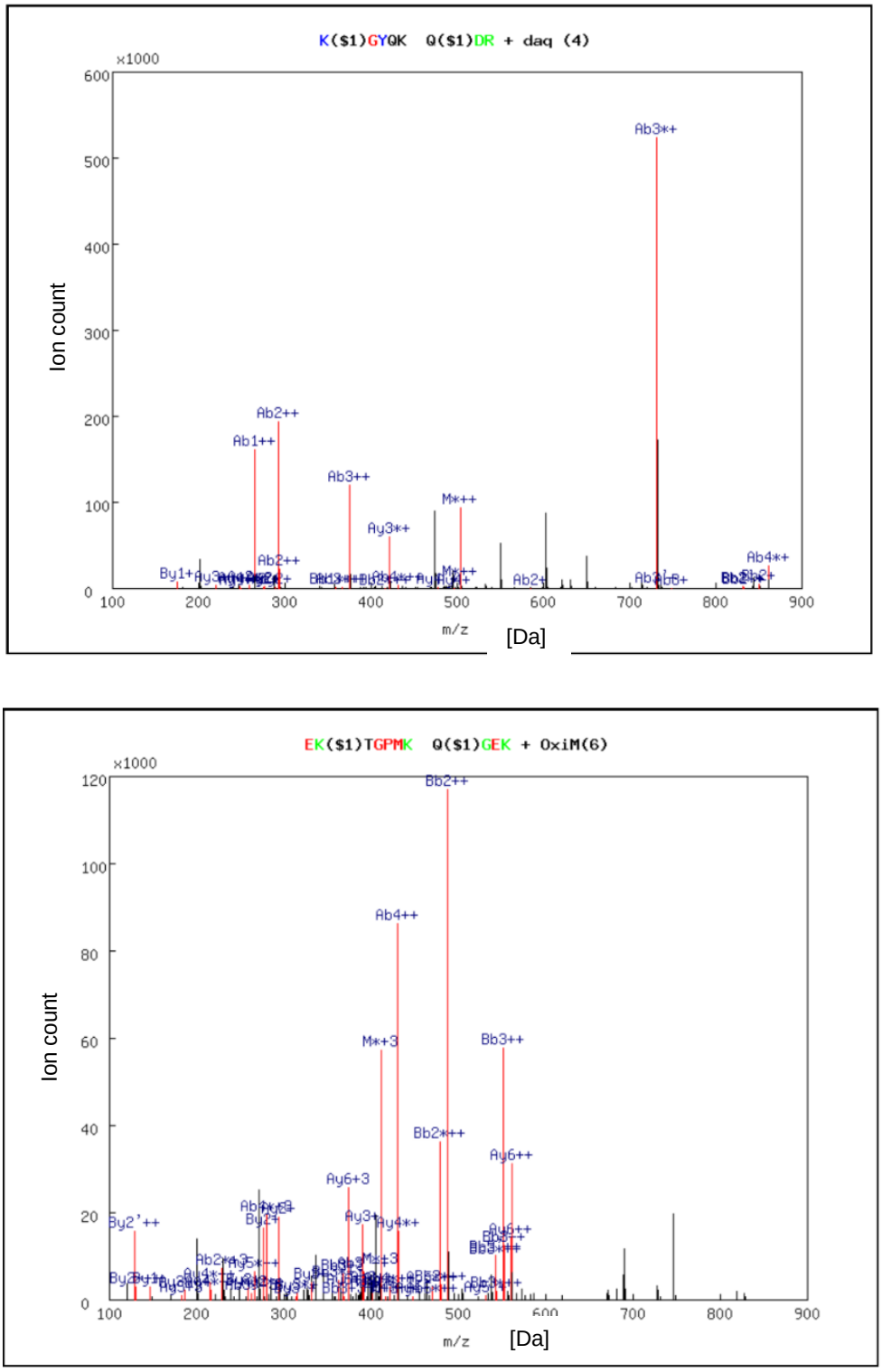



**Figure 5. 4 MS2 fragmentation spectra of the significant hits resulting from the MassMatrix search of transglutaminase cross-linked peptides in ECM extracted from the TSK2 mouse model.** The cross-link position is marked with the symbol ($) after the glutamine and lysine residues. The modifications found are labelled as daq (deamidated glutamine) and oxp (hydroxyproline). The exact position of the modification within the peptide sequence is shown in brackets. The signals are given in ion counts for a given m/z value.

| *charge* | score | pp | $pp_2$ | $pp_{tag}$ | m/z | MW(obs) | MW | delta | sequence + modifications | Cross-linked proteins |
|---|---|---|---|---|---|---|---|---|---|---|
| *+2* | 27 | 8.1 | 8.2 | 2.8 | 512.7509 | 1024.4946 | 1024.5058 | -0.0112 | **KGYdaQK QDR** | Collagen alpha-1(V) + Fibronectin |
| | 20 | 8.1 | 8.2 | 2.3 | 512.7495 | 1024.4917 | | -0.0141 | | |
| | 27 | 7.6 | 9.1 | 3.4 | 512.7409 | 1024.4944 | | -0.0114 | | |
| *+3* | 41 | 8.8 | 10.9 | 5.3 | 417.2120 | 1249.6214 | 1249.6093 | 0.0121 | **EKTGPoxMK QGEK** | Fibronectin + Collagen alpha-1(V) |
| | 36 | 7.7 | 10.8 | 5.3 | 417.2117 | 1249.6206 | | 0.0114 | | |
| | 29 | 6.2 | 11.2 | 5.4 | 417.2108 | 1249.6177 | | 0.0084 | | |

**Table 5. 3 Significant hits resulting from the MassMatrix search for transglutaminase cross-linked peptides in the TSK2 mouse model – TSK2 samples.** The cross-linked glutamine (Q) and lysine (K) residues in each peptide pair are shown in red text. Oxidised methionine and deamidated glutamine residues are labelled as oxM and daQ, respectively. All molecular weight values are shown in Da.

| *charge* | score | pp | $pp_2$ | $pp_{tag}$ | m/z | MW(obs) | MW | delta | sequence + modifications | Cross-linked proteins |
|---|---|---|---|---|---|---|---|---|---|---|
| *+2* | 25 | 6.4 | 8.0 | 4.2 | 512.7508 | 1024.4943 | 1024.5058 | -0.0115 | **KGYdaQK QDR** | Collagen alpha-1(V) + Fibronectin |
| | 25 | 6.4 | 7.7 | 3.1 | 512.7509 | 1024.4944 | | -0.0114 | | |
| | 27 | 8.1 | 7.9 | 2.2 | 512.7509 | 1024.4944 | | -0.0114 | | |
| | 28 | 9.4 | 7.9 | 5.4 | 512.7509 | 1024.4944 | | -0.0114 | | |
| *+3* | 41 | 7.2 | 12.3 | 5.4 | 417.2113 | 1249.6193 | 1249.6093 | 0.0100 | **EKTGPoxMK QGEK** | Fibronectin + Collagen alpha-1(V) |
| | 29 | 8.8 | 10.0 | 5.4 | 417.2118 | 1249.6208 | | 0.0115 | | |

**Table 5. 4 Significant hits resulting from the MassMatrix search for transglutaminase cross-linked peptides in the TSK2 mouse model - control samples.** The cross-linked glutamine (Q) and lysine (K) residues in each peptide pair are shown in red text. Oxidised methionine and deamidated glutamine residues are labelled as oxM and daQ, respectively.

## Collagen alpha-1(V)

AQPADLLEMLDFHNLPSGVTKTTGFCATRRSSSEPDVAYRVSKDAQLSMPTKQLYPESGFPEDFSILTTVKAKKGSQAFLVSIYNEQGIQQLGLELGRSPVFLYEDHTGKPGPEEYPLFPGINLSDGKWHRIALSVY
KKNVTLILDCKKKITKFLSRSDHPIIDTNGIVMFGSRILDDEIFEGDIQQLLFVSDNRAAYDYCEHYSPDCDTAVPDTPQSQDPNPDEYYPEGEGETYYYEYPYYEDPEDPGKEPAPTQKPVEAARETTEVPEEQTQP
LPEAPTVPETSDTADKEDSLGIGDYDYVPPDDYYTPPPYEDFGYGEGVENPDQPTNPDSGAEVPTSTTVTSNTSNPAPGEGKDDLGGEFTEETIKNLEENYYDPYFDPDSDSSVSPSEIGPGMPANQDTIFEGIGG
PRGEKGQKGEPAIIEPGMLIEGPPGPEGPAGLPGPPGTTGPTGQMGDPGERGPPGRPGLPGADGLPGPPGTMLMLPFRFGGGGDAGSKGPMVSAQESQAQAILQQARLALRGPAGPMGLTGRPGPMGPPGS
GGLKGEPGDMGPQGPRGVQGPPGPTGKPGRRGRAGSDGARGMPGQTGPKGDRGFDGLAGLPGEKGHRGDPGPSGPPGIPGDDGERGDDGEVGPRGLPGEPGPRGLLGPKGPPGPPGPPGVTGMDGQP
GPKGNVGPQGEPGPPGQQGNPGAQGLPGPQGAIGPPGEKGPLGKPGLPGMPGADGPPGHPGKEGPPGEKGGQGPPGPQGPIGYPGPRGVKGADGIRGLKGTKGEKGEDGFPGFKGDMGIKGDRGEIGPPG
PRGEDGPEGPKGRGGPNGDPGPLGPTGEKGKLGVPGLPGYPGRQGPKGSIGFPGFPGANGEKGGRGTPGKPGPRGQRGPTGPRGERGPRGITGKPGPKGNSGGDGPAGPPGERGPNGPQGPTGFPGPKG
PPGPPGKDGLPGHPGQRGETGFQGKTGPPGPPGVVGPQGPTGETGPMGERGHPGPPGPPGEQGLPGAAGKEGTKGDPGPAGLPGKDGPPGLRGFPGDRGLPGPVGALGLKGSEGPPGPPGPAGSPGERG
PAGAAGPIGIPGRPGPQGPPGPAGEKGLPGEKGPQGPAGRDGLQGPVGLPGPAGPVGPPGEDGDKGEIGEPGQKGSKGDKGEQGPPGPTGPQGPIGQPGPSGADGEPGPRGQQGLFGQKGDEGSRGFPGP
PGPVGLQGLPGPPGEKGETGDVGQMGPPGPPGPRGPSGAPGADGPQGPPGGIGNPGAVGEKGEPGEAGDPGLPGEGGPLGPKGERGEKGEAGPSGAAGPPGPKGPPGDDGPKGSPGPVGFPGDPGPPG
EPGPAGQDGPPGDKGDDGEPGQTGSPGPTGEPGPSGPPGKRGPPGPAGPEGRQGEKGAKGEAGLEGPPGKTGPIGPQGAPGKPGPDGLRGIPGPVGEQGLPGSPGPDGPPGPMGPPGLPGLKGDSGPKG
EKGHPGLIGLIGPPGEQGEKGDRGLPGPQGSSGPKGDQGITGPSGPLGPPGPPGLPGPPGPKGAKGSSGPTGPKGEAGHPGLPGPPGPPGEVIQPLPIQASRTRRNIDASQLLDDGAGESYVDYADGMEEIFGS
LNSLKLEIEQMKRPLGTQQNPARTCKDLQLCHPDFPDGEYWVDPNQGCSRDSFKVYCNFTAGGSTCVFPDKKSEGARITSWPKENPGSWFSEFKRGKLLSYVDAEGNPVGVVQMTFLRLLSASAHQNVTYNCY
QSVAWQDAATGSYDKAIRFLGSNDEEMSYDNNPYIRALVDGCATKKGYQKTVLEIDTPKVEQVPIVDIMFNDFGEASQKFGFEVGPACFLG

## Fibronectin (Fibronectin type-III 10 (cell attachment), Fibronectin type-III 14(heparin binding 2))

MLRGPGPGRLLLLAVLCLGTSVRCTEAGKSKRQAQQIVQPQSPVAVSQSKPGCFDNGKHYQINQQWERTYLGNALVCTCYGGSRGFNCESKPEPEETCFDKYTGNTYKVGDTYERPKDSMIWDCTCIGAGRGRI
SCTIANRCHEGGQSYKIGDKWRRPHETGGYMLECLCLGNGKGEWTCKPIAEKCFDHAAGTSYVVGETWEKPYQGWMMVDCTCLGEGNGRITCTSRNRCNDQDTRTSYRIGDTWSKKDNRGNLLQCVCTGNG
RGEWKCERHALQSASAGSGSFTDVRTAIYQPQTHPQPAPYGHCVTDSGVVYSVGMQWLKSQGNKQMLCTCLGNGVSCQETAVTQTYGGNSNGEPCVLPFTYNGRTFYSCTTEGRQDGHLWCSTTSNYEQDQ
KYSFCTDHAVLVQTRGGNSNGALCHFPFLYNNRNYTDCTSEGRRDNMKWCGTTQNYDADQKFGFCPMAAHEEICTTNEGVMYRIGDQWDKQHDLGHMMRCTCVGNGRGEWACIPYSQLRDQCIVDDITYNVN
DTFHKRHEEGHMLNCTCFGQGRGRWKCDPIDQCQDSETRTFYQIGDSWEKFVHGVRYQCYCYGRGIGEWHCQPLQTYPGTTGPVQVIITETPSQPNSHPIQWNAPEPSHITKYILRWRPKTSTGRWKEATIPGH
LNSYTIKGLTPGVIYEGQLISIQQYGHREVTRFDFTTSASTPVTSNTVTGETAPYSPVVATSESVTEITASSFVVSWVSASDTVSGFRVEYELSEEGDEPQYLDLPSTATSVNIPDLLPGRKYIVNVYQISEEGKQSLILS
TSQTTAPDAPPDPTVDQVDDTSIVVRWSRPQAPITGYRIVYSPSVEGSSTELNLPETANSVTLSDLQPGVQYNITIYAVEENQESTPVFIQQETTGTPRSDNVPPPTDLQFVELTDVKVTIMWTPPDSVVSGYRVEVL
PVSLPGEHGQRLPVNRNTFAEITGLSPGVTYLFKVFAVHQGRESNPLTAQQTTKLDAPTNLQFVNETDRTVLVTWTPPRARIAGYRLTAGLTRGGQPKQYNVGPLASKYPLRNLQPGSEYTVTLVAVKGNQQSPK
ATGVFTTLQPLRSIPPYNTEVTETTIVITWTPAPRIGFKLGVRPSQGGEAPREVTSDSGSIVVSGLTPGVEYTYTIQVLRDGQERDAPIVNRVVTPLSPPTNLHLEANPDTGVLTVSWERSTTPDITGYRITTTPTNGQ
QGTSLEEVVHADQSSCTFENLNPGLEYNVSVYTVKDDKESAPISDTVVPEVPQLTDLSFVDITDSSIGLRWTPLNSSTIIGYRITVVAAGEGIPIFEDFVDSSVGYYTVTGLEPGIDYDISVITLINGGESAPTTLTQQTAV
PPPTDLRFTNIGPDTMRVTWAPPPSIELTNLLVRYSPVKNEEDVAELSISPSDNAVVLTNLLPGTEYLVSVSSVYEQHESIPLRGRQKTGLDSPTGFDSSDITANSFTVHWVAPRAPITGYIIRHHAEHSVGRPRQDRV
PPSRNSITLTNLNPGTEYVVSIIAVNGREESPPLIGQQATVSDIPRDLEVIASTPTSLLISWEPPAVSVRYYRITYGETGGNSPVQEFTVPGSKSTATINNIKPGADYTITLYAVTGRGDSPASSKPVSINYKTEIDKPSQM
QVTDVQDNSISVRWLPSTSPVTGYRVTTTPKNGLGPSKTKTASPDQTEMTIEGLQPTVEYVVSVYAQNRNGESQPLVQTAVTNIDRPKGLAFTDVDVDSIKIAWESPQGQVSRYRVTYSSPEDGIRELFPAPDGED
DTAELQGLRPGSEYTVSVVALHDDMESQPLIGIQSTAIPAPTNLKFSQVTPTSFTAQWIAPSVQLTGYRVRVNPKEKTGPMKEINLSPDSSSVIVSGLMVATKYEVSVYALKDTLTSRPAQGVITTLENVSPPRRARV
TDATETTITISWRTKTETITGFQVDAIPANGQTPVQRSISPDVRSYTITGLQPGTDYKIHLYTLNDNARSSPVIIDASTAIDAPSNLRFLTTTPNSLLVSWQAPRARITGYIIKYEKPGSPPREVVPRPRPGVTEATITGLEP
GTEYTIYVIALKNNQKSEPLIGRKKTDELPQLVTLPHPNLHGPEILDVPSTVQKTPFITNPGYDTENGIQLPGTTHQQPSVGQQMIFEEHGFRRTTPPTAATPVRLRPRPYLPNVDEEVQIGHVPRGDVDYHLYPHVP
GLNPNASTGQEALSQTTISWTPFQESSEYIISCQPVGTDEEPLQFQVPGTSTSATLTGLTRGVTYNIIVEALQNQRRHKVREEVVTVGNAVSEGLNQPTDDSCFDPYTVSHYAIGEEWERLSDAGFKLTCQCLGFGS
GHFRCDSSKWCHDNGVNYKIGEKWDRQGENGQRMSCTCLGNGKGEFKCDPHEATCYDDGKTYHVGEQWQKEYLGAICSCTCFGGQRGWRCDNCRRPGAAEPSPDGTTGHTYNQYTQRYNQRTNTNVNCPI
ECFMPLDVQADRDDSRE

**Figure 5. 5 Transglutaminase cross-link positions in the sequences of the ECM proteins from the DNR samples.** The cross-linked lysine and glutamine residues are highlighted in red. The nonhelical regions of the collagen proteins are highlighted in yellow. The cross-linked glutamine residue found in fibronectin is located in the type-III 10 domain that is a part of the cell attachment region and the cross-linked lysine residue was found in the type-III 14 domain that is a part of the heparin binding domain (source: Uniprot). Both domains are highlighted in green.

## 5.3 Discussion

The initial aim of this chapter was to establish a proteomics workflow for the analysis of transglutaminase cross-linked sites in ECM proteins that would then be used for the qualitative comparison of control and fibrotic ECM extracts. In addition, we attempted to gather preliminary information about patterns of cross-linking site in different biological models as a response to pro-fibrotic stimulation. A major limitation of this study was the presence of multiple transglutaminase cross-links and other types of cross-links in the ECM that potentially interfered with the analysis of transglutaminase cross-linking analysis. As a result, after tryptic digestion of the ECM a fraction of the transglutaminase cross-linked peptides may have contained multiple different cross-links. These types of structures are not identifiable with the demonstrated method; therefore, this project cannot give a complete information about the positions of all transglutaminase cross-links. Nevertheless, we hypothesized that with the help of this method, progress in the discovery of unknown cross-linking sites could still be achieved.

Current information about the transglutaminase cross-linking sites in ECM proteins is limited to in-vitro studies [74, 180, 181]. Thus, due to several aspects in the respective experimental design in these studies, the biological significance of the corresponding discoveries is questionable. There are several reasons for this assertion. First, several transglutaminase cross-linking experiments have been performed on isolated proteins, for instance on collagen isolated from calf skin [74]. Although preliminary information about enzymatic preference for specific protein regions can be gathered in this way, the isolation procedure can lead to loss of binding partners and modification of the protein conformation that

can alter the relative rates of cross-linking at different sites in a subsequent transglutaminase cross-linking experiment. As a consequence, it is unclear whether or not the reported cross-linking sites are relevant in vivo. Second, these studies only address the question about the glutamine cross-linking site and lack any information about the location of the lysine donor. The cross-linking experiment is usually conducted using labelled amine donors such as fluorescent monodansylcadaverine or radioactively-labelled putrescine [171] to allow subsequent detection only of preferred glutamine cross-linking sites. Furthermore, the modified proteins are then subjected to proteolytic digestion followed by detection of the labelled peptides, not by identification of the modified amino acid residue [74]. Such approaches provide information about the regions where increased TG2 activity can be detected but not about the exact positions of the cross-linked sites. Other publications have used Edman degradation to determine the cross-linked glutamine site in a purified peptide [75]; however, this technique is not suitable for the analysis of complex protein samples. Moreover, the source of transglutaminase used in different cross-linking studies was often different, and typically either guinea-pig transglutaminase or microbial transglutaminase that might show differences in their preference for cross-linking sites. Finally, despite its far higher level of complexity compared to a purified protein, analysing whole ECM digests has the potential to give information about heterotypic partners. For these reasons, there is no reliable literature with which we could compare our discoveries in terms of cross-linking sites.

The protein database that we used to search for transglutaminase cross-links consisted of TG2 targets previously identified in vitro, including collagen I, III and V [44, 171, 172], fibronectin [135] and fibrinogen [173]. These previous studies

had demonstrated that fibronectin can be cross-linked to any of the above mentioned collagen types [178, 182], although the sites had not been identified. Hence the ECM cross-linking sites between fibronectin and collagen I and V that we identified using MassMatrix are consistent with earlier studies. A study demonstrating the formation of high molecular weight polymers of fibrinogen and type III procollagen by transglutaminase treatment [183] suggests that cross-links between fibrinogen and one or more collagens are also viable. However, there is as yet no evidence about the biological function of the cross-links between these proteins or their contribution to fibrotic progression.

Regarding the MassMatrix search of TGFβ-1-treated fibroblasts, it was inferred that the cross-link that was found between fibrinogen alpha and collagen alpha-3(V), and which was absent from the control samples, could be a new cross-link resulting from stimulation with the profibrotic cytokine TGFβ-1. However, the difference could alternatively have resulted from a difference in the amount of control and TGFβ-1-treated sample loaded onto the LC-MS system. This issue could not be resolved by loading equal peptide amounts on the LC-MS system due to the lack of a reliable method for peptide quantification at such low levels. The BCA assay showed insufficient sensitivity to detect the difference in the protein amount between the control and the fibrotic ECM digests (data not shown), although a possibility that the difference is negligibly small cannot be excluded. Absorbance measurements at 280 nm were unreliable as they showed a very high level of variation. Another way to determine the relative abundance of the cross-lined peptides is to normalise the signal intensity of the cross-linked peptides against the iBAQ values of the corresponding proteins. This would require further method development considering the number of

replicates required for a reliable quantification due to the high variation in mass spectrometry.

A major challenge in the analysis of cross-linked peptides is their low abundance in comparison with non-cross-linked peptides [77]. As a consequence, incomplete peptide isolation in MS1 leads to unidentified analytes in MS2. For this reason, chromatographic separation methods prior to LC-MS analysis are routinely exploited in chemical cross-linking experiments [161]. Previous isolation strategies for tryptic cross-linked peptide pairs have relied on the typically higher charge states and the larger size of these species in comparison with non-cross-linked peptides. Thus, separation has employed size-exclusion chromatography (SEC) or ion-exchange chromatography (IEC). However, these techniques would have lost the cross-linked peptides identified in this thesis, as they have short sequences and were often identified in their +2-charge state. These separation techniques could be used in the discovery of other transglutaminase cross-linked peptides with longer sequences and higher charge states that we were not able to identify with our method. However, tryptic cross-linked peptides with longer sequences are less likely to be identified by LC-MS and more likely to contain other cross-links that would not be identifiable using the available software.

In summary, we have demonstrated that transglutaminase cross-linking sites can be identified from complex ECM digests with the method that we developed. we were limited in our discovery of disease-specific cross-linking sites that may have been due to the relatively short exposure of our model to TGFβ-1 compared to the time for disease to progress from increased ECM to fibrosis. Cross-linking analysis of models with considerably longer exposure to pro-

fibrotic cytokines could lead to the discovery of additional cross-linking sites relevant to skin fibrosis.

The higher ECM complexity of skin tissues compared to model peptides and purified proteins probably caused an additional challenge in terms of peptide separation that could not be resolved with the current experimental workflow. This could explain the lower number of significant hits for the TSK2-mouse model compared to the cell-culture model. Therefore, further optimisation of peptide separation might be necessary for future studies of transglutaminase cross-linking sites in skin tissue. However, for the investigation of new disease-related cross-linking sites, a mouse model that shows a significant increase of the ECM transglutaminase cross-linking activity must be found, or better still, future studies should use the methods developed here to address cross-linking in patient skin samples.

# Chapter 6

# Discussion

As already stated in the introduction, the main objective of this thesis was to discover alterations of the ECM cross-linking pattern of skin fibrosis in SSc compared to control skin ECM that could be targets for new anti-fibrotic therapies. To achieve this goal, we focused our research on two types of cross-links that, according to previous literature, had shown increased levels in SSc patients, namely hydroxylysylpyridinoline (HP) and transglutaminase cross-links [40, 73]. These reported increases indicated a high probability that both types of cross-link would have an altered pattern in the fibrotic ECM. Prior to performing a comparative study of the cross-linking pattern between control and fibrotic ECM, a method providing reliable identification of cross-linking sites had to be developed. Due to the exploratory nature of this study, an automated method to screen for cross-linking sites in the large MS datasets that are typically generated from proteomics experiments had to be developed.

Regarding the method development, our objective was partially achieved, as an automated way to analyse transglutaminase cross-linked sites was developed, although with certain limitations in terms of the analyte charge state. The initial plan for the identification of transglutaminase cross-linked peptides was to find a characteristic feature (reporter ion) in the MS2 fragmentation spectrum of the tryptic peptide model that would allow me to select all the MS2 spectra belonging to transglutaminase cross-linked peptides from large proteomics datasets. In contrast to our hypothesis, we discovered that transglutaminase cross-linked peptides obeyed the common fragmentation rule of preferential cleavage of the peptide bond resulting in series of cross-linked and non-cross-linked peptides.

One characteristic feature of the fragmentation spectrum of the cross-linked peptides was a fragment with a 9-unit shift from the predicted m/z value of the precursor that appeared in every spectrum regardless of the charge state. As we could not build a hypothesis to explain the structure of this fragment, we did not undertake any further analysis of this fragment.

After having found a programme that was able to identify transglutaminase cross-linked peptides, another important aspect of the method development was to maximise the score that could be achieved. Several different properties of a peptide determine which fragment ions are formed from collision induced dissociation [139]. For instance, the presence of basic amino acid residues has a strong influence on the fragmentation efficiency and their distribution in the peptide sequence dictates the formation of certain fragments [184]. A transglutaminase cross-link modifies a lysine residue by forming a secondary amine at its ε-amino group. This modification reduces the basicity of lysine at this site. Thus, one can expect that the fragmentation behaviour of the cross-linked peptide will be different from the fragmentation behaviour of a non-cross-linked peptide with the same sequence. As a consequence, one cannot expect that a cross-linked peptide would have the same score as its non-cross-linked equivalent. For this reason it was important to analyse the score that could be achieved with defined peptides containing a transglutaminase-mediated cross-link. This information was then used to set a threshold with which to assess the confidence of novel cross-linked peptide identifications. In future, method optimization should include the analysis of more model peptides with different sequences and higher charge states that would allow a more complete understanding of the fragmentation behaviour of these species. This would ensure higher confidence in the identification of novel transglutaminase cross-

linked peptides. For instance, as we found that the pp value was strongly dependent on peptide length it would be useful to analyse how this parameter changes with a wider range of sequence lengths. A potential drawback of this approach would be that the pp value also depended on amino acid sequence. This latter effect could be minimized by analysing sets of peptides with single amino acid repeats outside the cross-linked sequence that would lead to a better prediction of the typical pp value range obtained from transglutaminase cross-linked peptides.

As mentioned previously, while the proposed method identified transglutaminase cross-linked peptides, it could not give a full definition of all cross-linked positions in the ECM, due to the presence of other types of cross-links. Consequently, a different strategy that could help identify more cross-linking sites would be to use a proteinase different from trypsin. Other enzymes that have been used for proteomics experiments include Glu-C and Asp-N. Glu-C is a serine proteinase that cleaves with high specificity at the C-terminus of aspartic and glutamic acids, while Asp-N is a metalloproteinase that preferentially cleaves at the N-terminus of aspartic and glutamic acid residues [185]. These techniques could help identify sequences that might not be identified from tryptic digests, for example due to the formation of very long peptides. A particular use could be in enabling the analysis of collagen helical regions which have very low densities of lysine and arginine residues. Digestion with the above mentioned proteases would generate peptides with shorter sequences. This method could be particularly useful for the analysis of transglutaminase cross-linked sites that are not identifiable from tryptic digests using the method we developed. This is the case when tryptic peptides containing not only transglutaminase cross-links but also other types of cross-

links are formed. In such cases, digestion with an alternative proteinase could lead to the formation of a dipeptide with only one transglutaminase cross-link that could be identified using the available software. This method would require the analysis of new peptide models, as the fragmentation behaviour of the analytes is likely to differ from that of tryptic peptides mainly due to the absence of a C-terminal basic residue. A major problem in this procedure would be the absence of a basic C-terminal residue what would deteriorate the fragmentation behaviour of the analyte. One possible solution would be to introduce positive charges via derivatisation of the carbonyl groups. Another approach suggested by Mendes et al [186] would be to perform a sequential digestion. In this procedure, proteins are first digested with trypsin followed by a shortening of the long tryptic peptides by performing a second digestion with other enzymes. However, this paper also demonstrates the importance of retaining a basic C-terminus for the optimal analysis of the cross-linked peptides.

As an alternative to a mass spectrometry based method, one can use cryo-electron microscopy (cryo-EM) to identify cross-linking positions in the ECM. This technique is used to determine the structure of proteins at near-atomic resolution without the need for sample crystallization [187]. Thus, cryo-EM could allow us to visualise the ECM cross-linking sites and serve as a validation method. However, as this technique depends on the computational averaging of thousands of images of identical particles, obtaining an image of particles with variable conformation and composition (heterogeneity) can be challenging [187]. This limitation could cause a problem in the analysis of ECM samples, as not all protein sites are cross-linked.

After having developed a method for the identification of transglutaminase cross-linked sites, it was necessary to ensure its suitability for the analysis of complex

protein digests prior to analysing ECM matrisomes. A major problem for the development of an efficient protocol remains the low abundance of cross-linked peptides in comparison with the rest of the protein digest. This issue was the main reason for the limited number of cross-linking sites that we managed to find in our fibrotic models. As mentioned in the previous chapter, other studies used different chromatographic purification techniques to isolate cross-linked peptides; however, this would most likely have led to a loss of cross-linked peptides with low molecular weight and charge. A more selective approach would be to perform affinity chromatography using an antibody to the transglutaminase cross-link [188]. However, based on the properties of this antibody, a main challenge during this purification step would be some loss of the low abundance target peptides, as non-specific binding of the antibody to other peptides may limit the specific interactions with the cross-link. The antibody has shown good results in immunohistochemistry studies [188] but it has not yet been tested for affinity chromatography.

Although using a fibrotic model was essential to developing the method, finding a suitable model that would allow me to compare differences between control and fibrotic matrisomes proved to be challenging. Chronic treatment of skin fibroblasts with TGFβ-1 showed the desired response of increased TG2 expression in the ECM; however, the exposure time with the fibrotic agent was apparently too short to detect multiple changes in the cross-linking pattern, as we found only one cross-link site in the TGFβ-1 treated cells that was absent from the control. We expect this model to prove unsuitable for the investigation of other cross-link types, as mature cross-links such as pyridinoline cross-links are a product of reaction cascades and their formation is likely to require extended times. Genetic mouse models and controls could show significant

differences in their ECM cross-linking pattern, as fibrogenesis can be extended to several months.

While mouse models may prove suitable for future site-mapping studies, there are several reasons why they are not suitable for the investigation of potential anti-fibrotic therapies. First, ECM protein sequence variation between different organisms must be taken into account. Accordingly, unless an identified cross-linking site is located in a region whose sequence is conserved in the orthologous human protein, it may be irrelevant to human disease. As all the identified cross-linking sites in the TSK2 mouse model were on regions identical with the human sequence, it is likely that they are also present in the human organism. Second, despite the longer time of exposure to fibrotic agents that can be achieved in mouse models in comparison with cell cultures, this time is still much shorter than the period of fibrotic development in SSc, which typically spans several years [20]. This difference can be particularly important, as a higher accumulation of cross-links and/or the appearance of specific cross-linking sites that cause the irreversible character of fibrosis might appear at later disease stages. Mouse models can mimic only early fibrotic stages, which implies that the cross-linking positions found in early fibrosis may not play a role in impaired ECM remodelling in irresolvable stages of fibrosis. The limited amount of time for which a mouse model experiment is conducted could mean that changes in the cross-linking pattern observed in these models are a result of normal wound-healing processes that are irrelevant to the progression of fibrosis and scarring. Thus, the existence of disease-relevant cross-linking sites can be proven only in a comparative study involving skin biopsies from SSc patients. Nevertheless, mouse models are useful for the development of a

method for the analysis of cross-links from tissue samples, as they provide a similarly high level of complexity to human skin tissue.

Moreover, a longitudinal analysis showing how the ECM cross-linking pattern changes during the progression of skin fibrosis could help identify targets for new disease-modifying therapies. Such studies could include the comparison of cross-linking sites between different healthy tissues, in normal wound healing, aging and different diseases in order to identify the sites relevant to fibrosis. This could be achieved if evidence were found that the accumulation of specific cross-linking sites in the course of disease progression inhibits ECM remodelling. As already stated in the Introduction, it is possible that the formation of a cross-link at a given protein site could lead to a conformational change that masks one or more protease recognition sites and thereby prevents proteolytic degradation. An essential part of future research will be to reveal the exact molecular mechanism by which each such cross-linking site prevents protein degradation by MMPs.

## References


1. Mehal, W.Z., J. Iredale, and S.L. Friedman, *Scraping fibrosis: expressway to the core of fibrosis.* Nat Med, 2011. **17**(5): p. 552-3.
2. Wynn, T.A., *Cellular and molecular mechanisms of fibrosis.* J Pathol, 2008. **214**(2): p. 199-210.
3. Rosenbloom, J., et al., *Human Fibrotic Diseases: Current Challenges in Fibrosis Research.* Methods Mol Biol, 2017. **1627**: p. 1-23.
4. Rosenbloom, J., F.A. Mendoza, and S.A. Jimenez, *Strategies for anti-fibrotic therapies.* Biochim Biophys Acta, 2013. **1832**(7): p. 1088-103.
5. Rockey, D.C., P.D. Bell, and J.A. Hill, *Fibrosis--a common pathway to organ injury and failure.* N Engl J Med, 2015. **372**(12): p. 1138-49.
6. Zeisberg, M. and R. Kalluri, *Cellular Mechanisms of Tissue Fibrosis. 1. Common and organ-specific mechanisms associated with tissue fibrosis.* American Journal of Physiology-Cell Physiology, 2013. **304**(3): p. C216-C225.
7. Varga, J. and D. Abraham, *Systemic sclerosis: a prototypic multisystem fibrotic disorder.* J Clin Invest, 2007. **117**(3): p. 557-67.
8. Denton, C.P., C.M. Black, and D.J. Abraham, *Mechanisms and consequences of fibrosis in systemic sclerosis.* Nat Clin Pract Rheumatol, 2006. **2**(3): p. 134-44.
9. Eckes, B., et al., *Molecular and cellular basis of scleroderma.* J Mol Med (Berl), 2014. **92**(9): p. 913-24.
10. Gabrielli, A., E.V. Avvedimento, and T. Krieg, *Scleroderma.* New England Journal of Medicine, 2009. **360**(19): p. 1989-2003.
11. Ferreli, C., et al., *Cutaneous Manifestations of Scleroderma and Scleroderma-Like Disorders: a Comprehensive Review.* Clin Rev Allergy Immunol, 2017. **53**(3): p. 306-336.
12. Nietert, P.J. and R.M. Silver, *Systemic sclerosis: environmental and occupational risk factors.* Current Opinion in Rheumatology, 2000. **12**(6): p. 520-526.
13. Pattanaik, D., et al., *Pathogenesis of Systemic Sclerosis.* Front Immunol, 2015. **6**: p. 272.
14. Radstake, T.R., et al., *Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus.* Nat Genet, 2010. **42**(5): p. 426-9.
15. Matucci-Cerinic, M., B. Kahaleh, and F.M. Wigley, *Review: evidence that systemic sclerosis is a vascular disease.* Arthritis Rheum, 2013. **65**(8): p. 1953-62.
16. Kisseleva, T. and D.A. Brenner, *Mechanisms of fibrogenesis.* Exp Biol Med (Maywood), 2008. **233**(2): p. 109-22.
17. Burbelo, P.D., et al., *Autoantibodies are present before the clinical diagnosis of systemic sclerosis.* PLoS One, 2019. **14**(3): p. e0214202.
18. Worda, M., et al., *In vivo analysis of the apoptosis-inducing effect of anti-endothelial cell antibodies in systemic sclerosis by the chorionallantoic membrane assay.* Arthritis Rheum, 2003. **48**(9): p. 2605-14.
19. Seibold, J.R., *Clinical trials: types, design, and end points.* Current Opinion in Rheumatology, 2011. **13**(6): p. 512-515.
20. Wu, W., et al., *Progressive skin fibrosis is associated with a decline in lung function and worse survival in patients with diffuse cutaneous systemic sclerosis in the European Scleroderma Trials and Research (EUSTAR) cohort.* Ann Rheum Dis, 2019. **78**(5): p. 648-656.

21. White, B.e.a., *Guidelines for clinical trials in systemic sclerosis (scleroderma). I. Disease-modifying interventions. The American College of Rheumatology Committee on Design and Outcomes in Clinical Trials in Systemic Sclerosis.* Arthritis Rheum., 1995. **38**(3): p. 351-60.
22. Denton, C.P. and C.M. Black, *Scleroderma--clinical and pathological advances.* Best Pract Res Clin Rheumatol, 2004. **18**(3): p. 271-90.
23. Perlish, J.S., G. Lemlich, and R. Fleischmajer, *Identification of Collagen Fibrils in Scleroderma Skin.* Journal of Investigative Dermatology, 1988. **90**(1): p. 48-54.
24. Svegliati, S., et al., *Platelet-derived growth factor and reactive oxygen species (ROS) regulate Ras protein levels in primary human fibroblasts via ERK1/2. Amplification of ROS and Ras in systemic sclerosis fibroblasts.* J Biol Chem, 2005. **280**(43): p. 36474-82.
25. Gabrielli, A., E.V. Avvedimento, and T. Krieg, *Scleroderma.* N Engl J Med, 2009. **360**(19): p. 1989-2003.
26. Kawakami, T., et al., *Increased expression of TGF-beta receptors by scleroderma fibroblasts: evidence for contribution of autocrine TGF-beta signaling to scleroderma phenotype.* J Invest Dermatol, 1998. **110**(1): p. 47-51.
27. Rajkumar, V.S., et al., *Shared expression of phenotypic markers in systemic sclerosis indicates a convergence of pericytes and fibroblasts to a myofibroblast lineage in fibrosis.* Arthritis Res Ther, 2005. **7**(5): p. R1113-23.
28. Hinz, B., et al., *The myofibroblast: one function, multiple origins.* Am J Pathol, 2007. **170**(6): p. 1807-16.
29. Bonnans, C., J. Chou, and Z. Werb, *Remodelling the extracellular matrix in development and disease.* Nat Rev Mol Cell Biol, 2014. **15**(12): p. 786-801.
30. Frantz, C., K.M. Stewart, and V.M. Weaver, *The extracellular matrix at a glance.* J Cell Sci, 2010. **123**(Pt 24): p. 4195-200.
31. Zhen, G. and X. Cao, *Targeting TGFbeta signaling in subchondral bone and articular cartilage homeostasis.* Trends Pharmacol Sci, 2014. **35**(5): p. 227-36.
32. Chang, C. and Z. Werb, *The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis.* Trends Cell Biol., 2001. **11**(11): p. 37-43.
33. Jablonska-Trypuc, A., M. Matejczyk, and S. Rosochacki, *Matrix metalloproteinases (MMPs), the main extracellular matrix (ECM) enzymes in collagen degradation, as a target for anticancer drugs.* J Enzyme Inhib Med Chem, 2016. **31**(sup1): p. 177-183.
34. Chung, L., et al., *Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis.* EMBO J, 2004. **23**(15): p. 3020-30.
35. Lu, P., et al., *Extracellular matrix degradation and remodeling in development and disease.* Cold Spring Harb Perspect Biol, 2011. **3**(12).
36. Sato, S., et al., *Function blocking autoantibodies against matrix metalloproteinase-1 in patients with systemic sclerosis.* J Invest Dermatol, 2003. **120**(4): p. 542-7.
37. Nishijima, C., et al., *Autoantibody against matrix metalloproteinase-3 in patients with systemic sclerosis.* Clin Exp Immunol, 2004. **138**(2): p. 357-63.
38. Kirk, T.Z., et al., *Myofibroblasts from scleroderma skin synthesize elevated levels of collagen and tissue inhibitor of metalloproteinase*

*(TIMP-1) with two forms of TIMP-1.* J Biol Chem., 1995. **270**(7): p. 3423-8.
39. Mattila, L., et al., *Activation of tissue inhibitor of metalloproteinases-3 (TIMP-3) mRNA expression in scleroderma skin fibroblasts.* J Invest Dermatol, 1998. **110**(4): p. 416-21.
40. Brinckmann, J., et al., *Interleukin 4 and prolonged hypoxia induce a higher gene expression of lysyl hydroxylase 2 and an altered cross-link pattern: important pathogenetic steps in early and late stage of systemic scleroderma?* Matrix Biol, 2005. **24**(7): p. 459-68.
41. Philp, C.J., et al., *Extracellular Matrix Cross-Linking Enhances Fibroblast Growth and Protects against Matrix Proteolysis in Lung Fibrosis.* Am J Respir Cell Mol Biol, 2018. **58**(5): p. 594-603.
42. Klingberg, F., B. Hinz, and E.S. White, *The myofibroblast matrix: implications for tissue repair and fibrosis.* J Pathol, 2013. **229**(2): p. 298-309.
43. Olsen, A.L., et al., *Hepatic stellate cells require a stiff environment for myofibroblastic differentiation.* Am J Physiol Gastrointest Liver Physiol, 2011. **301**(1): p. G110-8.
44. Fortunati, D., et al., *Cross-linking of collagen I by tissue transglutaminase provides a promising biomaterial for promoting bone healing.* Amino Acids, 2014. **46**(7): p. 1751-61.
45. Robins, S.P., *Biochemistry and functional significance of collagen cross-linking.* Biochem Soc Trans., 2007. **35**(5): p. 849-52.
46. Rodriguez-Pascual, F. and D.A. Slatter, *Collagen cross-linking: insights on the evolution of metazoan extracellular matrix.* Sci Rep, 2016. **6**: p. 37374.
47. Eyre, D.R., M.A. Weis, and J.J. Wu, *Advances in collagen cross-link analysis.* Methods, 2008. **45**(1): p. 65-74.
48. Eyre, D.R., M.A. Paz, and P.M. Gallop, *Cross-linking of Collagen and Elastin.* Ann. Rev. Biochem., 1984. **53**: p. 717-48.
49. van der Slot, A.J., et al., *Increased formation of pyridinoline cross-links due to higher telopeptide lysyl hydroxylase levels is a general fibrotic phenomenon.* Matrix Biol, 2004. **23**(4): p. 251-7.
50. van der Slot, A.J., et al., *Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis.* J Biol Chem, 2003. **278**(42): p. 40967-72.
51. Istok, R., et al., *Evidence for increased pyridinoline concentration in fibrotic tissues in diffuse systemic sclerosis.* Clin Exp Dermatol., 2001. **26**(6): p. 545-7.
52. van der Slot-Verhoeven, A.J., et al., *The type of collagen cross-link determines the reversibility of experimental skin fibrosis.* Biochim Biophys Acta, 2005. **1740**(1): p. 60-7.
53. Rimar, D., et al., *Brief report: lysyl oxidase is a potential biomarker of fibrosis in systemic sclerosis.* Arthritis Rheumatol, 2014. **66**(3): p. 726-30.
54. Vadasz, Z., et al., *Lysyl oxidase-a possible role in systemic sclerosis-associated pulmonary hypertension: a multicentre study.* Rheumatology (Oxford), 2019.
55. Bellaye, P.S., et al., *Lysyl Oxidase-Like 1 Protein Deficiency Protects Mice from Adenoviral Transforming Growth Factor-beta1-induced Pulmonary Fibrosis.* Am J Respir Cell Mol Biol, 2018. **58**(4): p. 461-470.

56. Barry-Hamilton, V., et al., *Allosteric inhibition of lysyl oxidase-like-2 impedes the development of a pathologic microenvironment.* Nat Med, 2010. **16**(9): p. 1009-17.
57. Wipff, P.J., et al., *Myofibroblast contraction activates latent TGF-beta1 from the extracellular matrix.* J Cell Biol, 2007. **179**(6): p. 1311-23.
58. Raghu, G., et al., *Efficacy of simtuzumab versus placebo in patients with idiopathic pulmonary fibrosis: a randomised, double-blind, controlled, phase 2 trial.* The Lancet Respiratory Medicine, 2017. **5**(1): p. 22-32.
59. Harrison, S.A., et al., *Simtuzumab Is Ineffective for Patients With Bridging Fibrosis or Compensated Cirrhosis Caused by Nonalcoholic Steatohepatitis.* Gastroenterology, 2018. **155**(4): p. 1140-1153.
60. Cox, T.R., et al., *The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase.* Nature, 2015. **522**(7554): p. 106-110.
61. Lorand, L. and R.M. Graham, *Transglutaminases: crosslinking enzymes with pleiotropic functions.* Nat Rev Mol Cell Biol, 2003. **4**(2): p. 140-56.
62. Davidenko, N., et al., *Control of crosslinking for tailoring collagen-based scaffolds stability and mechanics.* Acta Biomater, 2015. **25**: p. 131-42.
63. Mishra, S. and L.J. Murphy, *Tissue transglutaminase has intrinsic kinase activity: identification of transglutaminase 2 as an insulin-like growth factor-binding protein-3 kinase.* J Biol Chem, 2004. **279**(23): p. 23863-8.
64. Olsen, K.C., et al., *Transglutaminase 2 and its role in pulmonary fibrosis.* Am J Respir Crit Care Med, 2011. **184**(6): p. 699-707.
65. Wang, Z., et al., *Cardiac fibrosis can be attenuated by blocking the activity of transglutaminase 2 using a selective small-molecule inhibitor.* Cell Death Dis, 2018. **9**(6): p. 613.
66. Grenard, P., et al., *Transglutaminase-mediated cross-linking is involved in the stabilization of extracellular matrix in human liver fibrosis.* J Hepatol. , 2001. **35**(3): p. 367-75.
67. Siegel, M., et al., *Extracellular transglutaminase 2 is catalytically inactive, but is transiently activated upon tissue injury.* PLoS One, 2008. **3**(3): p. e1861.
68. Johnson, T.S., et al., *Transglutaminase transcription and antigen translocation in experimental renal scarring.* J Am Soc Nephrol., 1999. **10**(10): p. 2146-57.
69. Nunes, I., et al., *Latent transforming growth factor-beta binding protein domains involved in activation and transglutaminase-dependent cross-linking of latent transforming growth factor-beta.* J Cell Biol, 1997. **136**(5): p. 1151-63.
70. Verderio, E., et al., *Regulation of Cell Surface Tissue Transglutaminase:*

*Effects on Matrix Storage of Latent Transforming*

*Growth Factor-beta*

*Binding Protein-1.* The Journal of Histochemistry & Cytochemistry, 1999. **47**(11): p. 1417-32.
71. Gross, S.R., Z. Balklava, and M. Griffin, *Importance of tissue transglutaminase in repair of extracellular matrices and cell death of dermal fibroblasts after exposure to a solarium ultraviolet A source.* J Invest Dermatol, 2003. **121**(2): p. 412-23.
72. Mantero, J.C. and R. Lafyatis, *Expression data from skin biopsies in patients with systemic sclerosis.* Platform: GPL23080 2018.
73. Quan, G., et al., *TGF-beta1 up-regulates transglutaminase two and fibronectin in dermal fibroblasts: a possible mechanism for the*

*stabilization of tissue inflammation.* Arch Dermatol Res, 2005. **297**(2): p. 84-90.
74. Stachel, I., et al., *Cross-Linking of Type I Collagen with Microbial Transglutaminase: Identification of Cross-Linking Sites.* Biomacromolecules, 2010. **11**: p. 698–705.
75. Purves, L., M. Purves, and W. Brandt, *Cleavage of Fibrin-Derived D-Dimer into Monomers by Endopeptidase from Puff Adder Venom (Bitis arietans) Acting at Cross-Linked Sites of the gamma-Chain. Sequence of Carboxy-Terminal Cyanogen Bromide gamma-Chain Fragments.* Biochemistry, 1987. **26**(15): p. 4640-4646.
76. Sobel, J.H. and M.A. Gawinowicz, *Identification of the alpha Chain Lysine Donor Sites Involved in Factor XIIIa Fibrin Cross-linking.* THE JOURNAL OF BIOLOGICAL CHEMISTRY, 1996. **271**(32): p. 19288–19297.
77. Rappsilber, J., *The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes.* J Struct Biol, 2011. **173**(3): p. 530-40.
78. Recktenwald, C.V. and G.C. Hansson, *The Reduction-insensitive Bonds of the MUC2 Mucin Are Isopeptide Bonds.* J Biol Chem, 2016. **291**(26): p. 13580-90.
79. Fleckenstein, B., et al., *Molecular characterization of covalent complexes between tissue transglutaminase and gliadin peptides.* J Biol Chem, 2004. **279**(17): p. 17607-16.
80. Velez, V.F., et al., *Tissue transglutaminase is a negative regulator of monomeric lacritin bioactivity.* Invest Ophthalmol Vis Sci, 2013. **54**(3): p. 2123-32.
81. Hanson, D.A. and D.R. Eyre, *Molecular Site Specificity of Pyridinoline and Pyrrole Cross-links in Type I Collagen of Human Bone.* The Journal of Biological Chemistry, 1996. **271**(43): p. 26508–26516.
82. Kalli, A., et al., *Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers.* J Proteome Res, 2013. **12**(7): p. 3071-86.
83. Yates, J.R., et al., *Performance of a linear ion trap-Orbitrap hybrid for peptide analysis.* Anal Chem., 2006. **78**(2): p. 493-500.
84. Asano, Y. and S. Sato, *Animal models of scleroderma: current state and recent development.* Curr Rheumatol Rep, 2013. **15**(12): p. 382.
85. Avouac, J., M. Elhai, and Y. Allanore, *Experimental models of dermal fibrosis and systemic sclerosis.* Joint Bone Spine, 2013. **80**(1): p. 23-8.
86. De Langhe, E. and R. Lories, *Fibrogenesis, novel lessons from animal models.* Semin Immunopathol, 2015. **37**(5): p. 565-74.
87. Yamamoto, T., et al., *Animal model of sclerotic skin. I: Local injections of bleomycin induce sclerotic skin mimicking scleroderma.* J Invest Dermatol, 1999. **112**(4): p. 456-62.
88. Batteux, F., N. Kavian, and A. Servettaz, *New insights on chemically induced animal models of systemic sclerosis.* Curr Opin Rheumatol, 2011. **23**(6): p. 511-8.
89. Yoshizaki, A., et al., *Immunization with DNA topoisomerase I and Freund's complete adjuvant induces skin and lung fibrosis and autoimmunity via interleukin-6 signaling.* Arthritis Rheum, 2011. **63**(11): p. 3575-85.
90. Siracusa, L.D.e.a., *A Tandem Duplication within the fibrillin 1*

*Gene Is Associated with the Mouse Tight*

*skin Mutation.* Genome Res, 1996. **6**: p. 300-313.
91. Christner, P.J., et al., *The Tight Skin 2 mouse.* Arthritis & Rheumatism, 1996. **38**(12): p. 1791-1798.
92. Reich, N., et al., *The transcription factor Fra-2 regulates the production of extracellular matrix in systemic sclerosis.* Arthritis Rheum, 2010. **62**(1): p. 280-90.
93. Long, K.B., et al., *The Tsk2/+ mouse fibrotic phenotype is due to a gain-of-function mutation in the PIIINP segment of the Col3a1 gene.* J Invest Dermatol, 2015. **135**(3): p. 718-27.
94. Bayle, J., et al., *Increased expression of Wnt2 and SFRP4 in Tsk mouse skin: role of Wnt signaling in altered dermal fibrillin deposition and systemic sclerosis.* J Invest Dermatol, 2008. **128**(4): p. 871-81.
95. Long, K.B., C.M. Artlett, and E.P. Blankenhorn, *Tight skin 2 mice exhibit a novel time line of events leading to increased extracellular matrix deposition and dermal fibrosis.* Matrix Biol, 2014. **38**: p. 91-100.
96. Baxter, R.M., et al., *Analysis of the tight skin (Tsk1/+) mouse as a model for testing antifibrotic agents.* Lab Invest, 2005. **85**(10): p. 1199-209.
97. Avouac, J., et al., *Inhibition of activator protein 1 signaling abrogates transforming growth factor beta-mediated activation of fibroblasts and prevents experimental fibrosis.* Arthritis Rheum, 2012. **64**(5): p. 1642-52.
98. Denton, C.P., et al., *Fibroblast-specific expression of a kinase-deficient type II transforming growth factor beta (TGFbeta) receptor leads to paradoxical activation of TGFbeta signaling pathways with fibrosis in transgenic mice.* J Biol Chem, 2003. **278**(27): p. 25109-19.
99. Denton, C.P., *Advances in pathogenesis and treatment of systemic sclerosis.* Clinical Medicine, 2015. **15**(6): p. 58-63.
100. Distler, J.H., et al., *Review: Frontiers of Antifibrotic Therapy in Systemic Sclerosis.* Arthritis Rheumatol, 2017. **69**(2): p. 257-267.
101. Wynn, T.A. and T.R. Ramalingam, *Mechanisms of fibrosis: therapeutic translation for fibrotic disease.* Nat Med, 2012. **18**(7): p. 1028-40.
102. Distler, O. and J.H.W. Distler, *Tocilizumab for systemic sclerosis: implications for future trials.* The Lancet, 2016. **387**(10038): p. 2580-2581.
103. Murray, L.A., et al., *Targeting interleukin-13 with tralokinumab attenuates lung fibrosis and epithelial damage in a humanized SCID idiopathic pulmonary fibrosis model.* Am J Respir Cell Mol Biol, 2014. **50**(5): p. 985-94.
104. Saito, E., et al., *CD19-dependent B lymphocyte signaling thresholds influence skin fibrosis and autoimmunity in the tight-skin mouse.* Journal of Clinical Investigation, 2002. **109**(11): p. 1453-1462.
105. Salles, G., et al., *Rituximab in B-Cell Hematologic Malignancies: A Review of 20 Years of Clinical Experience.* Adv Ther, 2017. **34**(10): p. 2232-2273.
106. Giuggioli, D., et al., *Rituximab in the treatment of patients with systemic sclerosis. Our experience and review of the literature.* Autoimmun Rev, 2015. **14**(11): p. 1072-8.
107. Melissaropoulos, K., et al., *Targeting very early systemic sclerosis: a case-based review.* Rheumatol Int, 2019.
108. Bellando-Randone, S. and M. Matucci-Cerinic, *Very Early Systemic Sclerosis and Pre-systemic Sclerosis: Definition, Recognition, Clinical Relevance and Future Directions.* Curr Rheumatol Rep, 2017. **19**(10): p. 65.

109. Scotton, C.J. and R.C. Chambers, *Molecular targets in pulmonary fibrosis: the myofibroblast in focus.* Chest, 2007. **132**(4): p. 1311-21.
110. Iyer, S.N., G. Gurujeyalakshmi, and S.N. Giri, *Effects of Pirfenidone on Transforming Growth Factor-β Gene Expression at the Transcriptional Level in Bleomycin Hamster Model of Lung Fibrosis.* J Pharmacol Exp Ther. , 1999. **291**(1): p. 367-73.
111. Marshall, R.P., et al., *Angiotensin II and the fibroproliferative response to acute lung injury.* Am J Physiol Lung Cell Mol Physiol., 2004. **286**(156-64).
112. Arribillaga, L., et al., *Therapeutic effect of a peptide inhibitor of TGF-beta on pulmonary fibrosis.* Cytokine, 2011. **53**(3): p. 327-33.
113. Wehr, A., et al., *Pharmacological inhibition of the chemokine CXCL16 diminishes liver macrophage infiltration and steatohepatitis in chronic hepatic injury.* PLoS One, 2014. **9**(11): p. e112327.
114. Dees, C., et al., *Platelet-derived serotonin links vascular disease and tissue fibrosis.* J Exp Med, 2011. **208**(5): p. 961-72.
115. Kitaba, S., et al., *Blockade of interleukin-6 receptor alleviates disease in mouse model of scleroderma.* Am J Pathol, 2012. **180**(1): p. 165-76.
116. Hecker, L., et al., *Reversal of persistent fibrosis in aging by targeting Nox4-Nrf2 redox imbalance.* Sci Transl Med, 2014. **6**(231): p. 231ra47.
117. Beyer, C., C. Dees, and J.H. Distler, *Morphogen pathways as molecular targets for the treatment of fibrosis in systemic sclerosis.* Arch Dermatol Res, 2013. **305**(1): p. 1-8.
118. Dees, C., et al., *The Wnt antagonists DKK1 and SFRP1 are downregulated by promoter hypermethylation in systemic sclerosis.* Ann Rheum Dis, 2014. **73**(6): p. 1232-9.
119. Hilberg, F., et al., *BIBF 1120: triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy.* Cancer Res, 2008. **68**(12): p. 4774-82.
120. Huang, J.e.a., *Nintedanib inhibits fibroblast activation and ameliorates fibrosis in preclinical models of systemic sclerosis.* Ann. Rheum. Dis., 2015. **75**(5): p. 883-90.
121. Gerber, E.E., et al., *Integrin-modulating therapy prevents fibrosis and autoimmunity in mouse models of scleroderma.* Nature, 2013. **503**(7474): p. 126-30.
122. Reed, N.I., et al., *The alphavbeta1 integrin plays a critical in vivo role in tissue fibrosis.* Sci Transl Med, 2015. **7**(288): p. 288ra79.
123. Xu, H. and M.A. Freitas, *A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data.* BMC Bioinformatics, 2007. **8**: p. 133.
124. Gotze, M., et al., *StavroX--a software for analyzing crosslinked products in protein interaction studies.* J Am Soc Mass Spectrom, 2012. **23**(1): p. 76-87.
125. Laemmli, U.K., *Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4.* Nature, 1970. **227**: p. 680-685.
126. Fang, J., et al., *False EX1 signatures caused by sample carryover during HX MS analyses.* Int J Mass Spectrom, 2011. **302**(1-3): p. 19-25.
127. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.
128. Abramoff, M.D., P.J. Magalhaes, and S.J. Ram, *Image Processing with ImageJ.* Biophotonics International, 2004. **11**(7): p. 36-42.
129. Malesevic, M., et al., *A fluorescence-based array screen for transglutaminase substrates.* Chembiochem, 2015. **16**(8): p. 1169-74.

130. S.P., R., et al., *Standardization of pyridinium crosslinks, pyridinoline and deoxypyridinoline, for use as biochemical markers of collagen degradation.* Clinical Chemisty, 1996. **42**(10): p. 1621-1626.
131. Hermanson, G.T., *Bioconjugate Techniques*. 1 ed, ed. A. Press. 1996.
132. Clark, B.R., J. Datillo, and D. Pearson, *Chemical synthesis of urotensin 11, a somatostatin-like peptide in the caudal neurosecretory system of fishes.* Int. J. Peptide Botein Rex, 1982. **19**: p. 448-453.
133. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat Methods, 2007. **4**(9): p. 709-12.
134. Xu, H., et al., *Database Search Algorithm for Identification of Intact Cross-Links in Proteins and Peptides Using Tandem Mass Spectrometry.* Journal of Proteome Research, 2010. **9**: p. 3384–3393.
135. Forsprecher, J., et al., *Enhanced osteoblast adhesion on transglutaminase 2-crosslinked fibronectin.* Amino Acids, 2009. **36**(4): p. 747-53.
136. Shainoff, J.R., D.A. Urbanic, and P.M. DiBello, *Immunoelectrophoretic Characterizations of the Cross-linking of Fibrinogen and Fibrin by Factor XIIIa and Tissue Transglutaminase.* THE JOURNAL OF BIOLOGICAL CHEMISTRY, 1991. **266**(10): p. 6429-6437.
137. Murthy, S.N.P. and L. Lorand, *Cross-linked Aalpha.gamma chain hybrids serve as unique markers for fibrinogen polymerized by tissue transglutaminase.* Proc. Natl. Acad. Sci., 1990. **87**: p. 9679-9682.
138. Stieger, C.E., P. Doppler, and K. Mechtler, *Optimized Fragmentation Improves the Identification of Peptides Cross-Linked by MS-Cleavable Reagents.* J Proteome Res, 2019. **18**(3): p. 1363-1370.
139. Dongre´, A.R., et al., *Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model.* J. Am. Chem. Soc., 1996. **118**: p. 8365-8374.
140. Iacobucci, C., et al., *A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions.* Nat Protoc, 2018. **13**(12): p. 2864-2889.
141. Shao, C., Y. Zhang, and W. Sun, *Statistical characterization of HCD fragmentation patterns of tryptic peptides on an LTQ Orbitrap Velos mass spectrometer.* J Proteomics, 2014. **109**: p. 26-37.
142. Remst, D.F., et al., *TGF-β induces Lysyl hydroxylase 2b in human synovial osteoarthritic fibroblasts through ALK5 signaling.* Cell Tissue Res, 2014. **355**(1): p. 163-71.
143. Voloshenyuk, T.G., et al., *Induction of cardiac fibroblast lysyl oxidase by TGF-beta1 requires PI3K/Akt, Smad3, and MAPK signaling.* Cytokine, 2011. **55**(1): p. 90-7.
144. LeBlanc, D.C., *Statistics: Concepts and Applications for Science*. Vol. 2. 2004.
145. Tracy, L.E., R.A. Minasian, and E.J. Caterson, *Extracellular Matrix and Dermal Fibroblast Function in the Healing Wound.* Adv Wound Care (New Rochelle), 2016. **5**(3): p. 119-136.
146. Jack, R.S., *AN UNUSUALLY STABLE DNA BINDING PROTEIN CAN LOCATE ITS SPECIFIC BINDING SITE IN THE PRESENCE OF HIGH CONCENTRATIONS OF UREA.* BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, 1990. **169**(3): p. 840-845.

147. Schiller, H.B., et al., *Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair.* Mol Syst Biol, 2015. **11**(7): p. 819.
148. Naba, A., et al., *Quantitative proteomic profiling of the extracellular matrix of pancreatic islets during the angiogenic switch and insulinoma progression.* Sci Rep, 2017. **7**: p. 40495.
149. Chen, Y., et al., *Lysyl Hydroxylase 2 Is Secreted by Tumor Cells and Can Modify Collagen in the Extracellular Space.* J Biol Chem, 2016. **291**(50): p. 25799-25808.
150. Rautavuoma, K., et al., *Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1-3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity.* J Biol Chem, 2002. **277**(25): p. 23084-91.
151. Kim, N., et al., *Inter-molecular crosslinking activity is engendered by the dimeric form of transglutaminase 2.* Amino Acids, 2017. **49**(3): p. 461-471.
152. Bains, W., *Transglutaminse 2 and EGGL, the Protein Cross-Link Formed by Transglutaminse 2, As Therapeutic Targets for Disabilities of Old Age.* REJUVENATION RESEARCH, 2013. **16**(6).
153. van den Hoogen, F., et al., *2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative.* Arthritis Rheum, 2013. **65**(11): p. 2737-47.
154. Ezzoukhry, Z., et al., *TGF-beta1 promotes linear invadosome formation in hepatocellular carcinoma cells, through DDR1 up-regulation and collagen I cross-linking.* Eur J Cell Biol, 2016. **95**(11): p. 503-512.
155. Wang, Z., et al., *A novel extracellular role for tissue transglutaminase in matrix-bound VEGF-mediated angiogenesis.* Cell Death Dis, 2013. **4**: p. e808.
156. Piercy-Kotb, S.A., et al., *Factor XIIIA transglutaminase expression and secretion by osteoblasts is regulated by extracellular matrix collagen and the MAP kinase signaling pathway.* J Cell Physiol, 2012. **227**(7): p. 2936-46.
157. Martin, P., et al., *Abnormal collagen V deposition in dermis correlates with skin thickening and disease activity in systemic sclerosis.* Autoimmun Rev, 2012. **11**(11): p. 827-35.
158. Amano, S., et al., *Quantitative analysis of the synthesis and secretion of type VII collagen in cultured human dermal fibroblasts with a sensitive sandwich enzyme-linked immunoassay.* Exp Dermatol, 2007. **16**(2): p. 151-5.
159. O'Neill, J.D., et al., *Decellularization of human and porcine lung tissues for pulmonary tissue engineering.* Ann Thorac Surg, 2013. **96**(3): p. 1046-55; discussion 1055-6.
160. Quintanilla-Dieck, M.J., et al., *Expression and regulation of cathepsin K in skin fibroblasts.* Exp Dermatol, 2009. **18**(7): p. 596-602.
161. Fritzsche, R., et al., *Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis.* Rapid Commun Mass Spectrom, 2012. **26**(6): p. 653-8.
162. Rinner, O., et al., *Identification of cross-linked peptides from large sequence databases.* Nat Methods, 2008. **5**(4): p. 315-8.
163. Tatsukawa, H., et al., *Isozyme-specific comprehensive characterization of transglutaminase-crosslinked substrates in kidney fibrosis.* Sci Rep, 2018. **8**(1): p. 7306.

164. Canty, E.G. and K.E. Kadler, *Procollagen trafficking, processing and fibrillogenesis.* J Cell Sci, 2005. **118**(Pt 7): p. 1341-53.
165. Remst, D.F., et al., *Osteoarthritis-related fibrosis is associated with both elevated pyridinoline cross-link formation and lysyl hydroxylase 2b expression.* Osteoarthritis Cartilage, 2013. **21**(1): p. 157-64.
166. Iacobucci, C. and A. Sinz, *To Be or Not to Be? Five Guidelines to Avoid Misassignments in Cross-Linking/Mass Spectrometry.* Anal Chem, 2017. **89**(15): p. 7832-7835.
167. Barysz, H.M. and J. Malmstrom, *Development of Large-scale Cross-linking Mass Spectrometry.* Mol Cell Proteomics, 2018. **17**(6): p. 1055-1066.
168. Hoopmann, M.R., et al., *Kojak: efficient analysis of chemically cross-linked protein complexes.* J Proteome Res, 2015. **14**(5): p. 2190-8.
169. Yang, B., et al., *Identification of cross-linked peptides from complex samples.* Nat Methods, 2012. **9**(9): p. 904-6.
170. Barysz, H., et al., *Three-dimensional topology of the SMC2/SMC4 subcomplex from chicken condensin I revealed by cross-linking and molecular modelling.* Open Biol, 2015. **5**(2): p. 150005.
171. Kleman, J.-P., et al., *Transglutaminase-Catalyzed Cross-Linking of Fibrils of Collagen V/XI in A204 Rhabdomyosarcoma Cells.* Biochemistry, 1995. **34**: p. 13768—13775.
172. Bowness, J.M., J.E. Folk, and R. Timpl, *Identification of a Substrate Site for Liver Transglutaminase on the Aminopropeptide of Type III Collagen.* The Journal of Biological Chemistry, 1987. **262**(3): p. 1022-1024.
173. Lai, T.S. and C.S. Greenberg, *Histaminylation of fibrinogen by tissue transglutaminase-2 (TGM-2): potential role in modulating inflammation.* Amino Acids, 2013. **45**(4): p. 857-64.
174. GRIFFIN, M., R. CASADIO, and C.M. BERGAMINI, *Transglutaminases : Nature's biological glues.* Biochem. J. , 2002. **368**: p. 377–396.
175. Mizunoa, K., et al., *The fibril structure of type V collagen triple-helical domain.* Micron, 2001. **32**: p. 317–323.
176. Siegel, R.C., C.M. Black, and B. A.J., *Cross-linking of Collagen in the X-linked Ehlers-Danlos Type 5.* Biochemical and Biophusical Research Communications, 1979. **88**(1): p. 281-287.
177. Garcia-Pardo A. and G. L., *Further characterization of the binding of fibronectin to gelatin reveals the presence of different binding interactions.* Arch Biochem Biophys., 1993. **304**(1): p. 181-8.
178. Symoens, S., et al., *Identification of binding partners interacting with the alpha1-N-propeptide of type V collagen.* Biochem J, 2011. **433**(2): p. 371-81.
179. Wierzbicka-Patynowski, I. and J.E. Schwarzbauer, *The ins and outs of fibronectin matrix assembly.* J Cell Sci, 2003. **116**(Pt 16): p. 3269-76.
180. Chen, R. and R.F. Doolittle, *γ-γ Cross-linking sites in human and bovine fibrin.* Biochemistry, 2002. **10**(24): p. 4486-4491.
181. Ritchie, H., et al., *Cross-linking of plasminogen activator inhibitor 2 and alpha 2-antiplasmin to fibrin(ogen).* J Biol Chem, 2000. **275**(32): p. 24915-20.
182. Mosher, D.F. and P.E. Schad, *Cross-linking of fibronectin to collagen by blood coagulation Factor XIIIa.* J Clin Invest, 1979. **64**(3): p. 781-7.
183. Bowness, J.M., A.H. Tarr, and R.I. Wiebe, *Transglutaminase-catalysed cross-linking: A potential mechanism for the interaction of fibrinogen, low density lipoprotein and arterial type III procollagen.* 1989. **54**(4): p. 357-367.

184. Sun, F., et al., *A Unique Approach to the Mobile Proton Model: Influence of Charge Distribution on Peptide Fragmentation.* J. Phys. Chem. B, 2010. **114**: p. 6350–6353.
185. Tsiatsiani, L. and A.J. Heck, *Proteomics beyond trypsin.* FEBS J, 2015. **282**(14): p. 2612-26.
186. Mendes, M.L., et al., *An integrated workflow for cross-linking/mass spectrometry.* Mol Syst Biol., 2019. **15**(9).
187. Cheng, Y., et al., *A primer to single-particle cryo-electron microscopy.* Cell, 2015. **161**(3): p. 438-449.
188. Mizikova, I., et al., *Targeting transglutaminase 2 partially restores extracellular matrix structure but not alveolar architecture in experimental bronchopulmonary dysplasia.* FEBS J, 2018. **285**(16): p. 3056-3076.

# Appendix

1. Collagen proteins present in skin (from Protein Atlas) and sequence coverages found in each replicate from the murine proteomics experiment

| Gene name | Protein name | skin expression | type | RNA score | Sequence coverages [%] |
|---|---|---|---|---|---|
| COL1A1 | Collagen, type I, alpha 1 | Expressed in all | fibrillar | high | 62.6; 60.4 |
| COL1A2 | Collagen, type I, alpha 2 | Expressed in all | fibrillar | high | 64.3; 61.6 |
| COL3A1 | Collagen, type III, alpha 1 | Tissue enhanced | fibrillar | high | 58.9; 56.8 |
| COL4A1 | Collagen, type IV, alpha 1 | Tissue enhanced | network | medium | 13.2; 15.7 |
| COL4A2 | Collagen, type IV, alpha 2 | Tissue enhanced | network | high | 23.4; 22.7 |
| COL4A5 | Collagen, type IV, alpha 5 | Mixed | network | low | - |
| COL4A6 | Collagen, type IV, alpha 6 | Mixed | network | low | - |
| COL5A1 | Collagen, type V, alpha 1 | Mixed | fibrillar | medium | 27.6; 28.6 |
| COL5A2 | Collagen, type V, alpha 2 | Mixed | fibrillar | medium | 42.4; 40.5 |
| COL5A3 | Collagen, type V, alpha 3 | Tissue enhanced | fibrillar | low | - |
| COL6A1 | Collagen, type VI, alpha 1 | Mixed | network | medium | 41; 33.6 |
| COL6A2 | Collagen, type VI, alpha 2 | Expressed in all | network | high | 32.5; 28 |
| COL6A3 | Collagen, type VI, alpha 3 | Tissue enhanced | network | medium | - |
| COL6A5 | Collagen, type VI, alpha 5 | Group enriched | network | nd | 8.7; 4.8 |
| COL8A2 | Collagen, type VIII, alpha 2 | Mixed | network | low | 4; 5.6 |
| COL12A1 | Collagen, type XII, alpha 1 | Mixed | FACIT | medium | 7.9; 6 |
| COL13A1 | Collagen, type XIII, alpha 1 | Tissue enhanced | TM | low | - |
| COL16A1 | Collagen, type XVI, alpha 1 | Tissue enhanced | FACIT-like | low | 13.3; 15.5 |

| | | | | | |
|---|---|---|---|---|---|
| COL7A1 | Collagen, type VII, alpha 1 | Tissue enriched | (bridging) | low | 25.7; 24.8 |
| COL28A1 | Collagen, type XXVIII, alpha 1 | Tissue enhanced | (nerve) | low | - |
| COL15A1 | Collagen, type XV, alpha 1 | Tissue enhanced | endostatin | medium | 8.6; 23 |
| COL18A1 | Collagen, type XVIII, alpha 1 | Expressed in all | endostatin | medium | 2.9; 2.9 |
| COL9A2 | Collagen, type IX, alpha 2 | Tissue enhanced | FACIT | low | 3.8; 3.8 |
| COL14A1 | Collagen, type XIV, alpha 1 | Tissue enhanced | FACIT | medium | 2.5; 4.3 |
| COL21A1 | Collagen, type XXI, alpha 1 | Tissue enhanced | FACIT-like | low | - |
| COL27A1 | Collagen, type XXVII, alpha 1 | Mixed | fibrillar | low | - |
| COL17A1 | Collagen, type XVII, alpha 1 | Tissue enriched | TM | nd | 5.2; 6.7 |
| COL23A1 | Collagen, type XXIII, alpha 1 | Tissue enhanced | TM | low | - |

2. ECM protein identifications from the mouse skin proteomics experiment

| Gene name | Score | Peptides | | Unique peptides | | Sequence coverage [%] |
|---|---|---|---|---|---|---|
| **Il1rn** | 2.2514 | 1 | 0 | 1 | 0 | 4.7 |
| **Col15a1** | 4.1513 | 2 | 3 | 2 | 3 | 15.8 |
| **Ccl27;Ccl27a** | 2.9023 | 1 | 1 | 1 | 1 | 11.8 |
| **Cst3** | 6.276 | 2 | 2 | 2 | 2 | 26.5 |
| **Fbn1** | 323.31 | 57 | 54 | 55 | 52 | 25.6 |
| **Agrn** | 10.86 | 1 | 1 | 1 | 1 | 0.85 |
| **Col6a4** | 1.2485 | 0 | 2 | 0 | 2 | 0.7 |
| **Srpx** | 32.381 | 5 | 6 | 5 | 6 | 14.65 |
| **Col6a5** | 98.731 | 15 | 9 | 15 | 9 | 6.75 |
| **Lgals9** | 2.546 | 1 | 1 | 1 | 1 | 4 |
| **Col4a6** | 12.123 | 5 | 7 | 5 | 7 | 6.65 |
| **Papln** | 5.2773 | 3 | 5 | 3 | 5 | 4 |
| **Col14a1** | 110.41 | 3 | 5 | 3 | 5 | 3.4 |
| **Npnt** | 2.3535 | 0 | 1 | 0 | 1 | 1.65 |
| **Kng1** | 1.9818 | 0 | 0 | 0 | 0 | 0 |
| **Pzp** | 28.169 | 7 | 11 | 7 | 11 | 8.85 |
| **Ltbp4** | 16.174 | 3 | 2 | 3 | 2 | 2.8 |
| **Anxa11** | 1.2132 | 1 | 1 | 1 | 1 | 2.9 |
| **Fga** | 86.637 | 15 | 11 | 15 | 11 | 30.05 |
| **Itih3** | 1.7827 | 1 | 1 | 1 | 1 | 2.3 |
| **Col6a3** | 323.31 | 81 | 67 | 81 | 67 | 31.35 |
| **Col12a1** | 99.898 | 10 | 7 | 10 | 7 | 6.95 |
| **Vcan** | 9.7215 | 1 | 1 | 1 | 1 | 3.3 |
| **Hspg2** | 323.31 | 68 | 57 | 68 | 57 | 22.9 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Col6a6** | 61.944 | 15 | 15 | 15 | 15 | 9.6 |
| **Lamb1** | 252.48 | 13 | 13 | 13 | 13 | 8.85 |
| **Hrnr** | 2.5189 | 1 | 1 | 1 | 1 | 3.9 |
| **Anxa4** | 3.8462 | 2 | 2 | 2 | 2 | 12.1 |
| **Lamc1** | 323.31 | 24 | 21 | 24 | 21 | 18.45 |
| **Ecm1** | 3.767 | 2 | 3 | 2 | 3 | 4.65 |
| **Anxa6** | 16.023 | 10 | 7 | 1 | 0 | 14.3 |
| **Ctsa** | 1.0018 | 1 | 0 | 1 | 0 | 1.1 |
| **Lamc2** | 323.31 | 19 | 16 | 19 | 16 | 20.1 |
| **Tinagl1** | 4.1424 | 1 | 2 | 1 | 2 | 3.6 |
| **Col9a2** | 1.171 | 1 | 1 | 1 | 1 | 3.8 |
| **Serpinb8** | 6.7361 | 3 | 3 | 3 | 3 | 9.9 |
| **Tnxb** | 134.96 | 31 | 24 | 31 | 24 | 11.1 |
| **Anxa3** | 2.5756 | 1 | 1 | 1 | 1 | 2.8 |
| **Anxa8** | 81.394 | 5 | 3 | 5 | 3 | 15 |
| **Col5a1** | 323.31 | 29 | 31 | 25 | 27 | 28.1 |
| **Nid2** | 5.3359 | 2 | 2 | 2 | 2 | 3.1 |
| **Col4a1** | 47.438 | 12 | 13 | 11 | 12 | 14.45 |
| **Ctsl** | 5.6261 | 3 | 2 | 3 | 2 | 8.1 |
| **S100a4** | 3.9822 | 2 | 2 | 2 | 2 | 24.8 |
| **Anxa2** | 124.88 | 11 | 10 | 11 | 10 | 29.35 |
| **Serpina1;1b;1c** | 84.263 | 4 | 3 | 2 | 1 | 12.15 |
| **Serpina3c;3f;3g;3k;3m;3n** | 127.3 | 1 | 2 | 1 | 2 | 7.2 |
| **Col3a1** | 323.31 | 74 | 68 | 66 | 61 | 57.85 |
| **Col4a2** | 201.22 | 24 | 25 | 24 | 25 | 23.05 |
| **Anxa1** | 181.81 | 12 | 12 | 12 | 12 | 36.6 |
| **Ccl1** | 1.2462 | 1 | 0 | 1 | 0 | 8.15 |
| **Nid1** | 66.181 | 11 | 11 | 11 | 11 | 11.5 |
| **Ctsb** | 9.193 | 3 | 3 | 3 | 3 | 12.1 |
| **Col1a1** | 323.31 | 72 | 64 | 71 | 63 | 61.5 |
| **Fn1** | 48.94 | 15 | 9 | 15 | 9 | 7.85 |
| **Serpinb2** | 41.24 | 6 | 5 | 6 | 5 | 14 |
| **Anxa6** | 1.7571 | 10 | 8 | 1 | 1 | 16.4 |
| **Lgals1** | 43.092 | 6 | 5 | 6 | 5 | 40.75 |
| **Lgals3** | 5.3086 | 3 | 2 | 3 | 2 | 10.8 |
| **Serpinh1** | 10.32 | 5 | 4 | 5 | 4 | 15.6 |
| **Tgm2** | 232.38 | 8 | 7 | 8 | 7 | 13.55 |
| **Col8a2** | 1.4478 | 1 | 2 | 1 | 2 | 4.8 |
| **Col2a1** | 38.228 | 9 | 9 | 7 | 7 | 12.15 |
| **Bgn** | 230.11 | 11 | 8 | 11 | 8 | 34.95 |
| **Dcn** | 323.31 | 13 | 9 | 13 | 9 | 44.2 |
| **Mug2** | 6.2383 | 2 | 1 | 2 | 1 | 1.75 |
| **Fbln2** | 1.6875 | 2 | 0 | 2 | 0 | 1.7 |
| **Col18a1** | 14.592 | 2 | 2 | 2 | 2 | 2.9 |
| **Timp3** | 0.97945 | 1 | 1 | 1 | 1 | 4.75 |
| **Anxa5** | 243.45 | 6 | 7 | 6 | 7 | 25.55 |
| **Sftpd** | 2.8503 | 2 | 1 | 2 | 1 | 10.7 |
| **S100a11** | 111.45 | 2 | 1 | 2 | 1 | 21.95 |
| **Lum** | 36.455 | 7 | 7 | 7 | 7 | 26.3 |
| **Eln** | 68.737 | 7 | 11 | 7 | 11 | 25.3 |
| **Mfap2** | 84.234 | 1 | 1 | 1 | 1 | 5.4 |
| **Csta** | 39.89 | 3 | 4 | 3 | 4 | 41.25 |
| **Adamts10** | 0.98944 | 1 | 0 | 1 | 0 | 1.1 |
| **S100a3** | 5.6925 | 4 | 4 | 4 | 4 | 51.5 |
| **Serpinb5** | 7.6653 | 3 | 3 | 3 | 3 | 8.5 |
| **Tgfbi** | 207.74 | 18 | 14 | 18 | 14 | 30.35 |
| **Serping1** | 1.0079 | 2 | 2 | 2 | 2 | 6.9 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Ctsc** | 19.107 | 1 | 1 | 1 | 1 | 3 |
| **Loxl1** | 2.532 | 2 | 3 | 2 | 3 | 5.95 |
| **Lama4** | 110.15 | 11 | 11 | 11 | 11 | 9.4 |
| **Col8a1** | 2.599 | 0 | 2 | 0 | 2 | 2.3 |
| **Serpina1d** | 2.3219 | 2 | 2 | 0 | 0 | 7 |
| **Col1a2** | 323.31 | 68 | 59 | 68 | 59 | 62.95 |
| **Col6a2** | 323.31 | 28 | 25 | 28 | 25 | 30.25 |
| **Col6a1** | 323.31 | 31 | 25 | 31 | 25 | 37.3 |
| **Col10a1** | 2.0218 | 1 | 1 | 1 | 1 | 4.7 |
| **Anxa7** | 5.3439 | 4 | 4 | 4 | 4 | 7.9 |
| **Col17a1** | 15.821 | 5 | 6 | 5 | 6 | 5.95 |
| **Tgm3** | 88.687 | 5 | 4 | 5 | 4 | 10.9 |
| **Flg2** | 13.758 | 18 | 16 | 3 | 3 | 25.4 |
| **Col5a2** | 323.31 | 34 | 31 | 34 | 31 | 41.45 |
| **Lama2** | 162.07 | 29 | 21 | 29 | 21 | 10.45 |
| **Reln** | 1.3239 | 1 | 1 | 1 | 1 | 0.4 |
| **Serpinb6;Serpinb6a** | 167.45 | 4 | 5 | 4 | 5 | 14.55 |
| **Lama5** | 262.39 | 28 | 24 | 28 | 24 | 10.5 |
| **Lamb3** | 274.74 | 24 | 21 | 24 | 21 | 25.75 |
| **Col11a1** | 16.076 | 7 | 8 | 3 | 4 | 7.15 |
| **Serpinf2** | 2.7057 | 2 | 2 | 2 | 2 | 4.5 |
| **Lamb2** | 323.31 | 32 | 27 | 32 | 27 | 22.1 |
| **Fbn2** | 2.0344 | 3 | 3 | 1 | 1 | 1.15 |
| **Itih1** | 5.7317 | 2 | 4 | 2 | 4 | 4.75 |
| **Lama3** | 323.31 | 29 | 28 | 29 | 28 | 11.05 |
| **Prg2** | 3.1177 | 2 | 2 | 2 | 2 | 10.8 |
| **Ogn** | 323.31 | 5 | 5 | 5 | 5 | 18.1 |
| **Postn** | 323.31 | 18 | 19 | 18 | 19 | 31.45 |
| **Cstb** | 2.8282 | 1 | 1 | 1 | 1 | 11.2 |
| **Col7a1** | 323.31 | 48 | 47 | 48 | 47 | 25.25 |
| **Col4a5** | 49.575 | 3 | 4 | 2 | 3 | 4.3 |
| **Wnt8a** | 0.97979 | 1 | 1 | 1 | 1 | 3.4 |
| **Try10** | 1.1248 | 1 | 0 | 1 | 0 | 5.3 |
| **F13a1** | 13.969 | 7 | 6 | 7 | 6 | 10.6 |
| **Col16a1** | 155.05 | 12 | 15 | 0 | 1 | 14.4 |
| **Tgm6** | 77.019 | 8 | 8 | 8 | 8 | 12.9 |
| **Efemp1** | 24.482 | 7 | 4 | 7 | 4 | 13.15 |
| **Serpina3b** | 22.627 | 2 | 2 | 2 | 2 | 5.7 |
| **Ltbp1** | 33.704 | 5 | 5 | 5 | 5 | 4.4 |
| **Fgb** | 234 | 12 | 9 | 10 | 8 | 30.25 |
| **Emilin2** | 19.45 | 4 | 4 | 4 | 4 | 4.9 |
| **Vwa1** | 5.9794 | 2 | 0 | 2 | 0 | 5.4 |
| **Cd109** | 9.2568 | 6 | 5 | 6 | 5 | 5.65 |
| **Fgg** | 8.6538 | 7 | 3 | 7 | 3 | 14 |
| **Col26a1** | 1.1329 | 0 | 2 | 0 | 2 | 4.35 |
| **Hpx** | 7.0311 | 3 | 4 | 3 | 4 | 8.4 |
| **Lrg1** | 1.3813 | 1 | 1 | 1 | 1 | 4.1 |
| **Emilin1** | 268.92 | 11 | 12 | 11 | 12 | 18.75 |
| **Vwa5a** | 14.754 | 3 | 2 | 3 | 2 | 4.05 |
| **Aspn** | 172.16 | 15 | 12 | 15 | 12 | 39.4 |
| **Serpinb11** | 12.611 | 5 | 4 | 5 | 4 | 12.85 |
| **Lgals7** | 11.818 | 3 | 1 | 3 | 1 | 19.15 |
| **Lman1** | 2.6167 | 2 | 2 | 2 | 2 | 3.7 |
| **Serpinb1a** | 8.0901 | 0 | 1 | 0 | 1 | 1.45 |
| **Cst6** | 5.4458 | 4 | 4 | 4 | 4 | 32.9 |
| **Mfap4** | 122.11 | 4 | 3 | 4 | 3 | 21.05 |
| **Tchhl1** | 90.551 | 4 | 5 | 4 | 5 | 9 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Hrg** | 2.4788 | 2 | 2 | 2 | 2 | 4.6 |
| **Anxa9** | 1.1287 | 1 | 1 | 1 | 1 | 3.5 |
| **Prelp** | 10.968 | 3 | 2 | 3 | 2 | 7.3 |
| **Tgm1** | 5.9258 | 4 | 4 | 4 | 4 | 6.7 |
| **Col5a3** | 240.13 | 11 | 12 | 10 | 11 | 11.95 |
| **Col4a3** | 1.7653 | 3 | 3 | 3 | 3 | 3.8 |
| **Dpt** | 109.73 | 5 | 4 | 5 | 4 | 25.65 |
| **Ctsf** | 0.96014 | 1 | 1 | 1 | 1 | 3.9 |
| **Ctsj** | 1.1561 | 1 | 0 | 1 | 0 | 2.25 |
| **Fbln5** | 22.182 | 6 | 4 | 6 | 4 | 13.3 |