Targeting Mesenchymal Stem Cells to damaged cartilage in Osteoarthritis

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Anaïs Dabbadie

November 2019

Abstract: Targeting Mesenchymal Stem Cells to damaged cartilage in Osteoarthritis

Anaïs Dabbadie

Osteoarthritis is a condition which results in the degradation of cartilage, the protective tissue which enables movement of our joints with a low coefficient of friction. No treatment is currently available to cure the cause of the disease. The injection of Mesenchymal Stem Cells (MSCs) into the joints of osteoarthritic patients has been investigated. Although these injections were shown to improve pain, no cartilage repair was observed, and it was shown that MSCs did not engraft at the site of injection.

We hypothesised that increasing the number of cells which can accumulate onto the lesion would lead to better cartilage regeneration. This thesis describes the development of a method to specifically target MSCs towards type II gelatin, a molecule which is uniquely found in damaged cartilage, by coating them with a protein which binds to this substrate.

MSCs were successfully coated with a chimeric protein named 222, which was engineered following the expression and characterisation of the Collagen Binding Domain (CBD) of the Matrix Metalloprotease MMP-2. 222 was shown to bind to type II gelatin with a Kd of 2.46 nM, a value which was 14 times lower than that obtained for the CBD. The protein was shown to be present at the cell surface by confocal microscopy, and the viability of coated MSCs was similar to that of wild-type cells. Cells coated with 222 were shown to bind significantly better to type II gelatin than uncoated cells *in vitro*, with a median attachment of 224.4% for coated cells compared to 23.67% for wild-type cells.

Acknowledgements

I want to thank both Professor Anthony Hollander and Professor Lu-Yun Lian as well as Doctor Anna Salerno for their supervision and support during these past 4 years.

A lot of persons have helped me in the lab – too many to mention them all. Special thanks to Dr Teresa Almeida, Dr Katie Jameson, Dr Hannah Davies, Dr Mark Wilkinson, Dr Dave Mason, Dr Amy Woods, Mark Prescott, and Dr Philip Brownridge who have all helped me acquire some of the data presented in this thesis.

A very special mention to Dr Nordine Helassa for helping me so many times.

Finally, a big thank you to all the people I've shared the lab with and who have made life in the lab easier and funnier.

That same life in the lab which can be stressful at times – and that's when playing tennis becomes vital. So I need to thank all the tennis players I've hit the ball with during these past 4 years for offering me some much needed stress-relieving therapy.

To my friends,

Anna, thank you for helping me going through every single obstacle I've had to face over the past 4 years. I don't know what I would have done without you.

To Paulina and Zoé, thanks for being by my side since we met. I will be forever grateful for your support and friendship. Life has hit us pretty hard in the past year or so, but I am proud of the way we are here for each other.

Thanks to Maider for being my oldest friend, and for always making me see the positive side of life. From that bad joke in secondary school to that trip to Ireland which was a game changer in my PhD, you've always been there for me.

To the best of the best, Charlène, Danaëlle and Laure. 4 girls, each in a different country, and yet we managed to support each other over everything we had to go through during these past 4 years. Thanks for showing me what true friendship really means. I'll never be able to thank you enough, nor to explain to you what you all mean to me. [insert "I love you" GIF here]. Laure, you can stop reading now. Hope it wasn't too long.

To my family back home,

Grandma, thank you for being the strongest woman I know and for always believing in me since I was a little girl. I admire you more than I can say.

Mum and Dad, a thousand of "thank you" will never be enough for what you did and still do for us. Thank you for your daily support, for being strong and positive when I am not, for your sacrifices, and for your love.

Julien. You said it, we're not good at talking. I am not bad at writing, but you might not like a display of sentiments here. So let me just say that it was a privilege to grow up with you. And it will be another one to grow old with you. I'll be there.

To Delphine and her kids Emma, Enya and Louca. Thank you for making my brother happy.

To Chloé. I've never stopped thinking about you since that Christmas day. And I never will. "Tu n'es plus où tu étais, mais tu seras partout là où je suis"

Summary

List of abbreviations

ACI: Autologous Chondrocyte Implantation

ADAMTS: A Disintegrin and Metalloproteinase with ThromboSpondin motifs

AD-MSCs: Adipose-Derived Mesenchymal Stem Cells

BME: Beta-Mercaptoethanol

BM-MSCs: Bone Marrow Mesenchymal Stem Cells

BMPs: Bone Morphogenic Proteins

CBD: Collagen Binding Domain

CD: Circular Dichroism

CDCs: Cardiosphere-Derived Cells

CM: Conditioned Medium

CS: Chondroitin Sulfate

CV: Column Volume

DLS: Dynamic Light Scattering

DMEM: Dulbecco's Modified Eagles Medium

DMOADs: Disease Modifying Osteoarthritis Drugs

DMPA: N,N`-dimethyl-1,3-propanediamine

DMSO: Dimethylsulfoxide

DNA: Deoxyribonucleic Acid

DTNB: 5,5-dithio-bis-(2-nitrobenzoic acid)

DTT: Dithiothreitol

ECL: Enhanced Chemiluminescence

ECM: Extracellular Matrix

EDC: N-(3-Dimethylaminopropyl)-N`-ethylcarbodiimide hydrochloride

EDTA: Ethylenediaminetetraacetic acid

EGF: Epidermal Growth Factor

ESCs: Embryonic Stem Cells

FACIT: Fibril-Associated Collagens with Interrupted Triple-helix

FACS: Fluorescence Activated Cell Sorting

FBS : Foetal Bovine Serum

FDA: Food and Drug Administration

FF: Fast Flow

FGF: Fibroblast Growth Factor

FITC: Fluorescein-5-isothiocyanate

GAGs: Glycosaminoglycans

GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase

GFP: Green Fluorescent Protein

HA: Hyaluronan

HEK: Human Embryonic Kidney

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HGF: Hepatocyte Growth Factor

HLA: Human Leukocyte Antigen

HSQC: Heteronuclear Single Quantum Correlation

IEC: Ion Exchange Chromatography

IGF1: Insulin-like Growth Factor 1

IKDC: International Knee Documentation Committee

IL: Interleukin

IL-1β: Interleukin 1 beta

IL1-RA: Interleukin 1 Receptor Antagonist

iPSCs: Induced Pluripotent Stem Cells

IPTG: Isopropyl-ß-D-thiogalactopyranoside

KS: Keratan Sulfate

LB: Lysogeny Broth

LIF: Leukemia Inhibitory Factor

LP: Link Protein

LT: Large antigen T

MACI: Matrix-Associated Chondrocyte Implantation

MALDI-TOF: Matrix Assisted Laser Desoprtion Ionisation – Time Of Flight

MBP: Maltose Binding Protein

MMPs: Matrix Metalloproteases

MRI: Magnetic Resonance Imaging

MS: Mass Spectrometry

MSCs: Mesenchymal Stem Cells

MWCO: Molecular Weight Cut-Off

NaP: sodium phosphate

NMR: Nuclear Magnetic Resonance

NO: Nitric Oxide

NSAIDs: Non Steroidal Anti-Inflammatory Drugs

OA: Osteoarthritis

OD: Optical Density

PBS: Phosphate Buffer Saline

PCR: Polymerase Chain Reactions

PDB: Protein Databank

PEG: Polyethylene Glycol

PEI: Polyethylenimine

PISA: Proteins, Interfaces, Structures and Assemblies

PNPP: P-Nitrophenyl Phosphate

PPG: Palmitated Protein G

PRP: Platelet Rich Plasma

P/S: Penicillin/Streptomycin

PTOA: Post-Traumatic Osteoarthritis

RNA: Ribonucleic Acid

RT: Room Temperature

SASP: Senescent Associated Secretory Phenotype

SDS: Sodium Dodecyl Sulphate

SDS PAGE: Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis

SLRPs: Small Leucine-Rich Proteoglycans

SPARC: Secreted Protein Acidic And Cysteine Rich

SUMO: Small Ubiquitin-related Modifier

SVF: Stromal Vascular Fraction

SV40: Simian Virus 40

SZP: Superficial Zone Protein

TGE: Transient Genetic Expression

TGFβ: Transforming Growth Factor beta

TIMP1: Tissue Inhibitor of Metalloproteinase-1

TNFα: Tumour Necrosis Factor alpha

VAS: Visual Analogue Scale

VEGF: Vascular Endothelial Growth Factor

WGA: Wheat Germ Agglutinin

WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index

WORMS: Whole-Organ MRI Scoring Method

WT: Wild-Type

1. Introduction

1.1. Anatomy of the joint

A joint is defined as the point of contact between two bones. Depending on the movement they allow, joints can be classified as synarthrosis (immovable), amphiarthrosis (slightly moveable), and diarthrosis (freely moveable) [1]. Diarthrodial joints include the hip, ankle, elbow, shoulder and the knee [2], and all of them share common structural features (Figure 1- 1). They are enclosed in a fibrous capsule [3], whose inner surfaces are lined with the synovium [2]. This tissue, which contains fibroblast-like B cells and synovial macrophages (A cells), [3] is responsible for the secretion of the synovial fluid [2], a viscous liquid composed of a protein-rich plasma ultrafiltrate and hyaluronan [4]. The synovial fluid also contains electrolytes such as sodium, potassium, calcium and chloride [4], phospholipids [3] as well as the Superficial Zone Protein (SZP) [4], which helps with lubrication. The synovial fluid is also responsible for nutrients supply and waste removal [4]. Finally the articulating bones are covered with a smooth and thin tissue: the articular cartilage.

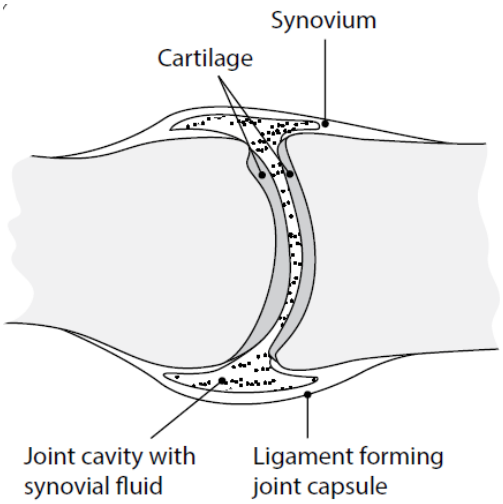


Figure 1-1: Anatomy of a diarthrodial joint. The joint capsule is lined with the synovium, whose cells are responsible for the secretion of the viscous synovial fluid. This fluid helps with the lubrication of the joint. The surface of each bone is covered with articular cartilage.

1.2. Cartilage

1.2.1. Overview

Cartilage can be described as elastic, fibrocartilage or hyaline, depending on the composition of its matrix [5]. Elastic cartilage, which has a high content in elastin fibres [4], is found in the ears and the trachea [6], where it helps maintain the shapes and flexibility of these structures [6,4]. Fibrocartilage can be found in the intervertebral discs, the meniscus and at the interface between ligaments and tendons with bones. It is rich in type I collagen and is the strongest of the three types of cartilage [4]. Finally, hyaline cartilage is present in the nasal cavity as well as in the articulating joints such as the elbow, shoulder, hip and knee joints, where it is referred to as articular cartilage [5,4].

1.2.2. Articular cartilage: structure and function

1.2.2.1. Function

Articular cartilage covers the end of opposing bones in a thin layer (2-4mm) [7], providing a smooth lubricating surface which allows movement with a low coefficient of friction [7,5,6,4]. Healthy articular cartilage appears white and smooth [6], and its ability to bear and dissipate loads throughout the joint [6,4] is directly related to its structure.

1.2.2.2. Structure

1.2.2.2.1. Chondrocytes

Cartilage contains only one cell type, the chondrocytes, which are sparsely distributed throughout the tissue [7,8]. These spheroidal cells [5,4] account for only 1 to 5 % of the volume of articular cartilage [8,5] and are not highly active metabolically [4]. They rarely form cell-to-cell contacts as they are embedded within the Extracellular Matrix (ECM) [7,8,5,4]. However, they play a major role in cartilage biomechanics as they are responsible for the synthesis, degradation and turnover of the components of this matrix [5,4], namely type II collagen and proteoglycans, which provide cartilage with its resistance to tensile strength and compression [5], respectively. Even though these two components form the basis of the biomechanical properties of cartilage, a matrix devoid of cells would not replicate the biomechanical properties of cartilage, highlighting the importance of the cellular content of cartilage to maintain the homeostasis of this tissue via responses to stimuli such as mechanical loads and hydraulic pressures [7,8,4].

1.2.2.2.2. Extracellular Matrix

The Extracellular Matrix (ECM) of cartilage is itself made of two major components: water (accounting for 70% of the wet weight) and the structural macromolecules, which include collagens and proteoglycans [7,8].

1.2.2.2.2.1. Collagens in articular cartilage

Collagens are the most abundant macromolecule in the ECM of articular cartilage [7], where they are organised in fibrils [9] (Figure 1-2). Type II collagen is the most abundant, as it represents 80 to 85% of the total collagen content [10]. This collagen form fibrils with type XI collagen, which is thought to regulate their diameter [10]. The fibrils are completed by the Fibril-Associated Collagens with Interrupted Triple-helix (FACIT) collagen type IX, which is exposed at their surface, where it can cross-link with other collagen molecules as well as interact with other proteoglycans in the ECM [11,10]. Finally, type X and VI collagens are also found in articular cartilage. The former is present in the calcified cartilage and is a marker of hypertrophic chondrocytes [10], while the latter is abundant in the pericellular matrix of chondrocytes [12]. Some reports have also described the presence of type I collagen at the surface of the tissue [4].

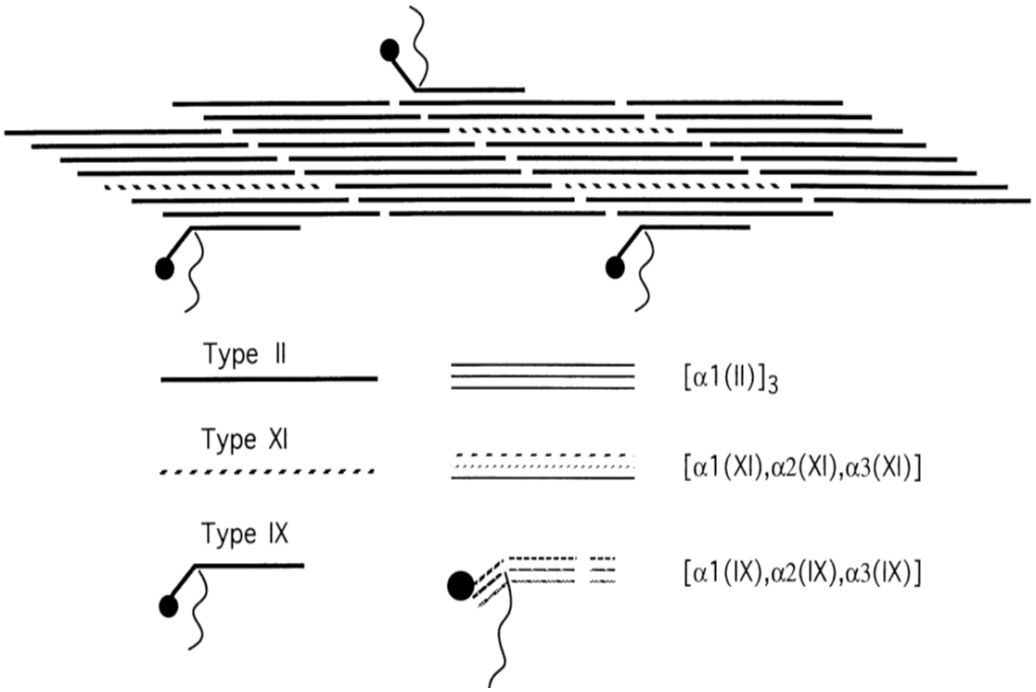


Figure 1-2: Composition of the collagen fibrils in articular cartilage. Collagen fibrils in articular cartilage are mostly made of type II collagen, which associates with type XI and type IX collagen. The latter is exposed at the surface of the fibrils to interact with other matrix components. The alpha chain composition of each type of collagen is also displayed. Each alpha chain has a different sequence, although all of them will be rich in glycine, proline and hydroxyproline [9]. Figure adapted from [10].

All collagens share the same structural features, as they are organised in a triple helix made of three alpha chains rich in glycine, proline and hydroxyproline [7,4]. All three residues confer a stability to the triple helix, which is reinforced by intra chains hydrogen bonds with the hydroxyproline [11] and electrostatic interactions between aspartate and lysines [9]. The helical structure of collagen molecules coupled to their organisation in a fribrillar meshwork provides cartilage with its resistance to shear and tensile strength [7,8].

1.2.2.2.2.2. Proteoglycans

Proteoglycans are made of a protein core to which Glycosaminoglycans (GAGs) are attached [8]. The most abundant proteoglycan in cartilage is aggrecan, which is made of a core protein of more than 200 kDa [13]. This protein core possesses more than 100 Chondroitin Sulfate (CS) and Keratan Sulfate (KS) chains [7,4] which form long chains of negatively charged groups that repel each other [7,5] (Figure 1-3, A). This structure is able to interact with Hyaluronan (HA), a non-sulphated glycosaminoglycan which is present as a coating agent at the surface of chondrocytes [14], with the stabilising support of a Link Protein (LP). They form large aggrecan molecules [7,5] which are entrapped within the collagen fibrils [8]. Up to 100 aggrecan molecules can bind to a single HA, and their association with the link protein is undissociable under physiological conditions [13]. Their high negative charge density provides cartilage with osmotic properties that trap water in the matrix, thereby enabling the tissue to withstand compressive loads [7,5,4].

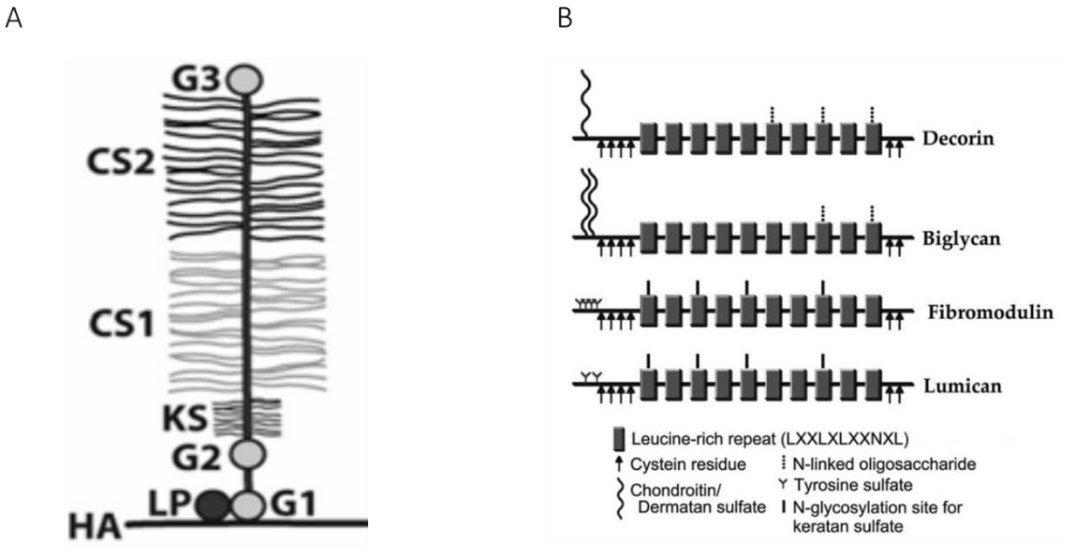


Figure 1-3: Structure of the proteoglycans aggrecan (A) and small leucine-rich proteoglycans (B). A: Aggrecan is made of a core protein with 3 interglobular domains (G1, G2 and G3) which present sites of attachment for Keratan Sulfate (KS) and Chondroitin Sulfate (CS). The Link Protein (LP) mediates its attachment to Hyaluronan (HA). Figure adapted from [14]. B: Small leucine-rich proteoglycans contain a protein core with leucine-rich repeats and sites of attachment for either chondroitin, dermatan or keratan sulfate. Figure from [15].

While aggrecan accounts for 90 % of the total proteoglycans mass within articular cartilage [8], the other 10 % correspond to small proteoglycans such as the dermatan sulfates decorin and biglycan and the keratan sulfates fibromodulin and lumican [14] (Figure 1-3, B). Their protein core, which is much smaller than that of aggrecan, with an average size of 40 kDa [15], contains 10 motifs rich in leucines, which are flanked by disulfide bridges [14]. Except for biglycan, the interaction of these Small Leucine-Rich Proteoglycans (SLRPs) with collagen fibrils have been demonstrated [14]. These SLRPs are therefore believed to regulate the diameter of the fibrils [13,16] as well as to protect them from degradation, by masking the recognition sites of collagenases [14]. Finally, all 4 of these SRLPs have been shown to be able to bind to other matrix proteins as well as cytokines and growth factors [14,15]. A particular role has been established for the binding of these SLRPs to Transforming Growth Factor beta (TGFβ), allowing control of its access to the cells, ultimately affecting cell proliferation and metabolism [14,16].

1.2.2.2.2.3. Water

70% of the wet weight of cartilage is due to water [7,5], which enables the transport of nutrients as well as the lubrication of the tissue [7,5,4] . More importantly, it allows for load-dependent deformation [5,4] through its interaction with the negatively charged proteoglycans [8,4].

1.2.2.2.3. Structural organisation

One important structural feature of cartilage is its heterogeneity throughout the depth of the tissue. Based on the shape of chondrocytes and the orientation of the collagen fibrils, 4 zones can be distinguished (Figure 1-4) [7,8].

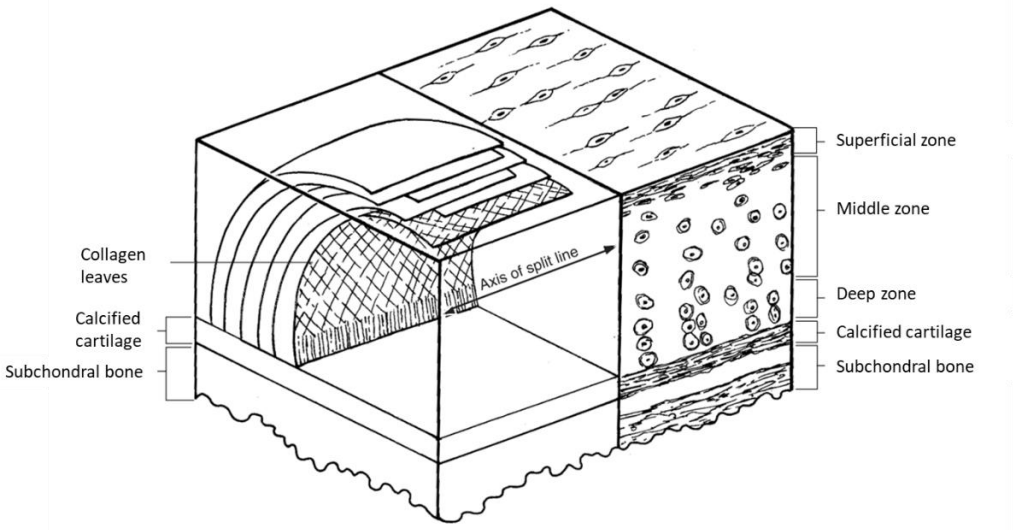


Figure 1-4: Structure of articular cartilage. The background displays the different zones of cartilage together with the shape and organisation of chondrocytes. The foreground represents the orientation of collagen fibrils throughout the depth of the tissue. Figure adapted from [6].

The superficial zone, which accounts for 10 % to 20 % of the depth of the tissue [7], contains flattened chondrocytes and collagen fibrils which run parallel to the surface [7,8,5]. The proteoglycan content is the lowest, and the amount of water the highest [8,5]. This zone also contains the protein lubricin or lamina splendens [8,5], which is secreted by both the chondrocytes and the synovium [17], and which reduces the coefficient of friction [17]. This superficial zone is responsible for withstanding sheer and tensile stresses [7,8,5]. The high

density of the collagen fibrils also limits the size of the molecules which can go through the matrix, hence isolating cartilage from the synovial tissue immune system [8,5].

The middle zone, which corresponds to 40 to 60 % of the tissue, is made of randomly/obliquely organised, thicker collagen fibrils, with a higher content in proteoglycans than the superficial zone [7,8,5], and a lower water content [8]. The chondrocytes, which are fewer than in the superficial zone [7], appear spherical [7,8,5] in this zone, which is the first line of resistance to compression [7].

The deep zone, which corresponds to 30 % of the thickness of the cartilage [7], contains chondrocytes organised in columns and the thickest collagen fibres, which are oriented perpendicular to the joint surface [8,5]. The cell density is lower than in the other two zones [5,4]. The proteoglycan content is at its highest and the water content at its lowest, providing this zone with the highest resistance to compression [7,8].

Below the deep zone is the calcified zone, which is rich in type X collagen and hypertrophic chondrocytes and which protects the subchondral bone [5]. The deep zone is separated from the calcified zone by the tide mark [5,7].

1.2.2.3. Challenges associated with cartilage repair

One particularity of articular cartilage is that it is aneural, avascular and alymphatic [7,5,4]. One direct consequence is that there is no early repair response to injury involving macrophages and monocytes [17]. The low cell density and low proliferative activity of chondrocytes coupled to the lack of blood supply mean that cartilage has a very limited capacity for intrinsic healing [7,4]. Finally, the complex three dimensional structure of cartilage, which is needed to ensure proper function, is challenging to recreate with regenerative techniques [7,4]. Cartilage injuries are therefore difficult to treat, and untreated lesions could lead to osteoarthritis.

1.3. Osteoarthritis

Osteoarthritis (OA) is a chronic, progressive and degenerative disease [18] which mostly affects the load- bearing joints such as the knee, the hip and the shoulder [19,20]. It is characterised by a narrowing of the joint space [19] which is the result of a thinning of the articular cartilage and a thickening of the calcified cartilage. It is often accompanied by the formation of osteophytes and, in some cases, an inflammation of the synovium [21,19] (Figure 1-5). OA causes pain, stiffness and movement limitation [22,20]. It is thought to be favoured by risk factors such as obesity, gender, genes, previous history of trauma [23,20] and age [23].

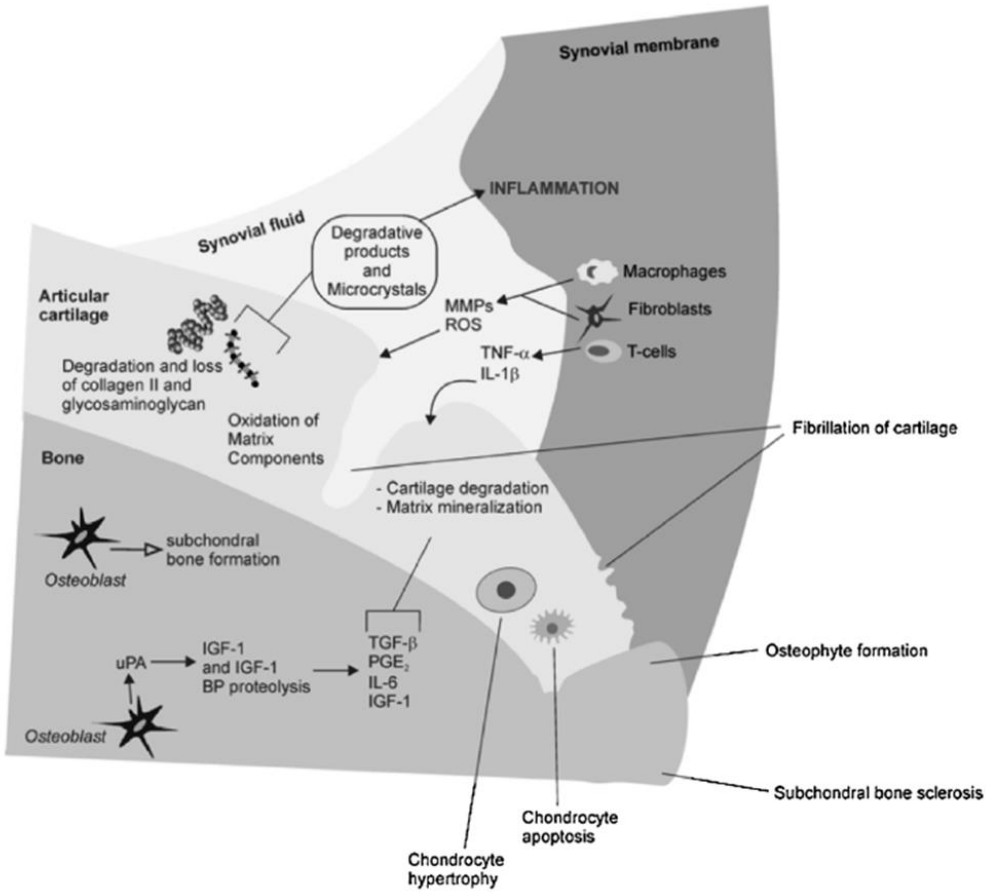


Figure 1-5: Molecular mechanisms of OA progression. OA is characterised by the degradation of the matrix of articular cartilage, which is mediated by the overexpression of several Matrix Metalloproteases (MMPs). Cartilage degradative products are also though to recruit the immune cells from the lining synovium, initiating an inflammatory response. Chondrocytes are also altered in OA, where they can die by apoptosis, de-differentiate (not shown) or exhibit a hypertrophic phenotype. These changes lead to cartilage fibrillation and calcification, with the formation of osteophytes. Figure from [23].

Normal articular cartilage is characterised by the presence of quiescent chondrocytes [20,21] which tightly control the expression of a cocktail of growth and transcription factors, as well as cytokines and enzymes, in order to maintain the tissue homeostasis and its function [20]. This homeostasis is largely disturbed in OA, where catabolic processes become prevalent. OA is indeed characterised by the overexpression of several matrix-degrading enzymes, such as the Matrix Metalloproteases (MMPs) MMP-2, MMP-3, MMP-8, MMP-9, MMP-13 and MMP-14, which are able to degrade most of the matrix molecules [20]. MMP-13 is thought to play the most crucial role in the degradation of type II collagen, while the aggrecanases A Disintegrin And Metalloproteinase with Thrombospondin Motifs (ADAMTS) 4 and 5 are thought to be involved in the degradation of aggrecan [21,20]. The degradation of these key components of the cartilage ECM leads to an increased permeability of the tissue and an alteration of its biomechanical properties.

These catabolic responses are further exacerbated by inflammation of the synovial membrane [20], which is thought to occur when cartilage fragments resulting from the matrix degradation fall in the lining synovium [21,20]. This results not only in the activation of both the innate and adaptive immune responses [24,21,19], but also drives the expression of inflammatory mediators, such as Tumour Necrosis Factor alpha (TNFα) and Interleukin 1 beta (IL-1β), which further increase the expression of matrix-degrading enzymes [21,20], and repress the synthesis of collagen and proteoglycan [25]. Even though some compensatory mechanisms have been observed, through the activation of Bone Morphogenic Proteins (BMPs) and Prostaglandins E2 [20], the overall process results in cartilage degradation which starts at the surface of the tissue, and advances to deeper zones as the condition progresses [20]. The expression of these inflammatory modulators is further exacerbated in ageing cartilage, where senescent chondrocytes secrete additional cytokines, chemokines and proteases belonging to the Senescent Associated Secretory Phenotype (SASP), which further contributes to cartilage degradation [23,26].

The phenotype of chondrocytes is also greatly altered in OA. Normally quiescent, chondrocytes adopt, in the deeper zones of the tissue [27], a proliferative phenotype to form clusters [22,21], in an attempt to initiate a repair of the damaged tissue [21]. However, most of them aberrantly express the Type X collagen, Vascular Endothelial Growth Factor (VEGF) and MMP-13 genes [28], which are typical of a hypertrophic phenotype [29], and will therefore contribute to the calcification of the ECM [21,29]. The calcified cartilage will be thickened, and its invasion by blood vessels from the subchondral bone will result in a duplication of the tidemark, leading to an invasion of cartilage by bone, further contributing to the thinning of this load-bearing tissue [22,21].

In parallel, a subpopulation of chondrocytes at the more superficial zones [30] seem to dedifferentiate as they express higher amounts of type I over type II collagen [31], a phenomenon which will lead to cartilage fibrillation [31].

Finally, OA chondrocytes also seem to be more subjected to apoptosis [22]. Studies in rabbits [32], horses [33] and humans [34] showed that OA cartilage contains higher numbers of apoptotic chondrocytes than healthy cartilage, and, in all three studies, a direct correlation between the grade of OA and the percentage of apoptotic cells was observed. This increased level of apoptosis was shown to start at the superficial and middle zones [32,33], and to co-localise with regions of degraded matrix [32,34] and secretion of Nitric Oxide (NO), which is thought to induce this programmed cell death [32,35]. Additional apoptotic chondrocytes were also observed at the tidemark [32] and in regions of chondrocytes clustering [32,34].The higher degree of apoptosis in OA might be due to a reduction in autophagy, although a direct link between the two still needs to be established [36]. The degradation of the ECM observed in OA might also play a role, as chondrocytes are dependent on their interaction with the matrix for survival [37]. However, other findings indicate that chondrocytes apoptosis might precede cartilage destruction [34], and therefore contribute to the initiation of OA [33].

The high level of apoptosis of the unique cells of cartilage has several consequences: this hypocellularity means that damage to the ECM are unlikely to be repaired. As the tissue is avascular, there are no macrophages present to ensure the clearance of apoptotic bodies [38,35] : they therefore remain in the tissue, releasing their content rich in proteases and further contributing to the matrix degradation [35]. Finally, these apoptotic bodies also favour the calcification of the tissue [39].

OA is the most common degenerative joint disease [19] and, according to reports from Arthritis UK, affected 8.75 million of people in the UK in 2013. OA was classified as the 11th highest cause of disability in 2010 [40] and, with increasing ageing population and obesity, is expected to affect 20 % of adults in Europe and Northern America by 2030 [23]. OA could therefore become an even bigger socioeconomic burden for healthcare systems, highlighting the importance of developing novel and effective treatments against this condition [23].

1.4. Osteoarthritis treatments

1.4.1. Drugs and intra-articular injections

1.4.1.1. Non Steroidal Anti-Inflammatory Drugs, corticoids and visco-supplementation

The first line of treatments against mild to moderate OA involve pain management with Non Steroidal Anti-Inflammatory Drugs (NSAIDs), acetaminophen or opioid analgesics [18,41]. These drugs present various unwanted side effects [41] and only manage the consequences of early OA, without addressing its underlying causes [42] nor initiating any repair processes [43]. Intra-articular injections of corticosteroids are an additional option for the treatment relief of pain associated with early OA [42], although their efficacy is reduced over time [42]. Visco-supplementation is also performed through intra-articular injections of hyaluronic acid [44], whose levels are reduced in OA [42], in an attempt to increase the viscosity of the synovial fluid and the motion of the joint [44]. However, the optimal dose and number of injections have yet

to be found [44], and this technique was shown to provide no additional benefit compared to a placebo [42].

Additional strategies for intra-articular therapies have been investigated with the use of Platelet Rich Plasma (PRP).

1.4.1.2. Platelet Rich Plasma

PRP is obtained following the centrifugation of whole blood, yielding a product which is enriched in platelets, by up to a factor 10 [45]. It is used in an attempt to restore the imbalance between anabolic and catabolic factors observed in OA [46]. Indeed, the granules contained in the platelet solution can release up to 800 mediators, including growth factors, anti-inflammatory cytokines such as interleukin IL-4, and the antagonist of the Interleukin 1 receptor IL1-RA, as well as proteins involved in angiogenesis [45,47]. The effects of this complex preparation are multiple: PRP is thought to inhibit the action of IL1-β and TNFα, and to promote the action of cartilage matrix molecules such as Fibroblast Growth Factor (FGF) and TGFβ [46]. PRP is also thought to enhance the recruitment of Mesenchymal Stem Cells (MSCs) as well as to favour their adhesion, proliferation and differentiation towards a chondrogenic phenotype [47]. PRP could also reduce the elevated levels of NO and exert anti-inflammatory properties, which are thought to be mediated by the Hepatocyte Growth Factor (HGF) [46].

However, in spite of these promises, clinical data on the injection of PRP for the treatment of OA have demonstrated only limited success. Although PRP was shown to provide more beneficial effects than HA or placebo at 12 months follow-up [48,49], especially on the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and International Knee Documentation Committee (IKDC) scores (see Table 1-1 for a description of these scores) [49], these benefits were shown to be clearer in patients suffering from early OA [48,50]. More adverse events such as pain, swelling and inflammation [46], were also observed following PRP injections compared to HA injections [50,49]. Finally, a longer term follow up of 24 months

showed that pain scores had worsened following PRP injection, highlighting the short term benefit of this treatment [46].

Several issues also need to be addressed before this method can be used as a treatment for OA. PRP can be injected as an inactivated product, allowing for the slow release of its bioactive molecules [46]. Alternatively, it can also be activated in order to release the platelet content before its delivery [46]. Whether activation is required and more efficient needs to be investigated [46]. Furthermore, the method of activation will lead to variable PRP content [45]. Other sources of variability include the method of preparation (*i.e.* number of centrifugations), the donor physical condition and its gender [45,47]. PRP can also exist as either poor or rich in leukocytes [46]. The latter will present anti-microbial effects, but may elicit an inflammatory response [46]. Clearly, standardisation in the preparation of PRP is needed to fully understand its efficacy and mode of action. Currently, however, the use of PRP on its own is not approved by the Food and Drug Administration (FDA) nor recommended for the treatment of OA, as no structural improvement has ever been demonstrated [42,44].

Table 1-1: Scores used to evaluate the progression of OA in pre-clinical and clinical trials

Score	Evaluation	Scoring system
Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [51]	Evaluates pain, stiffness, and physical function with 5, 2, and 17 questions, respectively	Ordinal scale of 0 to 4, with lower scores indicating lower levels of symptoms or physical disability
International Knee Documentation Committee (IKDC) [52]	Evaluates symptoms (pain, swelling, stiffness…) ; daily and sports activities ; and knee function with 7, 10 and 1 question, respectively	The response to each item is converted into an ordinal score, with lower scores representing worse symptoms or lowest levels of function. A total score is then calculated out of a 100, with the maximum score being equivalent to absence of symptoms and no limitation in daily or sporting activities.
Whole-Organ MRI Scoring Method (WORMS) [53]	Evaluates 14 features such as articular cartilage integrity, presence of osteophytes, cysts and loose bodies, meniscal and ligaments integrity, synovitis	Grading scales vary for each feature, lower scores indicate normal features.
Visual Analogue Scale (VAS) [54]	Evaluates pain intensity. Unidimensional measure	Scale from 0 to 100 with larger scores indicating greater pain intensity
Lysholm [52]	Evaluates symptoms of knee instability. 8 criteria: limp, support, locking, instability, pain, swelling, stair climbing, squatting	Individual scores for each criteria are added to give a total score on a scale of 0 to 100. A score of 100 indicates no symptoms or disability

In parallel to these pain management therapies, surgical procedures are available to repair focal cartilage lesions.

1.4.2. Surgical procedures

Three main cartilage repair procedures are currently clinically available: microfracture, Autologous Chondrocyte Implantation (ACI) and mosaicplasty [4].

1.4.2.1. Microfracture

Microfracture is the first-line of surgical treatment for full thickness chondral defects [5], where the subchondral bone is exposed [5]. This bone marrow stimulation technique relies on the perforation of the subchondral bone following debridement [5,17], which consists in removing, under arthroscopic vision, the damaged bone and cartilage fragments [55]. Perforation of the subchondral bone releases the resident MSCs into a fibrin clot [5,17], which favours their proliferation and differentiation into chondrocytes [5] in a remodelling process which can take up to 16 months [4]. However, it was reported that this remodelling usually leads to the formation of fibrocartilage [5,17], which is not suited to withstand the load associated with diarthrodial joints [4]. This fibrocartilage was also shown to deteriorate within 18 to 24 months, suggesting that this technique only delays the tissue degeneration observed in OA [43]. The penetration of the subchondral bone was also shown to lead to the formation of osteophytes [41,43], which might limit the durability of this treatment [41]. Finally, patients who had undergone microfracture presented higher failure rates when subjected to Autologous Chondrocyte Implantation (ACI) [43].

1.4.2.2. Autologous Chondrocyte Implantation

ACI is another surgical technique to treat focal cartilage lesions. It is a two-step technique where a biopsy of healthy tissue is obtained from the patient during a first arthroscopy [56,5,17]. Chondrocytes are then expanded in culture until reaching densities of between 12 to 48 million cells [43]. These cells are then re-implanted in the patient during a second surgery [4,17]. In order to avoid leakage, a periosteal patch is added to seal the surface [4]. ACI was shown to

provide clinical benefits for patients with large focal defects in the knee [43], and is usually prescribed following failure of microfracture [4]. However, its superiority to the latter was not demonstrated after a 15-year follow-up comparing both techniques [57], and neither of these methods seemed efficient at protecting patients from developing osteoarthritis [57]. This technique does not pose any immunogenic risks are the patient's own cells are used [43]; however, the two surgeries required and the expansion of chondrocytes *in vitro* make it expensive [18]. The tendency of chondrocytes to de-differentiate when expanded *in vitro* poses another problem [18,43], although chondrocytes were shown to be able to re-differentiate [43]. A hypertrophy of the flap was also often observed as an undesirable effect [43]. Replacement of this flap by synthetic collagen was adopted in subsequent techniques [43,4]. Other advancements were observed with the development of Matrix-Associated Chondrocyte Implantation (MACI), a technique which relies on the use of a type I/type III collagen scaffold [17]. With this technique, chondrocytes are seeded onto this matrix 3 days before their implantation, in an attempt to prevent their de-differentiation [43]. The matrix was designed to allow chondrocyte infiltration on one side and lubrication on the other [43], enabling better cell distribution throughout the defect [58]. In spite of these changes, the superiority of MACI over ACI was not demonstrated in a 1-year follow-up study [59], although studies with longer follow-up showed improvements in pain and function [60,61]. Importantly, however, MACI was shown to provide statistically significantly better outcomes than microfracture for pain and function in a 5-year follow-up study, although no statistical significant difference was observed structurally [62]. Seeding chondrocytes on the matrix for longer period of times, such as 4 to 6 weeks, was proposed as a way to improve the clinical benefits of MACI, in which the neo-formed tissue is still immature when implanted [43]. However, one of the remaining drawbacks of both ACI and MACI is the long postoperative treatment [4]. Even though MACI can be used on larger defects than ACI [58], both techniques are also only suited for focal

cartilage defects, which do not correspond to the generalised degeneration observed in OA [18].

1.4.2.3. Osteochondral autografts and allografts

Finally, the last available surgical method used to correct both chondral and osteochondral defects is mosaicplasty. The technique relies on the biopsy of cylindrical osteochondral plugs in regions of healthy cartilage [5,17]. These plugs are then used to fill the defect in a mosaic pattern [5,17]. Both auto and allografts can be used [4]. However, the latter presents a risk of immunogenicity [4], and the former can only be used to treat small sizes defects [42]. This technique provides a shorter postoperative treatment compared to ACI [4], but complications from the donor site are often observed [4], such as haematomas, inflammation or deterioration of the articular tissue [44]. The gaps between the plugs were also shown to be filled with fibrocartilage, creating zones of instability [5].

Overall, the only available therapies at the moment are either surgical procedures for the repair of focal lesions, or analgesics and anti-inflammatory drugs for the management of early to mild OA. None of the current OA treatments is ideal, and failure of one or a combination of them ultimately leads to total knee replacement [5]. However, this last resort solution is more suitable for older patients, usually aged above 60 [5], and presents risks such as infection and pulmonary embolisms [41].

Additional therapies are therefore needed, not only to restore the damaged articular cartilage caused by OA, but also to modify the course of the disease and to provide longer lasting effects [41,63]. To date, no Disease Modifying Osteoarthritis Drugs (DMOADs) are available [63], although several are under investigation and have shown promising initial results.

1.4.3. Disease Modifying Osteoarthritis Drugs

1.4.3.1. Gene therapy

Gene therapy relies on the delivery of nucleic acids to targeted tissues using either viral or non-viral vectors, in order to prevent, treat or cure diseases [64]. Their delivery can either rely on the direct injection of the modified vectors (*in vivo* approach), or require the extraction, transduction and re-implantation of cells of interest into the tissue (*ex vivo* approach) [64]. OA is particularly suited for such a local strategy as it is confined to a limited number of joints, has no extra-articular or systemic components, and no current alternative treatments are available [65]. Several types of transgenes can be used in an effort to limit inflammation, promote anabolic processes or inhibit catabolic ones [64].

One of the earliest strategies aimed at delivering the Interleukin 1 Receptor Antagonist (IL1-RA), in order to inhibit the IL-1 mediated pathway [64]. Direct intra-articular injection of adenoviruses expressing this transgene have shown to slow disease progression in an equine model of Post-Traumatic Osteoarthritis (PTOA) [64]. An Adeno Associated Virus expressing the human IL1-RA gene was shown to be safe and is currently being tested in a Phase I trial in the United States [64].

Another promising outcome of gene therapy for the treatment of OA has come from the development of Invossa. This gene therapy treatment relies on the use of allogeneic chondrocytes transduced with a retrovirus expressing TGFβ1, a growth factor which was shown to be involved in cartilage development and maturation [64]. In order to prevent the formation of tumour following the injection of these modified chondrocytes, these cells were first irradiated [65]. They were also mixed in a 1:3 ratio with untransduced, unirradiated cells of the same origin, which were used as a target to benefit from the paracrine effects of TGFβ1 [64,66]. Several preclinical studies have demonstrated that this strategy leads to cartilage repair in rabbits, and clinical trials have demonstrated the safety of the procedure [66]. Patients

suffering from moderate to severe OA treated with Invossa also showed better clinical outcomes in terms of pain and knee movement [67]. These encouraging results led to the approval of this therapy in Korea for patients with knee OA, but the approval has since been revoked following a mislabelling of one of the ingredients used. This therapy was also being tested as part of a Phase III clinical trial in the US [64], which has now been stopped. Early results from this trial seemed to suggest structural improvement, as cartilage thickness was increased by Magnetic Resonance Imaging (MRI), although no statistical significance was met [68].

In parallel, other molecules are also being investigated as potential DMOADs.

1.4.3.2. Other Disease Modifying Osteoarthritis Drugs

These include inhibitors of NO synthase or of Matrix Metalloproteases, both of which have the potential to modify the progression of the disease [63], although clinical trials for both drugs did not show structural improvements [69].

Interest has also been increasing for the development of the growth factors BMP7 and FGF-18 as potential drugs [63]. The first one exhibits both anabolic and anti-catabolic properties, and was shown to initiate cartilage repair in various animal models of OA [63]. The latter was shown to induce articular cartilage thickness in a rat model of joint degeneration, an observation which might be a consequence of its anabolic and chondrogenic properties [63]. Both drugs were tested in phase II clinical trials in recent years [69].

Finally, calcitonin was shown to inhibit the formation of osteophytes and to preserve healthy bone, and strontium ranelate, a drug initially developed for the treatment of osteoporosis, was shown to reduce joint space narrowing in knee OA [63,69], although improvements in pain scores were more moderate [69]. Glucosamine Sulfate and Chondroitin Sulfate are also classified as DMOADs. The former was shown to reduce pain and joint space narrowing in

both animal and human knee joints [44]. Other developments include inhibitors of MMP-13, of aggrecanses 5 and of cathepsin K [69], which are being tested in various clinical trials. Estrogen, parathyroid hormone and doxycycline have also raised interest as potential DMOADs, but are further away from clinical development [69].

The development of these drugs undoubtedly represents exciting opportunities for modifying the course of OA, but more clinical studies will be needed before these drugs become available to OA patients. As OA is a disease of the whole joint with different phenotypes, a correct selection of patients who are more likely to benefit from these treatments needs to be performed, and the identification of biomarkers for both phenotype identification and disease progression needs further development [69]. The complexity of the disease might also prevent its treatment with a single drug, and combinatorial treatments might be needed. A more direct route to clinical success for a disease modifying treatment might therefore lie in the use of Mesenchymal Stem Cells (MSCs).

1.5. Stem cells

Stem cells are defined, at the single cell level, as able to self-renew and differentiate into various cell types [70]. The zygote and blastomeres are considered the most potent cells, as they are able to differentiate into all tissues of the embryo as well as into the cells from the extra-embryonic tissues such as the placenta [70]. They are called totipotent [70]. Other types of stem cells have more limited differentiation potential, and include Embryonic Stem Cells (ESCs), Induced Pluripotent Stem Cells (iPSCs) and Adult Stem Cells.

1.5.1. Embryonic Stem Cells

Pluripotent embryonic stem cells can be isolated from the inner cell mass of the blastocyst and they are able to differentiate into cells from the three different germinal layers (ectoderm, mesoderm and endoderm) [70,71]. Their differentiation potential have made them interesting candidates for regenerative medicine, in which a damaged tissue can be restored thanks to the differentiation of stem cells into a specific cell type [71]. However, the ethical issues associated to their use have limited their therapeutic application, and favoured the emergence of cell reprogramming techniques, in which somatic cells are genetically altered to be restored to a pluripotent state capable of generating embryonic-like stem cells [71].

1.5.2. Induced Pluripotent Stem Cells

The generation of pluripotent stem cells has proven successful in 2007, when Yamanaka et al. managed to reprogram human adult skin fibroblasts into pluripotent stem cells by transducing them with the transcription factors Oct4, Sox2, c-myc and Klf4 [71]. Techniques for transduction include integrating methods relying on retroviruses and lentiviruses as well as non-integrating methods, which rely on the use of adenoviruses, plasmid Deoxyribonucleic Acid (DNA) and proteins [71]. The risk of genetic dysfunction and tumourigenesis associated with the former have encouraged safer methods to be developed, and small molecules have recently demonstrated to be able to induce reprogrammation in a mouse model [71]. The use of fibroblasts as the original cell type to be reprogrammed was another limitation, as the process was highly inefficient and prevented large-scale production [71]. Recently, iPSCs have successfully been generated from the widely available blood cells, therefore increasing their potential for future commercialisation [71,72]. The ability to alter the genome of these cells at very specific sites with the Crispr/Cas9 technique also makes their application possible for correcting disease-causing mutations [71,72].

The clinical applications of iPSCs are varied. They are used to model diseases and recapitulate the environmental, genetic and epigenetic factors leading to a pathology, which can sometimes be patient-specific [71]. They are also used for drug screening as they offer physiological human conditions, contrary to animal models [71]. The ability to study patient-specific diseases also enables the development of personalised medicine, with drugs being tested on a patient-by-patient basis [71]. Finally, iPSCs can be used for regenerative medicine, and recent applications have led to the use of iPSCs-derived retinal pigment epithelium in the treatment of macular degeneration in Japan [72]. Worldwide, several clinical trials are on-going to test the use of iPSCs-derived cells for the treatment of neurological disorders, heart failure and diabetes [72]. However, the broad applications of these cells in regenerative medicine is not yet a reality [72], as many issues still need to be solved. The best cell type for reprogramming needs to be selected, ensuring easy access, limited invasiveness and high efficiency of reprogramming [71]. iPSCs cells have also shown to be heterogeneous, as genetic and epigenetic variations have been observed between different iPSCs cell lines as well as between different passages of the same cell line [71]. These variabilities might affect the differentiation potential of iPSCs or increase their tumourigenecity [71]. Intensive characterisation, by karyotyping or whole genome sequencing, will certainly be needed before these cells can globally progress to the clinic [72]. Regulatory guidelines also need to be adopted in order to validate the methods of reprogramming, culture and differentiation, amongst others [72]. Finally, the commercial feasibility of generating individualised iPSC-derived therapeutics needs to be studied; the alternative would be to generate banks of allogeneic iPSCs matching each potential patient, a process which will need to be highly regulated and which seems unlikely to happen in the near future [71,72].

1.5.3. Adult Stem Cells

For all of these reasons, Adult Stem Cells constitute the best established existing clinical therapy at the moment [70]. Adult stem cells represent only a fraction of the cells in an individual, and have a more limited differentiation potential than iPSCs and ESCs [70]. They are called multipotent, as they can only differentiate into the cells from the tissue in which they reside [70]. One of the major sources of adult stem cells resides in the bone marrow, and encompasses hematopoietic and mesenchymal stem cells [73]. The former are able to restore the entire hematopoietic system, and have clinically be used for more than 30 years in patients suffering from blood disorders such as leukaemia and anaemias [73]. Mesenchymal Stem Cells (MSCs), on the other hand, display properties which make them an interesting candidate for cartilage regeneration [73].

1.5.3.1. Mesenchymal Stem Cells generalities

Mesenchymal Stem Cells (MSCs) are precursors of connective tissues [74] with a capacity for self-renewal and stemness maintenance [75] . They can be isolated from many different tissues, including bone marrow, adipose tissue, spleen, lungs [18], as well as muscle, dermis, both the synovial fluid and membrane, and articular cartilage [23]. Some MSCs can also be found in the umbilical cord blood and in the infrapatellar fat pad [75]. Although all MSCs will exhibit the same phenotypic characteristics, irrespective of their source, their proliferation and differentiation potential will depend on their origin [75]. The large variety of tissues they can be isolated from has led to the suggestion that MSCs could originate from a perivascular niche [74].

MSCs are adherent to plastic, which is one of the three features that are used to characterise them [20,75]. They should also, by definition, be positive for the cell markers CD105, CD73, and CD90 and negative for CD45, CD34, CD14 (or CD11b), CD79a (or CD19), and HLA-DR [75], although other markers can also be used (positive for CD 44, CD 145, CD 166 and STRO-1 and

negative for CD 31 and CD 117) [20,23]. Finally, they exhibit a tri-lineage differentiation potential, with an ability to differentiate into chondrocytes, osteoblasts and adipocytes [20,23,74]. They can also be differentiated into intervertebral disc, muscle and ligaments [23].

Although MSCs are very easily expanded [47,76], one of their main drawbacks is that they are a very heterogeneous population [76,75]. Only one third of primary MSCs are actually able to differentiate into the three main cell types mentioned above, while the other two thirds exhibit an either bipotent or unipotent phenotype [76].

1.5.3.2. Why use Mesenchymal Stem Cells for cartilage regeneration

In spite of these limitations, MSCs are of great interest in regenerative medicine. They are particularly attractive in cartilage regeneration, due to their chondrogenic potential [77], as they could replenish the stock of chondrocytes which is declining in OA.

However, their therapeutic effect could also come from paracrine and immunomodulatory effects, which could alter the progression of OA by restoring the balance between anabolic and catabolic factors [78].

Upon the expression of inflammatory cytokines such as IL-1 and TNFα, MSCs are known to secrete a range of bioactive molecules, such as the growth factors TGFβ, HGF, FGF2, Epidermal Growth Factor (EGF), Insulin-like growth factor 1 (IGF1) and VEGF, as well as NO, prostaglandins E2, and the anti-inflammatory interleukins IL-4, IL-6 and IL-10 [79]. These factors are thought to favour cell proliferation, prevent apoptosis and stimulate chondrogenic differentiation and cartilage matrix formation [78,79]. Recently, it was suggested that one additional mechanism of action of this paracrine effect could be the secretion of extracellular vesicles, called exosomes, by MSCs [41]. Their content is believed to induce the migration of endogenous stem cells into cartilage defects, leading to a partial repair of the damage [43].

MSCs were also shown to be immunosuppressive [77], directly interacting with immune cells to modulate their activities [23], especially their secretion of cytokines [41]. They are able to inhibit the proliferation of both T and B lymphocytes [80] as well as to suppress the activation of natural killer cells [75] and the maturation of monocytes and dendritic cells [78], therefore creating a zone of immune tolerance around the site of damage [77,23]. These effects are mostly mediated by prostaglandins E2, HGF, TGFβ1 [81] and IL-6 [80]. In addition, the secretion of IL-4 and IL-10 enables the restoration of the balance between the T cell helpers H1 and H2, which was shown to favour cartilage regeneration [79]. IL-4 is also responsible for restoring the balance between the pro inflammatory, anti-angiogenic M1 macrophages and the anti-inflammatory, pro remodelling M2 macrophages towards the latter [79].

MSCs were also shown to migrate selectively to the site of damage (homing potential), which adds to the rationale of using these cells in regenerative medicine [18,23,75,76]. They also express low levels of class II Human Leukocyte Antigen (HLA), which enables them to avoid recognition by the immune system and facilitates their use for allogeneic therapies [79].

Finally, MSCs from the joints of patients suffering from end-stage OA have been shown to have altered proliferative and differentiation capacities, with reduced chondrogenic and adipogenic activities compared to MSCs from healthy age-matched patients [74]. These observations add to the rationale of using MSCs as a potential treatment for OA, in order to replace the existing, defective population of MSCs present in the affected joints [74].

1.5.3.3. Sources of Mesenchymal Stem Cells

MSCs isolated from the Bone Marrow (BM-MSCs) have been, so far, the most commonly used for cartilage repair. But the painful harvesting procedure and the risks of infection and sepsis has led scientists to investigate the potential of MSCs from various other sources [74].

1.5.3.3.1. Adipose Derived Mesenchymal Stem Cells

Adipose-Derived MSCs (AD-MSCs) present the advantage of being easily harvested from liposuction waste [20,75], which is generally preferred by patients [41]. They are also present in much larger quantities, up to 500 fold more [76], than MSCs in the bone marrow, and their differentiation and proliferation potential is less likely to be affected by the donor's age [43]. They exhibit the same immunomodulatory and anti-inflammatory properties compared to BM-MSCs [43], and can be differentiated into chondrocytes, muscle cells and adipocytes [20,75]. However, their chondrogenic potential is more limited than that of BM-MSCs, although this limit may be overcome by the addition of appropriate growth factors [75]. Another feature which may limit their use is their higher propensity to undergo chondrocyte hypertrophy [20].

Notably, Stromal Vascular Fraction (SVF) is sometimes used as a therapeutic source of adipose derived MSCs. This product is obtained from liposuction aspirates, which are digested by collagenases, centrifuged and diluted prior to injection [82]. SVF is a highly heterogeneous product as it contains pericytes, fibroblasts, monocytes and macrophages as well as adipose-derived MSCs [82], which only account for 15 to 30 % of the total cell content [83].

1.5.3.3.2. Synovium Mesenchymal Stem Cells

Considering the limitations displayed by the adipose derived MSCs, those isolated from the synovium could represent an interesting alternative. They indeed present the same advantage as AD-MSCs as they can be easily harvested, directly extracted from OA patients during a classic arthroscopic procedure [18]. In addition to presenting the typical trilineage differentiation potential of MSCs [76], they also exhibit a higher chondrogenic potential compared to BM-MSCs [76]. More evidence will be needed to confirm their potential in cartilage repair [75].

1.5.3.3.3. Chondroprogenitors

Finally, the third most interesting candidate as a source of stem cells might be the surface of the articular cartilage itself. It has been shown that resident chondroprogenitors are present in this zone [20,23], although probably at a much lower density in adults compared to young patients. These cells have demonstrated a proliferative potential combined with an ability to differentiate into chondrocytes in 3D pellet cultures [20,23]. An additional advantage is that they do not present any risk of hypertrophic differentiation [20]. It was also suggested that using resident cells might be more relevant to treat cartilage defects, both biologically and physiologically [23].

Other potential sources, such as umbilical cord blood and skeletal muscles, may also be considered in the search for an ideal source of MSCs. They both present the advantage of easy harvesting procedures, offer a large availability of MSCs with a chondrogenic potential and, in the case of cord blood cells, immunomodulatory properties [23,76].

Overall, MSCs present many interesting characteristics for the treatment of OA: they can differentiate into chondrocytes and replenish this cell population which is declining in OA. Their immunomodulatory and anti-inflammatory properties mean that they could modulate the inflammation observed with this condition. The fact that they are easy to cultivate and can be harvested from multiple sites in the body makes them easier to work with than other cell types, such as chondrocytes. MSCs have therefore been used for the regeneration of cartilage in two different ways: either coupled to scaffolds for tissue engineering, or via direct intra-articular injections into the joints of patients suffering from OA.

1.5.3.4. Tissue engineering with Mesenchymal Stem Cells

Tissue engineering relies on a biological substitute to restore a damaged tissue or organ [44]. Traditional tissue engineering strategies consist in preparing the scaffold before seeding it with cells and culturing the mixed phase in a static manner *in vitro*, leading to the formation of immature tissue [4]. The tissue is then transferred into a bioreactor which will mimic the mechanical load found in physiological conditions [4]. The tissue in its more mature form is then implanted into the body, where it will still undergo remodelling to reach complete maturation [4]. The challenges associated with tissue engineering lie in the choice of the best cell population, the right scaffold and the appropriate signals to induce cell differentiation and proliferation [44,4].

MSCs are an interesting cell source for the reasons mentioned above, and seem more promising than articular chondrocytes, which are present in small numbers in cartilage, de-differentiate when expanded in culture, and seem to age faster [4]. However, there are still a number of challenges associated with the use of MSCs in tissue engineering.

The choice of the right scaffold is one of them. The ideal material should be biocompatible and biodegradable [44,4], and would provide cells with a 3D micro-environment that favours their proliferation, differentiation and the synthesis of an ECM [17]. However, this synthesis will be slow [17] and the biomaterial degradation should ideally happen at the same rate as the tissue formation [44,4]. The biomaterial should also withstand the load applied on cartilage until it is completely replaced by the neo-formed tissue [44,4]. In order to facilitate tissue formation, the material should be porous to allow for cell and fluid migration, and support cell proliferation [4].

Common biomaterials used as scaffolds are of two types: either natural, which comprises collagen, fibrin, gelatin, hyaluronic acid, alginate, chitosan and agarose, or synthetic polymers, such as polylactide, PolyGlycolic Acid (PGA) or polyethylene glycol. The former contain ligands which are recognised by the cell receptors, facilitating the adhesion of cells [4] as well as the insertion of the engineered grafts in the defect. They have also been shown to support chondrogenesis [4]. However, they lack the biomechanical properties needed to withstand the stresses associated with diarthrodial joints, are rapidly degraded [44,42], and their architecture is harder to modify [4]. On the other hand, the synthetic biopolymers provide mechanical strength and their degradation rate is much slower [44,42]. Their production is more reproducible [44], and their mechanical and chemical properties can be tuned to different needs [44]. However, they limit cellular interactions and cell-cell signalling compared to natural scaffolds [44], preventing cell attachment and proliferation [4].

The ideal scaffold candidate might lie in the combination of both natural and synthetic polymers [44,4]. The ability to tune the architecture of the engineered tissue with the use of stratified scaffolds [4] or, more recently, 3D bioprinting [76], which will likely help to mimic the complex architecture of native cartilage, are definite advantages for the use of such techniques for cartilage repair.

However, none of these tissue engineering techniques have yet made it to commercialisation. Some preclinical studies failed to show significant improvements compared to control when OA patients were treated with bone marrow MSCs seeded onto a collagen scaffold [76]. Another study showed that cartilage tissue engineered with bone marrow MSCs has led to a tissue with poorer biomechanical properties than when the same biomaterial was seeded with chondrocytes [4], a phenomenon which was thought to be due to the differentiation of MSCs towards the osteogenic lineage [4]. This finding reinforces the need to find adequate cocktails of growth factors to control the differentiation of MSCs *in vitro*, but also highlights that the best cell population for tissue engineering might still need to be found. Interestingly, a study by Kafienah et al. showed that BM-MSCs from OA patients could be used to successfully engineer hyaline cartilage, opening up the possibility to develop an autologous treatment for cartilage lesions [84]. OA BM-MSCs were seeded onto Poly Glycolic Acid scaffolds pre-coated with fibronectin, and appropriate molecular signals were applied using TGFβ3, FGF-2, insulin and parathyroid hormone related protein. The engineered tissue was shown to be histologically comparable to that engineered with nasal chondrocytes. However, the collagen content of the engineered tissue was shown to be lower than that found in native cartilage, a phenomenon which was observed with all the cartilage tissues grown *in vitro* [84]. The formation of immature tissue leading to mechanical failure is indeed the highest risk with tissue engineering [4], and the high cost associated with this technique, coupled to the need for assessing the safety of each component of the engineered construct, renders its commercialisation more complicated [17]. In comparison, direct injection of MSCs into an OA joint is technically easier, and the regulatory path of this technique towards commercialisation could be more straightforward [74].

1.5.3.5. Intra-articular injections of Mesenchymal Stem Cells: pre-clinical evidence

Considering all of these factors, it is not surprising that several groups have tested the efficacy of intra-articular injections of MSCs in different OA animal models, with various degrees of success.

1.5.3.5.1. General data

The injection of allogeneic adipose derived MSCs in the joints of rabbits where OA was surgically induced led to better gross appearance of damaged cartilage, which correlated with both higher histological and radiographic scores compared to a control, untreated group [85]. Similar positive outcomes were observed in a sheep model of OA, where the injection of MSCs led to partial cartilage regeneration both macroscopically and by histology [86]. Notably, when MSCs were cultured in a chondrogenic medium, they led to the additional partial repair of the meniscus [86]. However, these results were not confirmed in horses, where the injection of stromal vascular fraction and bone marrow MSCs were compared to a placebo. The only observed difference was a reduction in the levels of prostaglandins E2 in the group treated with BM-MSCs [87,75]. However, this difference did not lead to an improvement in pain or function, nor in an increase in proteoglycan synthesis [87]. In order to explain these findings, authors have suggested that the retention time of the injected MSCs may have not been long enough to lead to a therapeutic effect [87]. The use of hyaluronic acid as a cell carrier was suggested as a way to improve this outcome [87].

1.5.3.5.2. Results of cell tracking experiments

In order to further investigate the fate of injected MSCs, several studies have been developed in an attempt to track cells once injected. Using a method which enables cell tracking *in vivo*, human adipose MSCs injected in rats were shown to engraft on OA cartilage for 10 weeks, while they persisted for only 4 weeks on healthy cartilage [88]. This difference was thought to

be due to the presence of inflammatory cytokines in the joints of OA animals, which could attract MSCs. The persistence of MSCs for 10 weeks correlated with biological effects, namely, a higher proteoglycan content, lower collagen type I levels and increased cartilage thickness [88]. Injected MSCs were also shown to proliferate at the site of engraftment, and to be contained within the joint, as other tissues such as the brain, lungs and heart were shown to be devoid of expression of the human gene FOXP2 [88]. Interestingly, a similar observation was made by a different group who tracked human adipose MSCs in rabbits *ex vivo* [89]. Again, MSCs were shown to persist at the surface of articular cartilage for up to 10 weeks. The group treated with MSCs displayed better cartilage repair macroscopically, higher levels of proteoglycans and type II collagen, and lower levels of MMP-13 as well as thicker cartilage, compared to the control group which was treated with HA only [89]. The fact that, in both of these studies, the biological effects correlated with the persistence of MSCs in the damaged tissue led the authors to suggest that cartilage regeneration required a persistent cell activity [88]. However, both investigations were performed on non-upright walking animals, where the joint morphology differs from that of humans. Additionally, both models of OA were surgically induced, a procedure which does not necessarily replicate the inflammatory process happening in human OA joints.

Another study, however, was performed on Guinea pigs which spontaneously develop OA, in a model which resembles more the secondary OA observed in humans. Interestingly, authors reported the successful engraftment of injected human MSCs when coupled to HA as a vehicle [90]. Guinea pigs were treated with either MSCs + HA or MSCs+ Phosphate Buffer Saline (PBS), while control groups were treated with HA or PBS alone. The group treated with MSCs + HA showed significantly better macroscopic OA scores than all the other groups, and these results were corroborated with histological findings. More importantly, levels of type II collagen were significantly higher in the MSCs+HA group than in the MSCs+PBS group, and levels of MMP-13 were significantly lower in the former. The cell tracking evaluation revealed

that MSCs injected in PBS were present in low numbers and kept decreasing at each time point (1, 3 and 5 weeks). MSCs delivered with HA were present in higher numbers at all time points. Finally, double immunofluorescence staining showed expression of type II collagen around the injected MSCs as well as around the resident chondrocytes, suggesting that the delivered cells contribute to cartilage regeneration through both direct differentiation and stimulation of endogenous cells, potentially via trophic effects [90]. In this study, HA was therefore critical to allow cell adhesion, engraftment, and proliferation leading to cartilage regeneration. However, the benefits of this vehicle were not replicated in a larger animal model. Following-up on their early observations in rabbits, authors from the above mentioned study [89] investigated the effects of MSCs injection in sheep, in an attempt to determine whether MSCs could slow down OA progression in a large animal model [91]. Sheep were injected with allogeneic adipose MSCs and HA, HA alone or saline. Interestingly, and contrary to their observations in rabbits, MSCs were detected in the synovium 14 weeks after injection. However, sheep treated with MSCs + HA still displayed better cartilage both macroscopically and by histology. High levels of TGFβ and IGF1 were also detected, suggesting that the injected MSCs were promoting chondrogenesis via the secretion of these factors. Type II collagen levels were also increased, and the levels of Tissue Inhibitor of Metalloproteinase-1 (TIMP-1), an inhibitor of matrix metalloproteinases, decreased, although differences were not significant. Finally, the levels of expression of IL-6 and TNFα were also reduced, leading the authors to suggest that the injected MSCs were slowing down OA progression through their anti-inflammatory properties [91]. They however did not exclude the possibility that the cells might have engrafted onto cartilage before their 14-week time point, and recruited endogenous cells to perform cartilage repair [91].

Other pre-clinical studies, however, do suggest that MSCs act via paracrine effects only. In a study of collagenase induced OA in mice [92], adipose derived MSCs injected 7 days after induction led to a reduction in the thickness of the synovial layer, together with a reduction in the number of macrophages and in the levels of expression of ADAMTS-5. Interestingly, when cells were injected 14 days after induction of OA, when inflammation is reduced, no beneficial effect was observed, suggesting that the therapeutic window is important. The tracking of these cells showed their presence in the synovium 24 hours after their injection; cells were no longer detected 5 days after their injection, although the observed therapeutic effects persisted up to day 42.

There is clearly a debate at the moment as to whether MSCs act via direct differentiation or through the release of trophic factors, and, as suggested by the results of the studies mentioned above and by other authors [90], it is possible that both phenomena happen concomitantly to contribute to cartilage regeneration.

However, for MSCs differentiation to happen, it seems clear that these cells need to be retained for long enough periods of time in the joint. In an interesting comparative study of direct injection of MSCs + PRP versus their transplantation with fibrin glue [93], authors have reported that, even though both techniques led to positive clinical outcomes, the scaffold implantation led to significantly better results. Cartilage repair was also shown to be longer-lasting in the tissue engineering group and, interestingly, the correlation between clinical outcomes and the number of cells was only observed in the injection group [93]. Authors have suggested that the lack of retention, survival and engraftment of MSCs following their injection could be a potential explanation for the limited clinical benefits of this treatment [93]. These results strongly suggest that strategies should be implemented to favour the engraftment of MSCs onto cartilage lesions. Results obtained in clinical trials following the injections of MSCs in OA joints seem to support this statement.

1.5.3.6. Intra-articular injections of Mesenchymal Stem Cells: clinical trials results

1.5.3.6.1. Safety

The injection of MSCs was shown to be safe when using bone-marrow autologous cells [94,95,96] and long-term safety was demonstrated at 30 months following injection [97]. Autologous injections of adipose derived MSCs were also performed and showed no adverse effects, even after repeated injections, as mentioned in a review by Lopa et al [98], or with cell doses as high as 100 million [99]. Finally, allogeneic bone-marrow MSCs were also shown to be safe for doses as high as 50 million cells [100,101].

1.5.3.6.2. Clinical outcomes

These injections were also shown to provide reduction in pain and improved knee function, with both bone-marrow [94,95,102,103,96] and adipose derived MSCs [104] or the synovial vascular fraction [105]. These observations were sometimes associated with structural changes, such as increased cartilage thickness [95,96], for up to 24 months [104] or increased cartilage quality [102,97,103], which was again sustained at 24 months [106], as assessed by Magnetic Resonance Imaging (MRI) T2 mapping. However, these data were acquired on small numbers of patients, ranging from 4 [94] to 24 [104], and histological analysis of the newly formed cartilage were not provided. In addition, some of the early positive observations made by Davatchi et al. at a 1 year follow-up [94] were shown to have deteriorated at a 60 month follow-up, highlighting the short-term benefit of their treatment [107]. Other studies did not show structural improvements [105].

1.5.3.6.3. Cell dose

Further discrepancies were observed when investigating the effect of cell doses on clinical outcomes. Jo et al. tested the injection of adipose derived autologous MSCs at either 10 million, 50 million or 100 million cells per patient [99]. Significant improvements in pain and knee function were observed only in the high cell dose group, and were accompanied by a significant decrease in cartilage defect size. Interestingly, histological analysis showed the formation of a hyaline-like cartilage [99]. A 2-year follow-up further confirmed these early observations, as both the clinical and structural improvements were maintained only in the high cell dose group [108]. These findings were corroborated by the work of Song et al., who tested the effects of 3 repeated injections of adipose derived MSCs at either 10 million, 20 million or 50 million per patient. All the tested doses led to improvements in pain and knee function, but the increased cartilage volume was longer lasting in the high cell dose group, as reviewed by Lopa et al. [98]. However, contradictory results were obtained by Pers et al., who evaluated the effects of injecting adipose derived MSCs at either 2 million, 10 million or 50 million [109]. Again, all the scores evaluating pain were improved, but significance was only obtained in the low cell dose group. MRI at 4 months revealed possible improvements in only half of the patients, and cell engraftment was observed in only 1 patient [109]. Altogether, these data suggest that improvements in pain and knee function are not necessarily associated with cartilage repair, and clearly highlight a lack of reproducibility in the clinical trials.

1.5.3.6.4. Injections of Mesenchymal Stem Cells with vehicles

As mentioned in the preclinical data section, HA was shown to improve the engraftment of MSCs onto cartilage defects, and was associated with cartilage regeneration. Several groups have therefore tested the effects of injecting MSCs with HA as a vehicle in clinical trials, although with limited success. Gupta et al. tested the effect of a single injection of bone marrow MSCs using either 20 million, 50 million, 75 million or 150 million cells [100].

Improvements in pain scores were observed in the low dose group only, and did not reach significance [100]. Similarly, Lamo Espinosa et al. tested the injection of either 10 million or 100 million cells with HA [110]. Pain scores were improved in both groups, but, contrary to the observations made by Gupta et al., only the high dose group showed better improvements in pain and knee function [110]. The quality of cartilage was also improved in this group only. WORMS scores (see Table 1-1) were maintained at 12 months in the high dose group, while they were shown to be worse than baseline in the low dose and control groups [110].

PRP was also investigated as a vehicle for MSCs injection. Several case series studies have reported improvements in pain and knee function [111,112] for up to 24 months [112] when using SVF in combination with PRP. Additional reports described an increase in cartilage thickness by MRI associated with reduced pain [113], although this observation was not observed on all patients tested in the study by Bansal et al. [112]. An improved or maintained status of cartilage at 24 months was also observed by Koh et al. [114], as assessed by a second look arthroscopy. Finally, partial cartilage regeneration was observed by Bui et al. and associated with improvements in Visual Analogue Score (VAS) and Lysholm score (see Table 1-1) [113]. However, another study reported functional improvements with no relevant changes by MRI [115], and the use of PRP with bone marrow MSCs failed to provide a significant benefit in pain and function in a randomized control trial when compared to MSCs alone [116]. Finally, PRP was tested with SVF in combination with HA [117] and HA + dexamethasone [118] and led to improved clinical outcomes, along with increased thickness of cartilage [117]. However, these data were acquired on only 3 patients and, overall, the studies led on PRP do not provide convincing evidence that this adjuvant would contribute to better cartilage regeneration.

1.5.3.7. Why enhance cell engraftment

Overall, available data from clinical trials and case series studies on the injection of MSCs have only demonstrated short-term effects on pain and knee function. These improvements do not always correlate with structural changes of the damaged tissue, and histological analysis of the newly formed cartilage are lacking in all the reviewed studies but one [99]. Finally, the results obtained by several groups have shown inconsistencies, especially as to which cell dose would be optimal. The fact that the doses used in the above mentioned papers ranged from 2 million to 150 million highlight the need for further investigation, as the use of very high cell doses would increase the cost of such treatments and prevent their development into commercially available therapies.

These inconsistencies in both the outcomes of these clinical trials and the optimal cell dose could be explained by the lack of careful patient selection. Indeed, it seems critical to define, for each clinical trial, a set of criteria to select a homogeneous subpopulation of patients which are likely to respond to the treatment being tested. This would not only provide better clinical outcomes, but would also help define the best endpoints to measure therapeutic benefits. So far, no such work on patients stratification has been described.

Another limitation which could explain the lack of reproducibility in the results obtained from these clinical trials is that none of these studies have developed strategies to retain the stem cells at the site of their injection, where their properties are the most needed. The use of vehicles such as PRP and HA have not provided clear additional benefits, highlighting the need to develop more effective delivery methods to achieve better clinical outcomes, as suggested by others [119].

Generally, only 1 to 5 % of MSCs were shown to engraft onto their target tissue, irrespective of their delivery routes [120], and when these cells were injected in cartilage, they were shown

to disappear from the sites of injections only after a few days by Pers et al [120]. This poor cell retention could be due to mechanical damage during injections [119], cell death by apoptosis, recognition by the immune system [121] or leakage to surrounding tissues [119]. Several authors have therefore expressed the need to not only investigate the fate, behaviour and kinetics of MSCs following their injection [122,123], but also to develop methods to enhance their attachment onto the damaged tissue and promote their long-term engraftment [119,124]. Achieving this objective would allow for the full exploitation of the potential of MSCs, and might lead to a reduction of the required cell dose. Notably, improved engraftment of the injected MSCs in preclinical models led to drastic positive effects on cartilage repair, with better cartilage regeneration mediated by MSCs differentiation. Several methods have therefore been investigated in an attempt to artificially increase the retention of MSCs at the site of injection.

1.5.3.8. Strategies for artificial enhancement of Mesenchymal Stem Cells engraftment

In the work by Koga et al. [125], both rabbit and human MSCs were injected, *ex vivo*, on chondral defects facing upwards and immobilised for 10 minutes. This immobilisation led to the attachment of 60 % of the cells, a drastic improvement compared to previously observed adhesion efficiencies. The use of this method *in vivo* in rabbits showed a higher number of cells present in the defects compared to control rabbits which underwent normal injection. Cells were shown to be attached 1 day after injection following immobilisation of the joint, leading to a reduced cartilage defect at 24 days as well as to better histological scores.

Another strategy was developed for larger animals which relies on the use of magnetically-labelled MSCs coupled to an external magnetic device [124]. Mini pigs were submitted to this injection, while control groups were injected with either MSCs or saline with their patella facing upwards. An arthroscopy showed the immediate and specific accumulation of the

magnetically-labelled MSCs at the site of damage. Cells were retained after the external device was removed, and were shown to be fully integrated and differentiated in the tissue 1 week after their injection. These observations were associated with better cartilage repair compared to the control groups at 24 weeks, as the newly formed cartilage appeared white and smooth. Better histological scores were obtained for this hyaline-like tissue compared to that formed in the control groups, which were shown to be fibrous and hypertrophic [124].

Interestingly, in both of these studies, the improved attachment of either human MSCs on *ex vivo* cartilage implants, or of the magnetically labelled MSCs was shown to be mediated by an increase in the expression of cell adhesion molecules such as integrins [124] and intercellular adhesion molecule 1 [125].

Similarly, a paper published by the company Xintela [126] reported the use of Fluorescence Activated Cell Sorting (FACS)-sorted adipose derived equine MSCs presenting the collagen binding integrin α10β1 at their surface. The sorted cells were shown to retain the trilineage potential of unsorted MSCs as well as their immunomodulatory properties. Sorted cells also demonstrated a higher chondrogenic potential in pellet cultures, and a significantly higher adhesion capacity to both chondral and osteochondral defects *ex vivo* when compared to the unsorted population [126].

Altering the expression of molecules present at the cell surface was therefore pursued by other groups in an attempt to improve cell adhesion onto cartilage. Dennis et al. described the use of Palmitated Protein G as an intermediate to coat chondroprogenitors with various antibodies directed against matrix molecules [127]. The presentation of these antibodies at the cell surface led to a higher adhesion of the coated cells to a rabbit articular cartilage explant with a partial thickness.

However, these last two strategies do not specifically target MSCs towards damaged cartilage, as the matrix molecules used as baits to direct MSCs are also present in healthy

cartilage, although they might be less exposed. Other groups have therefore developed strategies to target MSCs to molecules which are only present in OA.

In a study by Leijs et al. [128], authors observed that, *in vitro*, MSCs migrated more towards OA synovium and OA cartilage Conditioned Medium (CM) than towards control medium. They thought that the inflammatory cytokines expressed in OA were recruiting MSCs. Interestingly, they also showed that MSCs migrated more towards OA synovium CM than OA cartilage CM [128]. Authors attempted to enhance the migration of MSCs towards OA tissues *in vivo* by pre-treating them with the inflammatory cytokine TNFα or by cultivating them in hypoxia, in an attempt to increase the expression of their cell surface receptors [128]. However, this pre-treatment did not influence the migration of cells injected *in vivo*. In addition, cells were shown to engraft to the synovium or to be present as free MSCs in the joints, and very few cells were observed onto the damaged cartilage [128].

This study demonstrates that the observed migration of MSCs *in vitro* is not recapitulated *in vivo*, and that the homing potential of MSCs is not enough to have them adhere onto the damaged tissue. It also shows that targeting inflammatory molecules expressed in OA might not be a successful strategy to improve their retention.

Taking all these observations together, our group decided to develop a strategy in which the adhesion of MSCs towards a matrix molecule present only in OA cartilage would be improved. Instead of targeting native type II collagen or other matrix molecules, as described by others, our strategy was to target a molecule which is specifically present in damaged cartilage: denatured type II collagen.

1.6. Type II collagen denaturation in Osteoarthritis

Indeed, several reports have described the increased cleavage, unwinding and subsequent degradation of type II collagen in OA [129,130,131]. Dodge et. al. were the first to demonstrate this phenomenon using an antibody directed against epitopes corresponding to the unwound helix of type II gelatin. Human arthritic tissues displayed stronger staining than healthy ones, especially at the superficial and middle zones [130]. These observations were later confirmed by Hollander et al., who have developed a new antibody, namely Col2-3/4-m, which specifically recognises an 11 amino acid-long epitope of type II collagen, which is only exposed when the helix is unwound [129]. This antibody was also able to recognise heat-denatured type II collagen [129]. The increased levels of denatured type II collagen were observed mostly at the surface of articular cartilage [132] in both ageing and OA cartilage, a localisation which also correlates with the presence of the matrix metalloproteinases MMP-1 and MMP-13 [131]. Both enzymes were shown to be able to cleave this substrate *in vitro*, at the same site (Gly-Leu/Ile bond), generating the characteristic ¼ and ¾ fragments [133,134] , but MMP-13 was shown to cleave this substrate more quickly [133] (Figure 1-6).

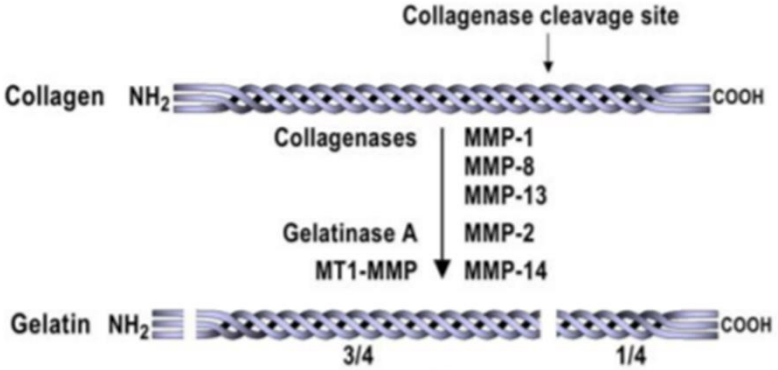


Figure 1-6: Cleavage of collagen by Matrix Metalloproteinases. The collagenases MMP-1, MMP-8 and MMP-13 have been reported to be the prevalent enzymes involved in the cleavage of type II collagen in OA, although MMP-14 and MMP-2 can also exhibit some collagenolytic activities. Cleavage of the collagen molecule generates a ¾ N terminal and a ¼ C terminal fragment, and several data have reported that this cleavage leads to the unwinding of its helix. Gelatin can be further degraded by the gelatinases MMP-2 and MMP-9. Figure adapted from [209].

The prevalence of these enzymes, together with MMP-8, in the degradation of type II collagen in OA was further confirmed in another study [135] which relied on the use of antibodies directed against epitopes found in the fragments generated by the cleavage of collagen by these enzymes. The authors observed increased levels of these epitopes in human OA samples compared to normal tissues, highlighting the importance of collagenases in the degradation of type II collagen in OA. Notably, a correlation was observed between the levels of degraded and denatured collagen [135], the latter being detected by the previously mentioned Col2-3/4-m antibody. The fact that this antibody was able to recognise the ¾ fragment of cleaved type II collagen [129] further reinforces the assumption that cleavage of type II collagen by collagenases leads to its denaturation (unwinding). The fact that chondrocytes in both normal and OA tissues express MMP-1, 8 and 13 [135], coupled to the observation that the staining of denatured type II collagen was mostly found in the pericellular region [132], suggest that cells from the cartilage are directly responsible for the increased breakdown of type II collagen observed in OA.

Altogether, these studies have demonstrated that increased amounts of denatured type II gelatin are present in OA cartilage. Importantly, in normal human articular cartilage, only 1.1 % of type II collagen is in its denatured form [136], against 6 % in OA cartilage [136], which justifies the use of this protein as a bait to target MSCs specifically to damaged cartilage. The fact that type II gelatin degradation starts at the articular surface, before progressing to deeper areas with disease progression [132] adds to the rationale of this strategy.

Interestingly for the purpose of targeting MSCs towards this substrate, collagen fragments generated after its cleavage naturally denature into the non-helical gelatin to become substrates of various MMPs [135].

1.7. Matrix Metalloproteases

1.7.1. Matrix Metalloproteases generalities

Matrix Metalloproteases (MMPs) are a family of 23 proteins [137,138] which belong to the larger group of metzincins [138]. They have been further divided into the subgroups of collagenases (MMPs-1, 8 and 13), stromelysins (MMPs-3 and 10), a more heterogeneous subgroup containing matrilysin (MMP-7), enamelysin (MMP-20), the macrophage metalloelastase (MMP-12) and MMP-19, membrane-type metalloproteinases (MT-MMPs-1 to 4 and stromelysin-3, MMP-11), and gelatinases (MMPs-2 and 9) [139]. Its members are characterized by the presence of a conserved zinc ion in the catalytic site of the enzymes [140,141]. MMPs are multidomain proteins, sharing the same structural features [138,142] (Figure 1-7).

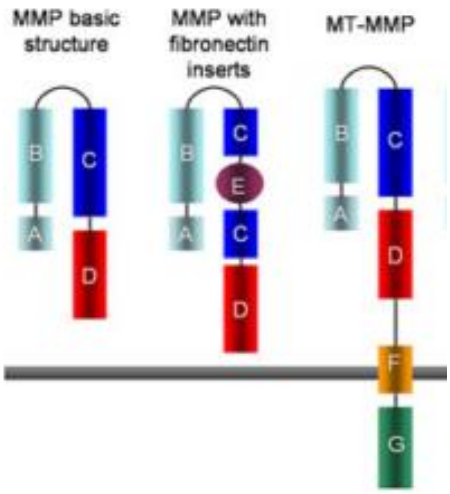


Figure 1-7: Structural organisation of MMPs. MMPs contain a signal peptide (A) followed by a prodomain (B), a catalytic domain (C) and a hemopexin domain (D). MMP-2 and MMP-9 contain three additional Fibronectin-like domains (E) which are inserted in the catalytic one. MT-MMPs also contain a transmembrane domain (F) and a cytoplasmic tail (G). Figure adapted from [138].

These common features include:

- A prodomain which keeps them in an inactive state through the interaction of a conserved cysteine with the zinc ion of their catalytic site; disruption of this interaction leads to the activation of the MMP via the so called cysteine-switch mechanism.
- The catalytic site which is responsible for the proteolysis activity of the proteases. It is connected to the hemopexin domain via a proline-rich linker.
- The C-terminal hemopexin domain, which confers substrate specificity [140]. It is absent from MMP-7 [139] and is required for the cleavage of triple helical collagens by collagenases [138].

Among this family, MMP-2 and MMP-9 (also known as gelatinase A and B, respectively) are unique in that they contain 3 additional and contiguous Fibronectin type II-like domains. These modules are also found in blood clotting factor XII, bovine seminal fluid proteins PDC 109 and BSP-A3, Insulin-like Growth factor II receptor, the mannose 6 phosphate receptor, the mannose receptor and the phospholipase A2 receptor [143].

Both gelatinases share the following substrates: type IV, type V, type VII and type X collagens, fibronectin, elastin, and all types of denatured collagens [144]. Their Fibronectin type II-like domains are inserts into their catalytic domain and resemble the collagen binding regions of the type II repeats found in Fibronectin [138]. In MMP-2, the Fibronectin type II-like domains 1, 2 and 3 span from residues 222 to 276, 279 to 333 and 337 to 391, respectively (Figure 1-8) and form the Collagen Binding Domain (CBD).

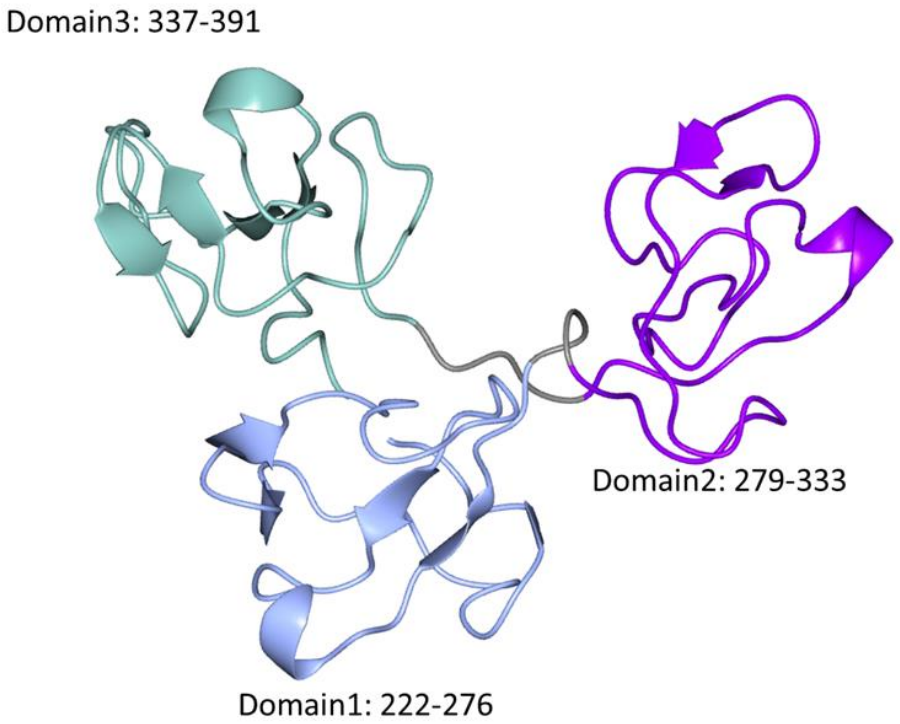


Figure 1-8: Cartoon representation of the CBD of MMP-2 derived from the published structure of MMP-2 (Protein Data Bank (PDB) Accession Number 1CK7) using Pymol. The sequence of the CBD (residues 222 to 391) was extracted from the published structure of MMP-2 to isolate the three Fibronectin-type II like domains.

1.7.2. The Collagen Binding Domain of Matrix Metalloprotease-2

The combined three Fibronectin-type II like domains of MMP-2, also referred to as the Collagen Binding Domain (CBD), have been extensively studied over the past few years. It has been shown to interact with both native and denatured type I collagen, denatured type IV and V collagen and elastin [141]. A weak interaction has also been identified for native type IV and V collagen as well as laminin and heparin [141]. Another paper from the same group [145] has also shown that the CBD is able to bind native and denatured type III collagen, denatured type IV, denatured and native type V and denatured type X collagens. However, the CBD does not bind to Secreted Protein Acidic and Cysteine Rich (SPARC), fibronectin, nor to

matrigels films or gels [141], although some binding to denatured matrigel films was observed [145].

The interaction of the CBD with these substrates is thought to be critical for the catalytic activity of the full length MMP-2. When deleted from the CBD, the activity of MMP-2 towards type I gelatin was indeed reduced to 10 % [146], and soluble CBD was shown to inhibit the gelatinolysis of MMP-2 [137,147], demonstrating the role of the CBD in substrate positioning.

It was also shown that all the collagen binding properties of MMP-2 could be attributed to the CBD, as a mutant deleted from this domain was unable to interact with type I collagen [146], or to inhibit binding of MMP-2 to gelatin [148]. The fact that soluble recombinant CBD inhibited the binding of MMP-2 towards gelatin in a concentration-dependent manner further reinforces this theory [137]. However, both full-length MMP-2 and 9 were shown to bind more tightly to gelatin than their isolated CBDs [148,149], suggesting that the other domains within the enzyme stabilize the interaction with the ligand.

Interestingly for the purpose of this project, the CBD of MMP-2 was shown to bind with a Kd of 50 nM to type II gelatin in a plate binding assay [145], but did not bind to native type II collagen [145]. This specificity towards denatured type II collagen justifies the choice of working with this CBD instead of that from MMP-9. The latter binds with a slightly higher affinity to type II gelatin (Kd of 8.0 nM), but also showed strong binding to its native form (Kd of 50 nM) [148].

1.8. Aim of the project

This project poses the hypothesis that targeting MSCs to damaged cartilage following their injection into OA knees would favour their engraftment and improve cartilage repair compared to previously published studies.

In order to do so, our aim was to target MSCs towards denatured type II collagen (*i.e.* type II gelatin), which is specifically found at the surface of OA articular cartilage. The strategy for targeting MSCs towards this substrate relies on coating their cell surface with a protein which binds avidly to type II gelatin: the CBD of MMP-2.

In order to achieve this aim, this protein was first expressed, purified, and its affinity for type II gelatin quantified. The following aim was to further characterise its interaction, as well as that of its individual modules, with type II gelatin by Nuclear Magnetic Resonance (NMR), in order to identify the most potent module for binding to this substrate. Next, chimeric proteins made of 3 identical modules were generated and a mutant protein with increased affinity for type II gelatin was obtained. The coating of MSCs with this protein was then tested, optimised and validated, before testing the ability of coated MSCs to bind to type II gelatin *in vitro.*

2. Materials and Methods

2.1. Materials

2.1.1. Equipment

Polymerase Chain Reactions (PCR) were performed using the Mastercycler from Eppendorf.

E.coli cells were grown in shaking incubators from New Brunswick Scientific. Optical Density (OD) measurements were performed using either a WPA Biowave instrument or the UV-Vis spectrometer CARY 300-BIO (Varian). Centrifugation steps were carried out using either the benchtop centrifuge Minispin (Eppendorf), the Beckman Coulter Allegra X-30R centrifuge or the Sorvall centrifuge from Thermo Scientific.

Proteins were purified using either the AKTA purifier or AKTA start systems from GE Healthcare.

Protein samples were denatured using a heat block CH-100 from BioSan. Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis (SDS PAGE) gels were ran using the Power Pac 300 from Biorad and scanned using the Image Scanner III from GE Healthcare.

Plates prepared for the binding assay were dried using a UNITEMP incubator. Absorbance signals for the assay were read using the Multiskan Ascent plate reader from Labsystems.

The pH of all buffers was measured using the JENWAY 3505 pH meter.

All cell culture was performed in class II laminar flow hoods (NUAIRE) and incubated in humidified incubators (NUAIRE) at 37°C and 5 % carbon dioxide.

2.1.2. DNA sequences

The following DNA sequences were ordered from Invitrogen. They contain no start and stop codons as both of them were provided by the vectors used for expression.

Module 1 (Residues 223 - 282):
gtgaaatatggtaatgccgatggcgaatattgcaaatttccgtttctgtttaacggcaaagagtataatagctgtaccgataccggt
cgtagtgatggttttctgtggtgtagcaccacctataactttgagaaagatggcaaatatggcttttgtccgcatgaagcactgttta
ccatg

Module 2 (Residues 278 – 336):
gcactgtttaccatgggtggtaatgcggaaggtcagccgtgtaaatttccttttcgctttcagggcaccagctatgatagttgtacca
ccgaaggtcgtaccgatggttatcgttggtgcggtacaaccgaagattatgatcgtgacaaaaaatacggtttctgtccggaaacc
gca

Module 3 (Residues 337– 394):
atgagcaccgttggtggtaatagtgaaggtgcaccgtgtgtttttccgtttacctttctgggtaacaaatatgaaagctgtaccagtg
caggtcgttcagatggtaaaatgtggtgcgcaaccaccgcaaattatgatgatgatcgtaaatggggtttttgcccagatcagggt

CBD (Residues 222 – 396):
Cgtgtgaaatatggtaatgccgatggcgaatattgcaaatttccgtttctgtttaacggcaaagagtataatagctgtaccgatacc
ggtcgtagtgatggttttctgtggtgtagcaccacctataactttgagaaagatggcaaatatggcttttgtccgcatgaagcactgt
ttaccatgggtggtaatgcggaaggtcagccgtgtaaatttccttttcgctttcagggcaccagctatgatagttgtaccaccgaag
gtcgtaccgatggttatcgttggtgcggtacaaccgaagattatgatcgtgacaaaaaatacggtttctgtccggaaaccgcaatg
agcaccgttggtggtaatagtgaaggtgcaccgtgtgtttttccgtttacctttctgggtaacaaatatgaaagctgtaccagtgcag
gtcgttcagatggtaaaatgtggtgcgcaaccaccgcaaattatgatgatgatcgtaaatggggtttttgcccagatcagggttata
gc

111:
gaaggtcaggttgttcgtgtgaaatatggtaatgcagatggcgagtattgcaaatttccgtttctgtttaacggcaaagagtataat
agctgtaccgataccggtcgtagtgatggttttctgtggtgtagcaccacctataactttgagaaagatggcaaatatggcttttgtc
cgcatgaagcactgtttgtgatgtatggcaatgccgatggtgaatactgtaaattcccatttctgttcaatggtaaagaatacaactc
atgcaccgatacaggccgttcagatggctttcgttggtgttcaaccacctacaatttcgaaaaagacggtaagtatggtttctgtccg
gaaaccgcagtgcgcgttaaatatggcaacgcggacggggaatattgtaagtttccgttcttattcaacgggaaagaatataaca
gttgcacagacaccggtcgctcagatggttttttatggtgctcaacaacgtataacttcgaaaaggatgggaagtacggattttgtc
cggatcagggttatagcctg

222:
Gaaggtcaggttgtgtttaccatgtatggtaatgccgaaggtcagccgtgtaaatttccgtttcgttttcagggcaccagctatgata
gttgtaccaccgaaggtcgtaccgatggttatcgttggtgtggtacgaccgaagattatgatcgtgataaaaagtatggcttttgtcc
gcatgaagccctgtttacaatgggtggcaatgcagagggccagccttgcaaattccctttttcgcttccagggtacatcttatgattca
tgcacaacggaaggtcgcacagatggctaccgctggtgcggcaccacagaggattatgaccgcgacaaaaaatacggtttttgtc
cggaaaccgcactgttcaccatgggtggtaatgcggaaggacaaccatgcaagtttccattccgctttcagggaacctcatatgat
agctgcacaacagagggacgtacggatggatacagatggtgcggtacaaccgaggactacgatagagataagaaatatggtttc
tgtcccgatcagggttatagcctg

333:
Gaaggtcaggttgttagcaccgtttatggtaatagcgaaggtgcaccgtgtgtttttccgtttacctttctgggtaacaaatatgaaa
gctgtaccagcgcaggtcgtagtgatggttttctgtggtgtgcaaccaccgcaaattatgatgatgatcgtaaatggggtttttgtcc
gcatgaagcactgtttaccatgggtggcaattctgaaggtgccccttgcgtgtttccttttacatttttaggcaacaagtacgaaagc
tgcacctcagccggtcgttcagatggtaaacgttggtgcgccaccacagccaactatgatgacgacagaaaatggggcttctgtcc
tgaaaccgcaatgagcaccgttggtggcaacagtgaaggcgctccatgcgttttcccgttcacattcctgggcaataaatacgaat
catgtacctctgcaggtcgctctgatggcaaaatgtggtgcgcgacaacggccaattacgacgatgaccgcaagtggggctttgc
ccagatcagggttatagcctg

2.1.3. Protein sequences

The DNA sequences presented in section 2.1.2 encoded for the following protein sequences

CBD: R V K Y G N A D G E Y C K F P F L F N G K E Y N S C T D T G R S D G F L W C S T T Y N F E K D G K Y G F C P H E A L F T M G G N A E G Q P C K F P F R F Q G T S Y D S C T T E G R T D G Y R W C G T T E D Y D R D K K Y G F C P E T A M S T V G G N S E G A P C V F P F T F L G N K Y E S C T S A G R S D G K M W C A T T A N Y D D D R K W G F C P D Q G Y S

Module 1: VKYGNADGEYCKFPFLFNGKEYNSCTDTGRSDGFLWCSTTYNFEKDGKYGFCPHEALFTM

Module 2: ALFTMGGNAEGQPCKFPFRFQGTSYDSCTTEGRTDGYRWCGTTEDYDRDKKYGFCPETA

Module 3:MSTVGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGKMWCATTANYDDDRKWGFCPDQG

Trimodular proteins containing repeats of identical modules were then designed with the aim of maintaining the overall structure observed in the CBD. In the sequences below, residues in blue are those conserved from the original CBD in order to maintain intramolecular interactions, residues in black are linkers and residues in red are from each module.

111: E G Q V V R V K Y G N A D G E Y C K F P F L F N G K E Y N S C T D T G R S D G F L W C S T T Y N F E K D G K Y G F C P H E A L F V M Y G N A D G E Y C K F P F L F N G K E Y N S C T D T G R S D G F R W C S T T Y N F E K D G K Y G F C P E T A V R V K Y G N A D G E Y C K F P F L F N G K E Y N S C T D T G R S D G F L W C S T T Y N F E K D G K Y G F C P D Q G Y S L

222: E G Q V V F T M Y G N A E G Q P C K F P F R F Q G T S Y D S C T T E G R T D G Y R W C G T T E D Y D R D K K Y G F C P H E A L F T M G G N A E G Q P C K F P F R F Q G T S Y D S C T T E G R T D G Y R W C G T T E D Y D R D K K Y G F C P E T A L F T M G G N A E G Q P C K F P F R F Q G T S Y D S C T T E G R T D G Y R W C G T T E D Y D R D K K Y G F C P D Q G Y S L

333: E G Q V V S T V Y G N S E G A P C V F P F T F L G N K Y E S C T S A G R S D G F L W C A T T A N Y D D D R K W G F C P H E A L F T M G G N S E G A P C V F P F T F L G N K Y E S C T S A G R S D G K R W C A T T A N Y D D D R K W G F C P E T A M S T V G G N S E G A P C V F P F T F L G N K Y E S C T S A G R S D G K M W C A T T A N Y D D D R K W G F C P D Q G Y S L

2.2. Methods

2.2.1. Cloning

The InFusion (Clontech) cloning technology is a ligase independent cloning method which enables an insert to be inserted into a linearized vector if each share 15-nucleotide-long complementary sequences at both extremities (Figure 2-1).

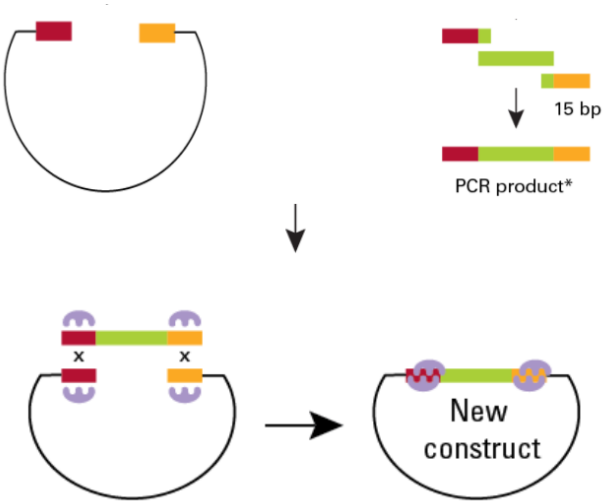


Figure 2-1: Principle of the InFusion cloning. The Infusion cloning is a ligase independent method that relies on the recombination between the insert and the linearised vector when both share 15-nucleotide-long complementary sequences at their extremities. Figure adapted from the Infusion manufacturer's manual

This was the cloning method chosen for all constructs and it was used according to the manufacturer's recommendations. The vectors used were all from the pOPIN suite developed by OPPF, Oxford.

2.2.1.1. Construct design

The synthetic genes encoding for the above mentioned protein sequences were purchased from Invitrogen and codon-optimised for expression in *E.coli*. Primers were ordered from Sigma and were used to subclone the genes into either the pOPINM or pOPINS vectors (Figure 2-2), which were provided by OPPF, Oxford. The sequences of the primers can be found in Table 2-1.

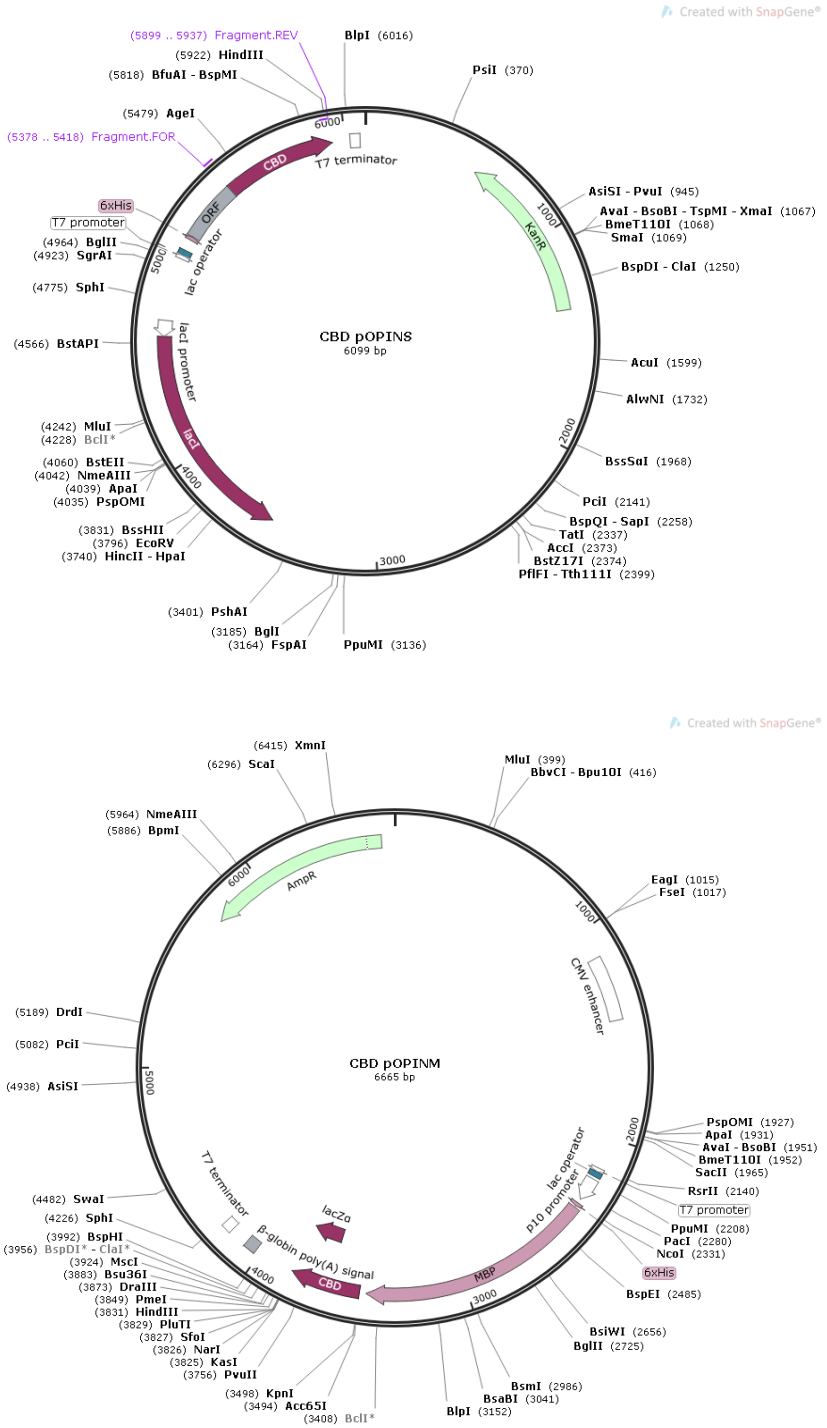


Figure 2-2: Maps of the vectors pOPINS and pOPINM. Both vectors were obtained from OPPF, Oxford. The pOPINM vector carries a hexahistidine tag followed by the Maltose Binding Protein (His_6-MBP tag) and the cleavage site of the 3C protease. The pOPINS carries the SUMO tag downstream of the SUMO protease cleavage site. The gene encoding for the protein was cloned downstream of each of these tags, as indicated in this Figure for the CBD gene.

Table 2-1: Sequences of the primers used for cloning. The sequences in red are vector specific and were used as instructed on the OPPF website. Sequences in black are insert-specific.

Protein	Vector	Forward (5' to 3')	Reverse (5' to 3')
CBD	pOPINM	AAGTTCTGTTTCAGGGCCCGCGTGTGAAATATGGTAATGCCGAT	ATGGTCTAGAAAGCTTTAGCTATAACCCTGATCTGGGCA
	pOPINS	GCGAACAGATCGGTGGTCGTGTGAAATATGGTAATGCCGAT	
Module 1	pOPINS	GCGAACAGATCGGTGGTGTGAAATATGGTAATGCC	ATGGTCTAGAAAGCTTTACATGGTAAACAGTGCTTC
Module 2	pOPINS	GCGAACAGATCGGTGGTGCACTGTTTACCATG	ATGGTCTAGAAAGCTTTATGCGGTTTCCGGACAGAA
Module 3	pOPINS	GCGAACAGATCGGTGGTATGAGCACCGTTGGT	ATGGTCTAGAAAGCTTTAACCCTGATCTGGGCAAAA
111	pOPINS	GCGAACAGATCGGTGGTGAAGGTCAGGTTGTTCGTGTG	ATGGTCTAGAAAGCTTTACAGGCTATAACCCTGATCCGG
222	pOPINS	GCGAACAGATCGGTGGTGAAGGTCAGGTTGTGTTTACC	
333	pOPINS	GCGAACAGATCGGTGGTGAAGGTCAGGTTGTTAGCACC	

The amplification of the synthetic gene with the correct extremities for cloning into each vector was performed with the PCR program described in Table 2-2.

Table 2-2: PCR program used for amplification of the gene of interest with the correct extremities. The annealing temperature was adjusted for each PCR reaction, and set 5 degrees below the lowest melting temperature of both primers.

Step	Temperature (°C)	Time (min)	
Initial denaturation	98	2	
Denaturation	95	0.5	29 cycles
Annealing	N/A	0.5	
Amplification	72	1.5	
Final amplification and clearance	72	7	

The amplified products were then ran on an agarose gel electrophoresis (see section 2.2.1.3).

2.2.1.2. Vector linearisation

5 µg of the pOPINS vector were digested with the enzymes HindIII HF and KpnI HF (NEB) for 6 hours at 37°C in the provided CutSmart buffer (NEB). The reaction was terminated by the addition of 0.5M Ethylenediaminetetraacetic acid (EDTA), pH 8.0. The success of the linearisation was confirmed with an agarose gel electrophoresis (see section 2.2.1.3). The pOPINM vector had already been linearised before.

2.2.1.3. Agarose gel electrophoresis

A 1 % agarose gel was run to confirm either the successful amplification of the construct or the linearisation of the plasmid. Briefly, 1 g of agarose (Bioline) was dissolved into 100 mL of TAE buffer (2M Tris, 0.1M EDTA, 1M acetic acid) and heated. Once cooled down, the solution was used to cast a gel and either 10 µL of 0.5 µg/ml ethidium bromide or 5 µL of Mildori Green (NIPPON genetics) was added and mixed. Once set, 5 µL of each sample, previously mixed with the same amount of loading buffer (NEB), was loaded and the gel was run at 120 V for 30 minutes. The visualisation was performed under UV light.

2.2.1.4. In Fusion cloning (Clontech)

Briefly, 5 µL of the PCR product (amplified synthetic gene) was purified by adding 2 µL of Cloning Enhancer and incubating for 15 minutes at 37°C and for a further 15 minutes at 80°C.

The In Fusion reaction was then prepared with the corresponding 50-100 ng of linearised pOPIN vector, 1 µL of cleaned PCR product, 1 µL of In Fusion enzyme, in a final volume of 5 µL.

The mixture was incubated for 15 minutes at 50°C before proceeding to bacteria transformation.

2.2.1.5. Bacteria transformation

Briefly, 50 µL aliquots of competent cells were thawed on ice before adding 1.5 µL of DNA to the cells. After incubating for 30 minutes on ice, bacteria were heat-shocked for 30 seconds at 42°C. Cells were put back on ice for 3 minutes before adding 450 µL of either S.O.C or Lysogeny broth (LB, Melford) medium and incubating for 1 hour at 37°C with shaking.

The samples were then spun at 4 000 g for 30 seconds and resuspended in 50 µL before being spread onto LB agar plates containing the appropriate antibiotic for selection (Kanamycin: Gibco; Ampicillin: Melford). The plates were incubated at 37°C overnight and then kept at 4°C for a maximum of one week.

Either Stellar (Clontech) or XL1 blue (Agilent) *E.coli* competent cells were submitted to this process for plasmid amplification. DNA was extracted using the miniprep kit from Qiagen. All constructs were confirmed by sequencing.

The same protocol was followed with various expression strains for protein expression.

2.2.2. Protein expression in *E.coli*

2.2.2.1. pET expression system

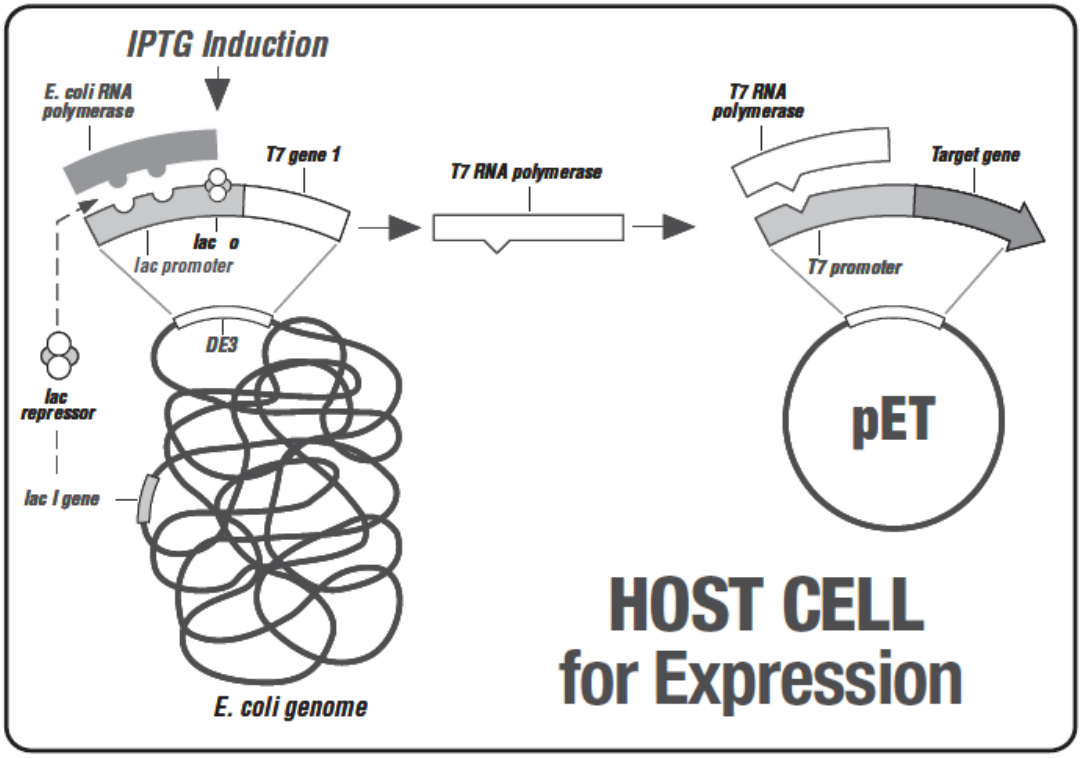


Figure 2-3: Principle of the pET expression system. The gene of interest (target gene) is placed under a T7 promoter. The gene encoding for the T7 Ribonucleic Acid (RNA) polymerase is itself under the control of a lac promoter, which is inducible with Isopropyl-ß-D-thiogalactopyranoside (IPTG). Addition of IPTG will lead to the transcription of the gene encoding for the polymerase, and the subsequent transcription of the gene of interest. Figure from Novagen user's manual.

One of the most commonly used vector systems is the pET expression system, which places the gene of interest under the control of a T7 promoter, which is itself recognised by the phage T7 RNA polymerase (Figure 2-3). The gene encoding for the polymerase is usually inserted in the bacterial genome through a prophage named lambda DE3, and is placed under the control of a lac UV5 promoter, which is inducible with lactose or Isopropyl-ß-D-thiogalactopyranoside (IPTG). Addition of IPTG to the culture will enable expression of the T7 RNA polymerase, which will then be able to start the transcription of the gene of interest. In order to avoid basal leaky expression of the T7 RNA polymerase, a T7 lysozyme can be

provided by an additional plasmid (either pLys S or pLys E), which will limit the transcription by the T7 RNA polymerase without inducer [150].

2.2.2.2. Protein expression trials – Small scale

Either BL21 (DE3) (Novagen) (for expression with the pOPINB and pOPINS vectors), Rosetta-pLysS (DE3) (Novagen) (for expression with the pOPINM vector) or Shuffle T7 express (NEB) were transformed on selective LB agar plates.

When required, minimal media was prepared by adding a filter-sterilised solution B (0.1 M $MgSO_4$, 7.5 mM $CaCl_2$, 0.75 mM $MnCl_2$, 0.25 mM $FeSO_4$ supplemented with Glucose (20% w/v)) to an autoclaved solution A (100 mM Na_2HPO_4, 40 mM KH_2PO_4, 18.3 mM ^{15}N NH_4Cl).

1 colony from the transformation plate was then picked and inoculated into 10 mL of either LB or minimal media containing the appropriate antibiotics and the culture was left to grow overnight at 37°C. 10 µL of the overnight culture was then inoculated into 10 mL fresh media and grown at 37°C before inducing with various concentration of IPTG (Apollo Scientific) when the optical density at 600 nm (OD_{600}) reached 0.6. Cultures were then maintained at 18°C overnight or at 37°C for 5 hours before being harvested by spinning at 5000 rpm for 5 minutes. The cells were then subjected to one cycle of freeze-thaw before being lysed in 1 mL of lysis buffer (20 mM Na_2HPO_4, 500 mM NaCl, 20 mM Imidazole, 1mg/mL lysozyme (Sigma), 1 % Tween 20 (v/v) (Sigma), DNAse I (Roche) and protease inhibitors (Roche), pH 7.4) for 1 hour. The lysate was then cleared by spinning at 13 300 rpm for 10 minutes. The insoluble pellet was resuspended in 1 mL lysis buffer containing 8 M urea and spun at 13 300 rpm for 10 minutes. The supernatant was kept and both the soluble and insoluble fractions were purified using the Ni-Nta agarose resin (see section 2.2.4.1).

2.2.2.3. Protein expression– Large scale

E.coli expression strains were transformed as described above with the appropriate expression vector and selected on agar plates containing the corresponding antibiotics. One colony was inoculated in 10 mL of LB containing antibiotics overnight. 5 mL of the overnight culture was used to inoculate 500 mL of media containing appropriate antibiotics into a 2L flask. Growth was carried out at 37°C, 180 rpm, until reaching an OD at 600 nm of 0.7-0.9. Cells were induced with the following concentrations of IPTG:

- CBD: 600 µM
- 222: 1000 µM
- 333: 800 µM,
- Module 1: 400 µM
- Module 2: 20 µM
- Module 3: 20 µM

and the production performed overnight at 18°C, except for 333, which was left for 5 hours at 25°C following induction. Cells were harvested at 8 000 g for 15 minutes and lysed by sonication in 20 mL lysis buffer (20 mM Na_2HPO_4, 500 mM NaCl, 20 mM Imidazole, 1mg/mL lysozyme (Sigma), 1% Tween 20 (v/v), DNAse I (Roche) and protease inhibitors (Roche), pH 7.4). The lysate was cleared by spinning at 18 000 g for 40 minutes before being filtered and purified on a His Trap Fast Flow 5 mL column (HiFIIQ NiNTA) using an AKTA system (GE healthcare) (see section 2.2.5.1).

2.2.2.4. Protein refolding

As 111 was shown to be highly insoluble, the refolding protocol described by Steffensen et al. [141] was tested. Briefly, the insoluble pellet was sonicated before being washed in 100 mM Tris·HCl, pH 7.0, 5 mM EDTA, 2 M Urea, 2 % (v/v) Triton X-100 and spun for 30 min at 22 000 g, 4°C. This washing step was repeated 4 times. The pellet was then resuspended in 100 mM Tris·HCl, pH 7.0, 5 mM EDTA, sonicated and spun at 22 000 g, 4°C for 30 min. The washed pellets were then stored at -80°C until use.

Once thawed, the pellets were then resuspended in 6M GdnHCl, 0.1M sodium phosphate, 10 mM Tris, pH 8.0, sonicated and incubated for 1 hour at room temperature rotating. The non-dissolved aggregates were removed by centrifugation at 12 000 g for 15 min at 4°C. The supernatant was then dialysed from 2 hours to overnight against aerated 0.1 M sodium borate buffer, pH 10.0. The sample was then diluted and dialysed against 4L of 100 mM NaP, 0.5 M NaCl, pH 8.0 at 4°C. The sample was then spun at 12 000 g for 15 min at 4°C, dialysed against His Trap A buffer (20 mM Na_2HPO_4, 20 mM Imidazole, 500 mM NaCl, pH 7.4) and subjected to the same purification protocol as the other constructs (see section 2.2.5).

2.2.3. Protein expression in HEK cells

One of the main cell lines used for protein expression is the Human Embryonic Kidney (HEK) 293. It was established by the transformation of human embryonic kidney cells with sheared human adenovirus DNA. The HEK 293T cell line was derived from the original HEK 293 one by integrating genes encoding the Simian Virus 40 (SV40) Large antigen T (LT). This genetic modification has enabled transient gene expression methods to be developed. The SV40 LT is indeed able to bind to expression vectors containing the SV40 origin of replication, enabling activation of the DNA replication machinery and episomal replication of the recombinant plasmid, leading to high copy numbers of recombinant DNA and, subsequently, higher protein yields [151,152]. These elements are also thought to favour nuclear import of the expression plasmid, making more of it available for transcription [152].

2.2.3.1. Cell transfection with Lipofectamine

HEK 293T cells were seeded in 10-cm dishes at 3.5 x 10^6 cells. 48 hours later, they were transfected with Lipofectamine 2000 (ThermoFisher Scientific), according to the manufacturer's instructions. Cells were harvested 24 hours after transfection.

Considering the cost of Lipofectamine 2000, more transfection trials were carried out using either Calcium Chloride or Polyethylenimine (PEI) as the transfection reagents, which are cheaper, in preparation of large scale experiments.

2.2.3.2. Cell transfection with Polyethylenimine

HEK 293T cells were seeded at 600 000 cells/well in 6-well plates 24 hours before being transfected. The PEI solution (25kDa, linear, Polysciences Inc.) was prepared by dissolving 100 mg in 70 mL of milliQ water, heating to 80 ^oC and allowing to cool down at Room Temperature (RT). The pH was adjusted to 7.0, the volume brought up to 100 mL and the mixture filter sterilized. 3.5 µg of DNA was mixed with OptiMEM while PEI was added to OptiMEM in a different tube. 3 ratios of DNA:PEI were tested. PEI was then added to the DNA, the solution mixed and incubated for 10 minutes at RT before adding the mixture to the cells (300 µL/well).

Cells were harvested either 24, 48 or 72 hours after transfection.

2.2.3.3. Cell transfection with Calcium Chloride

HEK 293T cells were seeded at 600 000 cells/well in 6-well plates 24 hours before being transfected. The media was also changed 3 hours before transfection.

Either 1, 3 or 5 µg of DNA/well was used and mixed with a 2M $CaCl_2$ solution. The mixture was added dropwise to 420 µL of 2X HBS solution (50mM HEPES, 280 mM NaCl, 1.5 mM Na_2HPO_4) and gently tapped before adding the mixture directly to the cells (210 µl/well). The plate was tilted back and forth and cells were put back into the incubator. The media was

changed either 2, 6 or 16 hours post transfection and cells were harvested 48 hours after transfection.

Protein expression was scaled up accordingly using the Calcium Chloride transfection method and 20 hours incubation before changing the media. The purification was carried out as outlined below.

2.2.4. Protein purification: Small scale

The first stages of purification exploit the affinity of the histidine tag for nickel. The mode of binding of nickel to histidine is shown in the figure below (Figure 2-4, **(1)**) and the elution with imidazole is also displayed (Figure 2-4, **(2)**)

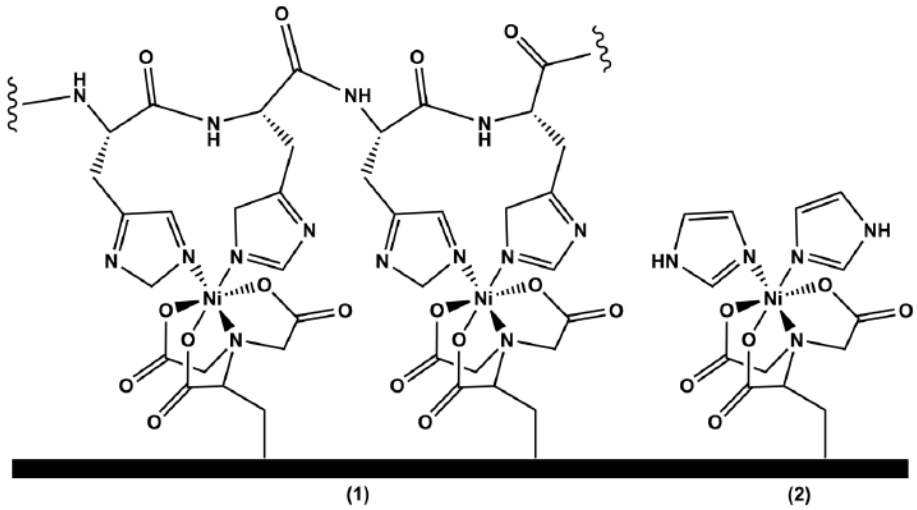


Figure 2-4: Mode of binding of histidine to nickel (1) and its elution with imidazole (2). Figure from [153].

2.2.4.1. Small scale purification from E.coli cells

All spinning steps were performed at 3000 rpm for 30 seconds.

60 µL of Ni-Nta agarose resin (Qiagen) was spun and the ethanol removed. 60 µL of milliQ water was then added before spinning the resin again and discarding the water. Finally, 60 µL of binding buffer (20 mM Na_2HPO_4, 500 mM NaCl, 20 mM Imidazole, pH 7.4) was added before applying the sample. The flowthrough was recovered by centrifugation, before

washing the column twice with 60 µL of the binding buffer. The bound fraction was recovered from the washed resin by adding 30 µL of 2X SDS loading buffer (0.05 M Tris HCl pH 6.8, 2 % Sodium Dodecyl Sulfate (SDS) (w/v), 10% glycerol (v/v), 0.1 % bromophenol blue (w/v), 1.4 M Beta-Mercaptoethanol (BME)) and was loaded on a SDS PAGE gel (see section 2.2.9.1).

2.2.4.2. Small scale purification from HEK 293T cells

Cells pellets were kept at -20 °C. When thawed, they were washed 2 times with PBS (Sigma) and collected by spinning at 1000 g for 5 minutes. Cells were then lysed with either 900 µL (10-cm dish) or 175 µL (6-well plate) of NPI-10-TIg buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, 0.05 % Tween 20 (v/v), 1 % Igepal (v/v), pH 8.0) containing DNAse I and protease inhibitors (Roche). Cells were incubated on ice for 15 minutes and on a roller shaker at 4°C for a further 15 minutes. The lysate was then cleared by spinning at 10 000 g for 10 minutes at RT. The supernatant was saved while the insoluble fraction was solubilized with the lysis buffer supplemented with 8M urea. The sample was cleared as previously described. Both fractions were then subjected to purification with Ni-Nta agarose, as described for the bacterial expression (section 2.2.4.1), with the following changes: the sample was allowed to bind to the resin overnight at 4°C on a roller shaker; the washes were performed with NPI-20T Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM Imidazole, 0.05 % Tween 20 (v/v), pH 8.0) ; the bound fraction was eluted with NPI-500-T elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM Imidazole, 0.05 % Tween 20 (v/v), pH 8.0) and the resin was also solubilized in 2X SDS loading buffer.

2.2.5. Protein purification: large scale

All buffers were prepared fresh, filtered through 0.22 µm filters (Millipore) and degassed prior to being used on the AKTA systems.

2.2.5.1. His Trap and Reverse His Trap

The strategy of purification is outlined in Figure 2-5.

Briefly, either a His Trap Fast Flow (FF) 5mL column (for purification from expression in *E.coli*) or His Trap FF 1mL column (for purification from expression in HEK 293T cells) (Generon) was washed with milliQ water before being equilibrated with 5 Column Volumes (CV) of His Trap buffer B (20 mM Na_2HPO_4, 500 mM NaCl, 500 mM Imidazole, pH 7.4) followed by 5 CV of His Trap buffer A (20 mM Na_2HPO_4, 500 mM NaCl, 20 mM Imidazole, pH 7.4). The filtered clarified supernatant was then loaded at 2.0 mL/min and eluted with a linear gradient of 100 % His Trap buffer B applied on 10 CV. Fractions of 2 mL were collected and those containing the target protein pooled and dialyzed overnight (Medicell dialysis membrane, 3500 Da molecular weight cut-off) at 4°C in 20 mM Na_2HPO_4, 500 mM NaCl, pH 7.4 in the presence of either a histidine-tagged SUMO protease (to remove the SUMO-histidine tag) or a histidine-tagged 3C protease (to remove the 3C-MBP-histidine tag). Both proteases were produced in-house.

The following day, samples were spun at 6000 rpm for 30 minutes at 4°C to remove any precipitate prior to the reverse purification. The sample was then passed back down the equilibrated His Trap column at 2.0 mL/min, and the flowthrough containing the native protein was collected. This step was repeated a second time to ensure efficient removal of contaminants.

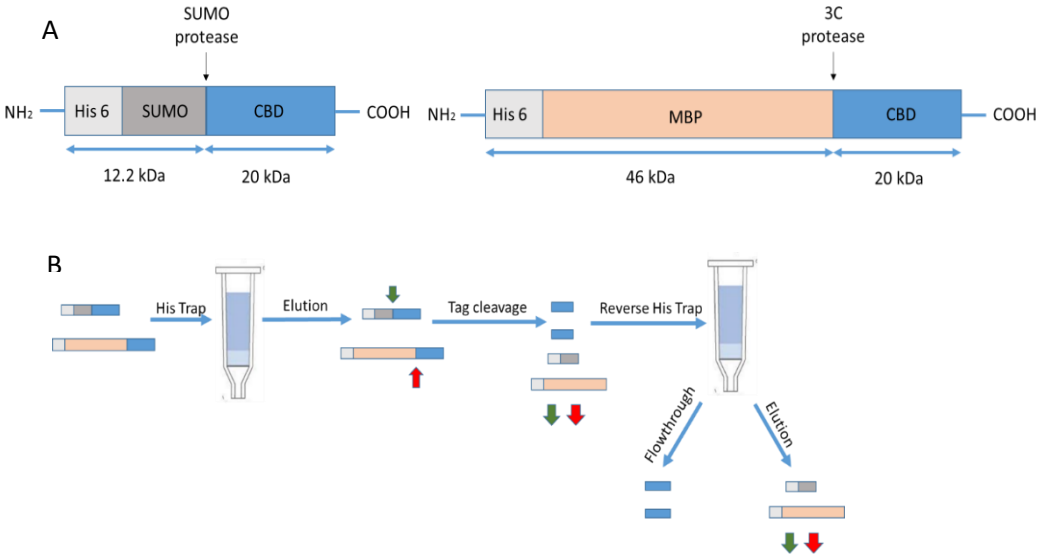


Figure 2-5: Strategy of purification of both MBP and SUMO CBD constructs. (A) The construct are designed so that the CBD sequence is preceded by the hexahistidine and either the MBP or SUMO tag. The cleavage site of either of the SUMO or 3C protease is inserted between the tag and the CBD. (B) The protein carrying the hexahistidine tag is captured on a nickel column (His Trap) and eluted with a gradient of imidazole. The cleavage of the tag is mediated by histidine tagged 3C and SUMO proteases (red and green arrows, respectively) while the concentration of imidazole is lowered down by dialysis. The mixture is then passed back through the nickel column, which retains the tags and the proteases, while the CBD is collected in the flowthrough.

2.2.5.2. Gel Filtration

Either a HiLoad 16/60 s75 (GE Healthcare) or a Superdex 75 10/300 column (GE Healthcare) were used. For both of them, the column was equilibrated with milliQ water for 0.75 CV and with buffer (either 150 mM NaCl, 50 mM Tris, pH 7.4 or 300 mM NaCl, 50 mM Tris, pH 7.4) for 1 CV. The sample, concentrated in less than 2 mL, was then loaded, and the isocratic elution allowed to take place. 2.5 mL fractions were collected and those containing the protein pooled.

2.2.5.3. *Ion Exchange Chromatography*

A Q FF 5 mL column (GE Healthcare) was equilibrated with 5 CV of Ion Exchange Chromatography (IEC) A buffer (20 mM Tris, 10 mM NaCl, pH 8.0), 5 CV of IEC B buffer (20 mM Tris, 1 M NaCl, pH 8.0) and a final 5 CV of IEC A buffer. The sample, which had been previously desalted on a HiPrep 26/10 column (GE Healthcare), was loaded at 2.0 mL/min.

A step gradient was applied, using 24 % of IEC B buffer on 5 CV to specifically elute the CBD, before applying a gradient to 100 % of IEC B buffer on 10 CV. A 12% gradient of IEC B buffer was used for module 1 while modules 2 and 3, as well as 111, 222 and 333 were eluted with linear gradients. 2 mL fractions were collected for all proteins, and pooled after running a SDS PAGE gel (see section 2.2.9.1).

2.2.6. Protein concentration

Once the correct fractions were pooled, the protein concentration was measured with a Nanodrop by reading the absorbance at 280 nm, using the Beer-Lambert formula:

$$A = C * \varepsilon * l$$

Where A is the absorbance, C the concentration (mol/L), ε the molar absorption coefficient ($L.mol^{-1}cm^{-1}$) and l the path length (cm). When needed, the molar concentration was converted into a concentration in mg/mL using the molecular weight.

The values for the molar absorption coefficient and the molecular weight were obtained for each protein using the prot param online tool from Expasy. These values are reported in Table 2-3 below.

Table 2-3: Absorption extinction coefficient and molecular weight of the expressed proteins

	ε ($L.mol^{-1}cm^{-1}$)	Molecular weight (Da)
CBD	40 630	19 640
Module 1	13 200	6 902
Module 2	11 710	6 665
Module 3	14 230	6 265
111	41 090	20 714
222	38 110	20 650
333	45 670	19 750

If needed, the protein concentration was increased by centrifugation using centricons (Millipore, molecular weight cut off of 3500 Daltons). Briefly, the centricons were equilibrated in the corresponding buffer, spun for 10 min at 4500 g, before discarding the buffer. A maximum volume of 15 mL of protein sample was then added and spun for 20 min at 4500g, 4°C. The process was repeated until reaching the desired concentration. The protein sample was pipetted up and down between each spin to prevent protein aggregation.

2.2.7. Protein storage

Protein samples were stored in aliquots in the IEC buffer, snap frozen in liquid nitrogen and stored at -80°C until use.

2.2.8. Buffer exchange with PD10

A PD 10 Sephadex G-25 M column (GE Healthcare) column was used for buffer exchange procedures. The column was first rinsed with 30 mL of milliQ water before being equilibrated with 25 mL of final buffer.

2.5 mL of protein in the initial buffer were applied to the column and allowed to run through.

3.5 mL of final buffer were added to elute the bound protein in the final desired buffer.

2.2.9. Protein Characterisation

2.2.9.1. Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis

The levels of expression and purity of the expressed protein were analysed by Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis (SDS-PAGE) using 15 % polyacrylamide gels in a Bio-Rad gel electrophoresis system. A volume of 5 µl of protein samples were added to 5 µl of 2X SDS loading buffer (0.05 M Tris HCl pH 6.8, 2 % SDS (w/v), 10 % glycerol (v/v), 0.1 % bromophenol blue (w/v), 1.4 M BME), boiled at 100°C for 5 minutes and loaded onto the gel. All gels were run at 180 V for 60 minutes, stained with Magic Stain (60 mg Coomassie G-250, 3.4 mL HCL concentrated, 1L solution) and de-stained with milliQ water.

2.2.9.2. 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB) assay

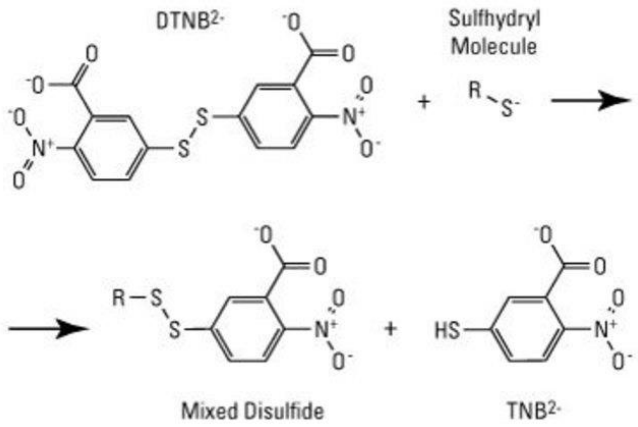


Figure 2-6: Principle of the DTNB assay. The DTNB reagent reacts with molecules containing sulfhydryl to produce TNB2-, a coloured product, whose absorbance is measured at 412 nm. Figure from the manufacturer's instructions manual (Thermo Scientific).

The principle of the assay is outlined in Figure 2-6.

Briefly, 4 mg of Ellman's reagent (Thermo Scientific) were dissolved in 1 mL Reaction Buffer (0.1 M sodium phosphate, pH 8.0). A set of standard concentrations was prepared using Dithiothreitol (DTT), ranging from 0.001 to 0.1 mM. Each standard or sample reactions were mixed as follows: 250 µL sample, 2.5 mL reaction Buffer, 50 µL Ellman's reagent. The reaction

was allowed to develop for 15 minutes at Room Temperature (RT) before reading the absorbance at 412 nm. The number of free cysteines in the sample was determined using the standard curve obtained from the DTT standards.

2.2.9.3. Western Blot

15 % SDS-PAGE gels were run as described previously before transferring onto a nitrocellulose membrane in transfer buffer (25 mM Tris, 190 mM Glycine, 20 % methanol, pH 8.3) at 100 V for 1 hour. The membrane was washed with PBS (Sigma), blocked for 40 minutes to 1 hour in 5 % milk (Marvel)/PBS-T (8 mM Na_2HPO_4, 150 mM NaCl, 2 mM KH_2PO_4, 3 mM KCl, 0.05 % Tween (v/v), pH 7.4). The primary mouse anti-histidine antibody was used at a 1:1000 dilution in 5 % milk/PBS-T and incubated for either 1 hour at RT or overnight at 4°C. The membrane was then washed 3*10 minutes with PBS-T before adding the anti-mouse secondary antibody, diluted 1:1000 in 5 % milk/PBS-T. The membrane was washed again 3 times for 10 minutes with PBS-T and revealed with the Enhanced Chemiluminescence (ECL) substrate (Pierce).

2.2.9.4. Mass Spectrometry analysis

The two most common uses of Mass Spectrometry (MS) are the determination of the identity of a protein after its digestion, and that of its intact mass. Both rely on the ionization of either the protein or of the peptides obtained from its digestion. The mass to charge ratio of the sample is then measured. When digested, the sequence of the peptides can be determined, and matched to the protein sequence to confirm that the host has expressed the protein of interest. When using accurate instruments, the determination of the mass of the intact protein can not only confirm its identity, but also its post-translational modifications [154,155].
A MALDI TOF (Matrix Assisted Laser Desorption Ionisation – Time Of Flight) instrument was used to confirm the identity of the digested protein, which was extracted from a SDS PAGE gel band.

The intact mass analyses were run on an Electrospray instrument. Samples were first reverse desalted using C4 trap prior to mass analysis. 1 to 3 pmol of proteins were applied to the column.

Both MS analyses were carried out by internal collaborators (Centre for Proteome Research, University of Liverpool).

2.2.9.5. Circular dichroism

Another important aspect of protein characterization is to check for the correct protein structure. Circular Dichroism (CD) enables to check the secondary structure of proteins by relying on the differential absorption of right and left circularly polarised light by their chromophores. When the circular dichroism is measured against the wavelength of the light, a CD spectrum is obtained and can be matched to specific secondary structural elements of a protein, namely its alpha helix and beta sheet content [156,157].

Protein solutions (0.1 to 0.15 mg/mL) in 2.5 mM to 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer, pH 6.5 were applied to a 0.5 cm path length cuvette. The signal was measured between 260 and 190 nm using a JASCO CD instrument. Ellipticity data were obtained in millidegrees and converted into molar ellipticity to correct for protein concentration, using the formula:

$$\theta = m * M/(10 * l * C)$$

Where θ: Molar Ellipticity (deg.cm^2/dmol), m: ellipticity (degrees), M: average molecular weight (g/mol), l: path length (cm), C: concentration in protein in g/L.

2.2.10. Binding assays

2.2.10.1. Gelatin sepharose binding assay

The binding to gelatin was first qualitatively assessed using a gelatin sepharose 4B resin (GE Healthcare). All spinning steps were performed at 3000 rpm for 30 seconds.

A volume of 100 µL of resin was spun down and the ethanol removed. 100 µL of milliQ water was then added before spinning the resin again and discarding the water. Finally, 100 µL of binding buffer was added before applying 50 µg of the protein sample. The flowthrough was recovered by centrifugation, before washing the column twice with 100 µL of the binding buffer. The bound fraction was recovered from the washed resin by adding 30 µL 2X SDS loading buffer and was loaded on a SDS PAGE gel.

2.2.10.2. Plate binding assay

2.2.10.2.1. Protein biotinylation

Protein samples were dialysed overnight in 1L of 0.1M $NaHCO_3$, pH 7.4, at 4°C with stirring. Biotin (EZ-Link™ Sulfo-NHS-LC-Biotin, Thermo Fisher) was then added with a 20-fold molar excess, according to the manufacturer's instructions. The samples were incubated for 20 minutes at 22°C and then for 2 hours at 4°C. Samples were then dialysed against 1L of PBS (145 mM NaCl, 11.3 mM Na_2HPO_4, 3.97 mM KH_2PO_4, 7.69 mM NaN_3), pH 7.4 overnight at 4 °C, with stirring. Both dialysis were performed with a cellulose membrane with a molecular weight cut-off of 3500 Daltons.

The biotinylation efficiency was tested with a HABA/Avidin assay according to the manufacturer's instructions (Pierce™ Biotin Quantitation Kit, ThermoFisher) and the number of biotin added per protein was quantified.

2.2.10.2.2. Plate preparation

Either type I (Collagen Solutions) or type II collagen (Bioiberica) solution or skin gelatin (Sigma) was heat-denatured at 80°C for 20 minutes, vortexing every 5 minutes. The collagen was then diluted to 0.01 mg/mL in 0.1 M carbonate buffer, pH 9.2. 50 µL were added to each well of a 96 well plate (ThermoScientific, Immulon 2HB, high binding), except the outer ones, giving a coating at 0.5 µg/well.

The plates were wrapped in cling film and left at 4°C for 3 days and then washed 3 times with PBS-Tween, blocked with 100 µL of PBS-2.5% BSA (Sigma) for 30 minutes at RT, washed one more time with PBS-Tween and dried in an oven at 50 °C for 15 minutes. Plates were kept at 4°C and used within 4 weeks.

2.2.10.2.3. Binding assay

Biotinylated proteins were added over a concentration range on plates coated with 0.5 µg of gelatin/well and incubated for 1 hour at 22°C. Plates were washed 3 times with 200 µL PBS-2.5% BSA before being blocked with BSA for 30 minutes at RT. Plates were washed again 3 times with 200 µL PBS-2.5% BSA before alkaline phosphatase conjugated to streptavidin (diluted 1:10 000 in PBS, Pierce) was added for 30 minutes at 22°C. Plates were washed 3 times with 200 µL PBS-2.5% BSA and 100 µL of P-Nitrophenyl Phosphate (PNPP) solution was added. It was incubated for 8 minutes at RT before reading the plate absorbance at 405 nm.

The signal from the blank wells was removed from the sample wells and a binding curve was generated using a non-linear regression fit (four parameters) in GraphPad Prism.

2.2.11. Nuclear Magnetic Resonance experiments

2.2.11.1. 1D Nuclear Magnetic Resonance

Protein samples were concentrated to a maximum volume of 400 µL. D_20 was added (10 % (v/v)) and 8 to 32 spectra were acquired on an 800 MHz Bruker, at 298 K.

2.2.11.2. 2D Nuclear Magnetic Resonance

Proteins were expressed in ^{15}N ammonium chloride minimal media. They were expressed and purified as previously described (see sections 2.2.2.3 and 2.2.5). Proteins were then buffer exchanged in 25 mM sodium phosphate at pH 6.5 using a PD10. NMR data were acquired at 298K on an 800 MHz Bruker. The data were processed using TopSpin and CCPN. The CBD chemical shift assignments as well as those of the individual modules were identified using previously published results [140,158,159,160].

2.2.12. Protein conjugation

When appropriate, a Fluorescein-5-isothiocyanate (FITC) tag (Sigma) was added to the 222 protein according to the manufacturer's instructions, giving FITC-222. Protein samples were then buffer exchanged against either 20 mM HEPES, pH 6.5 or 20 mM sodium phosphate, pH 6.5 using a PD 10 Sephadex G-25 M column (GE Healthcare). The protein was then concentrated up to 0.7 mg/mL before proceeding to the cationisation.

2.2.12.1. Cationisation

The ratio of (N,N`-dimethyl-1,3-propanediamine) (DMPA, Sigma) and N-(3-Dimethylaminopropyl)-N`-ethylcarbodiimide hydrochloride (EDC, Sigma) to the native protein anionic residues were kept consistent at 150 and 34, respectively, to calculate the required amounts of reagents to add.

Under a fume hood, the required amount of DMPA was added to either 20 mM HEPES buffer, pH 6.5 or 20 mM sodium phosphate, pH 6.5. The pH of the solution was adjusted to 6.5

before adding it to the protein sample. The required amount of EDC was then weighed and added. The sample was left at 4°C for 24 hours with stirring.

The sample was then purified from the excess reagents using 50 mL concentrators - Amicon Ultra 15 - Regenerated cellulose 10 NMWL (Merck) with a 10 kDa molecular weight cut-off. The protein was spun five times for 15 minutes at 4 900 g. The volume of the protein sample was brought up to 20 mL by adding either 20 mM HEPES buffer, pH 6.5, or 20 mM sodium phosphate, pH 6.5 between each spin.

2.2.12.2. Surfactant addition

A ratio of 1.4 surfactant chains per cationic residue was kept constant. The amount of surfactant needed was calculated based on a cationisation efficiency of 100 % and added to the protein sample, which was left overnight at 4°C with stirring. The resulting protein will be referred to as Surfactant-222 or, when appropriate, Surfactant-FITC-222.

2.2.12.3. Dynamic Light Scattering

Particle size and zeta potential analyses were performed using a ZetaSizer Nano ZS (Malvern Instruments, UK) and ZetaSizer software (Malvern Instruments, UK). Samples were filtered through a 0.22 µm filter prior to the analysis. Zeta potential measurements were made after a 120 second equilibration with 10-100 runs per sample. Size measurements were made after a 120 second equilibration, with the backscatter measured at 173°. The values reported in this thesis correspond to the radius expressed in number of the total sample. Each zeta potential and particle size measurements were performed in triplicates.

2.2.13. Cell coating

2.2.13.1. Cell culture and conditions

All cells were cultured in monolayer using either T25, T75 or T175 flasks (Corning) and, respectively, 5, 15 or 20 mL of expansion medium (low glucose (1000 mg/L) Dulbecco's Modified Eagles Medium (DMEM) (Sigma) with 100 units/ml penicillin, 100 µg/mL streptomycin (Sigma) (later referred to as P/S), 2mM GlutaMAX supplement (Gibco), 10% (v/v) Foetal Bovine Serum (FBS, Gibco) and 5 ng/ml freshly supplemented basic human fibroblast growth factor (Peprotech). The media was replenished every 2 to 3 days. Cell confluence was monitored using bright field microscopy.

2.2.13.2. Cell counting and viability

Cells were harvested by aspiring the expansion medium, washing with PBS (Sigma) and adding a trypsin/EDTA solution (Sigma). Cells were incubated for 3-5 minutes at 37°C before the flask was tapped. Cell detachment was monitored under a bright field microscope. Expansion medium was added to stop the action of the trypsin, and cells were then spun for 5 min at 1500 rpm. The supernatant was removed and cells were resuspended in a known volume of expansion medium. A 10 µL aliquot was taken and mixed in a 1:1 ratio with trypan blue (Sigma) before counting the cells with a haemocytometer.

Both white and blue cells were counted and the viability was obtained as follows:

$$Viability\ (\%) = \frac{Number\ of\ white\ cells}{Number\ of\ total\ cells} * 100$$

2.2.13.3. Coating adherent MSCs with Surfactant-FITC-222

The outline of the experiment is illustrated in Figure 2-7.

50 000 cells were seeded onto 35-mm dishes (Corning) 4 hours prior to the coating. The expansion medium was aspirated before adding 0.5 mL of 18 µM of Surfactant-FITC-222 in 20 mM HEPES, pH 6.5. Cells were incubated with the protein conjugate for 30 minutes in the

incubator. The protein solution was then removed and cells were washed with 0.5 mL of 0.04 mg/mL heparin ammonium salt (Sigma), resuspended in PBS (Sigma).

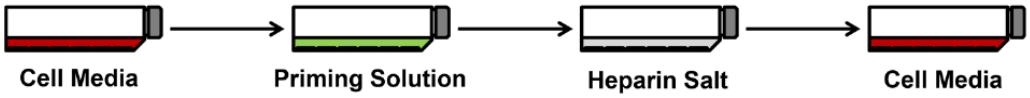


Figure 2-7: Coating adherent MSCs. The media is removed before adding the Surfactant-FITC-222 protein solution at 18 µM. Cells are incubated with the protein for 30 minutes at 37°C before being washed with heparin salt. Figure from [153].

Imaging medium (phenol-free, low glucose (1000 mg/L) DMEM (D5921) with 100 units/mL penicillin, 100 µg/mL streptomycin, 2 mM GlutaMAX supplement and 10 % (v/v) FBS, buffered with 20 mM HEPES (Sigma) was then added and cells were observed by confocal microscopy.

2.2.13.4. Coating MSCs in suspension with Surfactant-FITC-222

The outline of the experiment is illustrated in Figure 2-8.

Cells cultivated in proliferation medium were trypsinised, counted, and resuspended at 1 million cells/mL. The cells were then spun down for 5 min at 1500 rpm and resuspended in 300 µL of 5 µM of either Surfactant-222 or Surfactant-FITC-222. The protein samples were either in 20 mM HEPES buffer, pH 6.5 or 20 mM sodium phosphate, pH 6.5. Control samples were also incubated with 300 µL of either 5 µM of native 222, or buffer. Cells were incubated for 30 minutes at 37°C with shaking and manually agitated every 10 minutes to prevent cell aggregation.

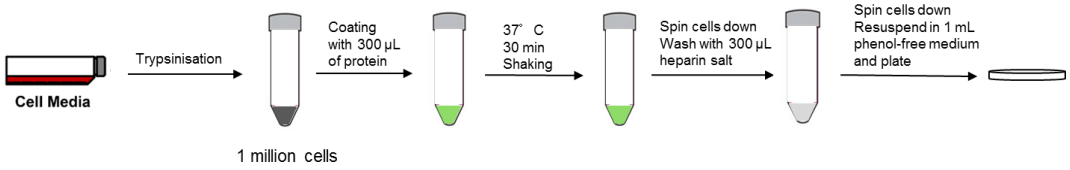


Figure 2-8: Coating MSCs in suspension. Cells were trypsinised and counted. Each million of cells was incubated with 300 µL of protein for 30 min at 37°C. Cells were then spun down and washed with heparin salt before being spun down and plated.

Following the incubation, cells were spun down for 5 minutes at 1500 rpm and washed once with 300 µL of 0.04 mg/mL heparin ammonium salt (Sigma), resuspended in PBS (Sigma), before being spun down again and resuspended in 1 mL of imaging medium. 250 µL were plated in glass-bottomed dishes and observed by confocal microscopy.

2.2.13.5. Cell adhesion assay

96-well Tissue culture plates (Costar) were coated with type II gelatin as described for the binding assay, under sterile conditions. Cells were coated in suspension with either Surfactant-FITC-222 or Surfactant-222 as described in the previous section. Coated cells were resuspended in low glucose (1000 mg/L) DMEM (Sigma) with 100 units/mL penicillin, 100 µg/mL streptomycin (Sigma), 2mM GlutaMAX supplement (Gibco) at a density of 4 000 000 cells/mL. 200 µL of this cell suspension were added in the first 2 wells of the 96-well plate, before performing 1:2 serial dilutions in the same medium. Dilutions were performed on alternating columns only. This process was repeated a second time on the empty columns, giving 4 replicates for each cell density (200 000, 100 000, 50 000 and 25 000 cells). Control cells were treated with native 222. Cells were also seeded on an untreated tissue culture plate for comparison purposes. The layout of the plate is outlined in Figure 2-9.

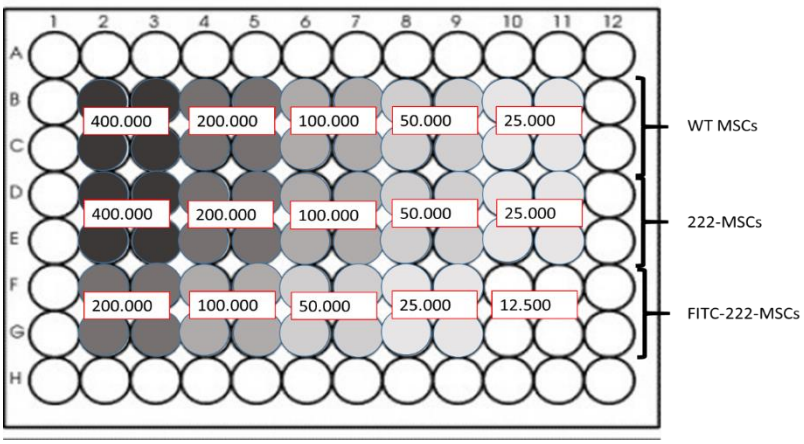


Figure 2-9: Layout of the plate used for the cell adhesion assay. The numbers indicate cell densities. Cells were either uncoated (Wild Type (WT) MSCs, coated with Surfactant-222 (222-MSCs) or with Surfactant-FITC-222 (FITC-222-MSCs)

Plates were incubated overnight in the incubator. The media was then removed, and each well was washed with 200 µL sterile PBS (Sigma). Imaging medium containing FM-4-64 (InVitrogen) was then added to each well, and plates were imaged using an epifluorescent microscope.

2.2.13.6. Microscopy

All confocal and epifluorescence microscopy experiments were performed with internal collaborators (Centre for Cell Imaging, University of Liverpool). The LSM 710 microscope was used for confocal microscopy purposes. When required, cells were imaged without and with the membrane dye FM-4-64 (InVitrogen).

3. New expression system identified for the expression of a soluble, folded and functional Collagen Binding Domain of MMP-2

3.1. Introduction: Characterisation of the interaction between the Collagen Binding Domain of Matrix Metalloprotease-2 and gelatin

The type of interactions which mediate the binding of the CBD to gelatin was studied by Steffensen et al [141], who showed that dimethylsulfoxide (DMSO), but not sodium chloride, could displace either the bound CBD or the full length MMP-2 from gelatin sepharose. This suggests that hydrophobic interactions are prevalent in the binding of the protein to the substrate, an observation which was confirmed by others, on both MMP-2 [137], MMP-9 [161] and Fibronectin [145].

Another important feature of the binding of the CBD to gelatin is the structural integrity of the protein. This relies mostly on the presence of six disulfide bonds; each individual module contains 4 cysteines, and each one is involved in the formation of an intramodule disulfide bridge [141,149] . While preliminary data [141] suggested that the binding to gelatin was maintained after reduction of the disulfide bridges, others have shown the opposite [143]. It is important to note that the study by Steffensen et al [141] has, however, shown that the reduction and carboxymethylation of the disulfide bridges abrogated the binding to gelatin. One of the main aims of the project was, therefore, to express a correctly folded and functional CBD in a suitable host for recombinant expression.

3.2. Recombinant protein expression: generalities

Recombinant protein expression relies on using a genetically-modified, easy to grow host for the expression of a heterologous protein. The aim of recombinant protein expression is to obtain high yields of pure and functionally folded protein [150].

Various hosts are available for this purpose, such as bacteria, yeasts, insects or mammalian cells [162]. Each of them offers advantages and disadvantages, and the choice of the ideal host very much depends on the characteristics of the protein to express, such as its size, origin, and post-translational modifications [162]. The cost of culturing the hosts and the yield of protein that can be obtained are other factors which influence the choice of the organism used for protein expression [162].

3.2.1. Recombinant protein expression in *E.coli*

E.coli has been the workhorse of recombinant protein expression. This gram-negative, mesophilic prokaryote has a doubling population time of 20 minutes [150]. It is easy to grow, inexpensive and easy to scale-up [150,163]. Various vectors have been developed to enable the transformation of *E.coli* strains with the gene of interest, with various promoters and plasmid copy number [150]. One of the most commonly used is the pET vector system, which enables the induction of protein expression upon addition of Isopropyl-ß-D-thiogalactopyranoside (IPTG) [150].

3.2.1.1. Protein insolubility and inclusion bodies formation

It is not unusual to encounter solubility issues when expressing recombinant proteins. High rates of protein synthesis, coupled to high levels of expression, can saturate the folding machinery of the host, leading to the accumulation of misfolded protein intermediates [164]. These intermediates present exposed hydrophobic patches which facilitate their self-

aggregation into the so-called inclusion bodies [165], which mostly contain heterogeneous populations of misfolded proteins [165,166].

This problem is further exacerbated when expressing eukaryotic proteins in *E.coli*, as this host lacks the folding machinery of higher organisms. In particular, the cytoplasm of *E.coli* is maintained in a reduced environment by both the glutaredoxin and thioredoxin pathways, which prevents the formation of disulfide bridges [166]. Bacterial proteins that need disulfide bridges are naturally secreted to the periplasm, where the enzymes of the Dsb family, especially the oxidase DsbA and the isomerase DsbC, perform this post-translational modification [167].

3.2.1.2. Recobinant expression of the CBD of MMP-2

The CBD of MMP-2 contains 12 cysteines, each being involved in an intramolecular disulfide bond. Previous publications on the production of the CBD of MMP-2 in *E.coli* have described the accumulation of the CBD in inclusion bodies. Authors have therefore developed a strategy of denaturation/refolding of the CBD from these inclusion bodies [141,144,145,148]. This approach was also adopted for the expression of single modules [168] and the CBD of MMP-9 [161]. The advantage of this strategy is that it prevents the heterologous protein from being degraded, usually leading to high levels of pure protein. However, the procedure can be tedious, time consuming, and does not guarantee that the protein is folded in a native-like structure (Chapter 12 of reference 176). In order to circumvent this limitation, the protocol described for the expression of the CBD added a purification step onto a gelatin sepharose column, in order to select only the functional protein [141]. However, eluting the protein from the column involved disrupting the binding with DMSO, which should be avoided in order to maintain the structural integrity of the protein, since DMSO as a propensity to denature proteins even at modest concentrations (>5 %). Finally, this method involved the use of a Histidine-tagged CBD instead of a protein with a native N-terminus.

Overall, it was decided that a more preferable strategy would be to ensure that the native CBD is expressed as a correctly folded and soluble form. Several strategies were adopted in that attempt.

3.2.1.2.1. Codon optimisation

As the abundance of tRNA varies from one organism to another, the sequence of the gene of interest needs to be codon-optimised, to favour the use of tRNAs which are present in large concentration in the host [169]. This strategy was shown to favour, in some cases, soluble expression of recombinant proteins in *E.coli* [169].

3.2.1.2.2. Addition of a tag

In addition to facilitating protein purification [150], the use of a fusion partner can help improve protein solubility by preventing the fusion protein from aggregating [150,165]. The Maltose Binding Protein (MBP) (40 kDa [163]) and Small Ubiquitin-related Modifier (SUMO) (11.5 kDa) tags [170, 171] are two commonly used fusion partners. They have been shown to improve the solubility of proteins normally accumulated in inclusion bodies [172,171]. MBP was also shown to act as a chaperone by improving the folding of proteins it was fused to [172], while the SUMO tag offers advantages regarding its removal by proteolytic cleavage: the SUMO protease recognises the folded 3D structure of SUMO, preventing it from cleaving an identical sequence in the protein of interest [171]. The cleavage also generates a native N-terminal protein [171,173].

3.2.1.2.3. Use of engineered strains

Several *E.coli* strains have been engineered to improve the levels of expression or solubility of recombinant proteins. BL21 star strains are mutated for the Lon and OmpT proteases, which protects the recombinant protein from being degraded after cell lysis. They also carry a hsdSB mutation which prevents plasmid loss and

degradation [150] as well as a mutation in the rne131 allele, which attenuates the activity of the RNAse E, leading to increased mRNA stability and higher expression levels [174]. Rosetta cells have the additional advantage of expressing rare tRNA to overcome codon bias (Chapter 2 of reference 176).

The thioredoxin and glutaredoxin pathways are responsible for maintaining the cytoplasm in a reduced environment in *E.coli*. Mutations in both of these pathways cause the two thioredoxins in *E.coli* to be maintained in their oxidised state and to catalyse the oxidation of disulfide bridges, in a reversal of their physiological function [167,174]. Strains mutated in their thioredoxin and glutareodxin pathways were commercialised under the name Origami [167,174]. In order to facilitate the correct oxidation of cysteines in the cytoplasm, the isomerase DsbC was also overexpressed in the cellular compartment of another commercial strain with the Origami background, which was named the SHUffle cells (trxB, gor, ahpC* strain) [167,174]. Interestingly, authors have previously successfully expressed a soluble and functional module of the CBD using *E.coli* strains mutated for thioreductases [175].

3.2.1.2.4. Optimisation of growth and expression conditions

Optimal growth conditions are required to favour high levels of protein expression. The induction of protein expression is usually performed in the mid-log phase, when protein synthesis is optimal [150]. The concentration of IPTG can also be optimised. Finally, expression of the protein at low temperature (15-25°C) can favour proper protein folding by slowing down the rate of expression [150]. This slower expression rate reduces the concentration of folding intermediates [165] and prevents hydrophobic interactions which are responsible for inclusion bodies formation [163].

3.2.2. Recombinant protein expression in mammalian cells

Mammalian cells are often considered less attractive than *E.coli* for recombinant protein expression, as cultivating them is more expensive [162], their transfection requires large amounts of DNA, and the yield obtained after expression is generally low (Chapter 14 of reference 151).

However, they are a particularly useful host for expressing proteins with post-translational modifications, as they have the required machinery to add these modifications to yield a fully folded and functional protein [176,162].

Consequently, several methods for Transient Genetic Expression (TGE) have been developed, which rely on either chemical, physical or biological means of delivering the expression plasmid to the nucleus of mammalian cells [177], such as Lipofection, Calcium Chloride and Polyethyleneimine-mediated methods.

3.2.3. Strategy of expression of the CBD

Considering all these aspects, we first designed a strategy for expressing the CBD in *E.coli*, and aimed to maximise its solubility using some of the strategies outlined earlier. BL21 Star and Rosetta cells were tested in the first instance, in order to determine if these strains could express the CBD with its 6 disulfide bridges. This chapter describes the results obtained with these initial strains, as well as in the mammalian cell line HEK 293T, which was used in an attempt to favour the folding of the human CBD, using all three transfection methods mentioned earlier.

Finally, this chapter describes the successful use of the engineered *E.coli* strain Shuffle for the expression of a fully folded CBD, as well as its purification to 90 % purity and its characterisation.

3.3. Materials and Methods

3.3.1. Cloning

The CBD of MMP-2 had the following sequence:

R V K Y G N A D G E Y C K F P F L F N G K E Y N S C T D T G R S D G F L W C S T T Y N F E K D G K
Y G F C P H E A L F T M G G N A E G Q P C K F P F R F Q G T S Y D S C T T E G R T D G Y R W C G
T T E D Y D R D K K Y G F C P E T A M S T V G G N S E G A P C V F P F T F L G N K Y E S C T S A G
R S D G K M W C A T T A N Y D D D R K W G F C P D Q G Y S

The DNA encoding for this protein sequence was codon optimised and cloned into either the pOPINS or pOPINM (OPPF-UK) vector using the ligase independent InFusion cloning methodology. The two constructs which were generated carried either a His_6-SUMO or a His_6-MBP tag, followed by a cleavage site for a SUMO and 3C protease, respectively (Figure 3-1).

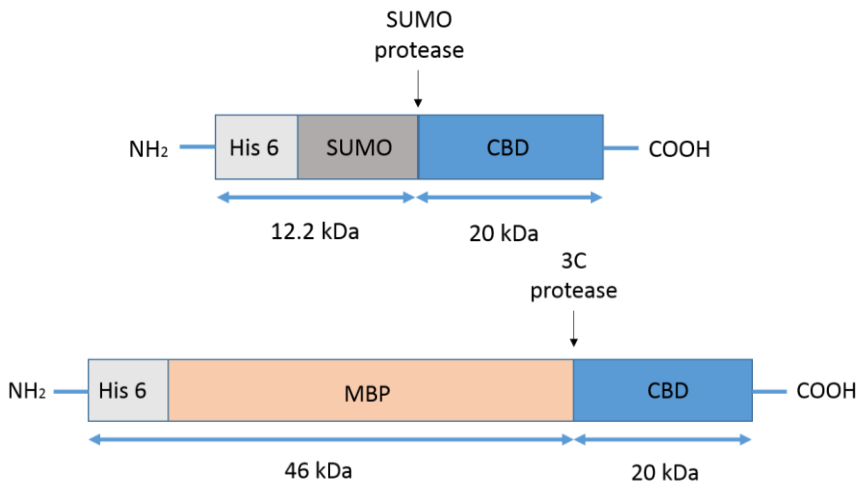


Figure 3-1: Schematic representation of the two constructs used for expression of the CBD. Two constructs were generated, one carrying a His_6-SUMO tag, and one a His_6-MBP tag, each at the N-terminus of the protein. The tags were separated from the sequence of the CBD by the cleavage site of the SUMO and 3C proteases, respectively.

3.3.2. Expression in *E.coli* strains

E.coli cells were transformed with the sequenced expression plasmids. Conditions of expression such as the temperature and length of the production phase as well as the concentration of inducer were optimised on a small scale. On a larger scale, protein expression was carried out in 2L flasks. Cells were grown either in LB or minimal media at 37°C, induced for protein expression with IPTG, harvested following the production phase and stored at -80°C.

3.3.3. Expression in mammalian cells

HEK 293T cells were cultivated as mentioned in the Materials and Methods section. They were transfected with the recombinant pOPINM construct using either Lipofectamine, Calcium chloride or Polyethylenimine. Conditions of transfection such as the amount of DNA, the length of incubation of cells with the DNA:transfection reagent complex or the time of harvesting were optimised.

3.3.4. Purification strategy

Following cell lysis, the released recombinant protein was captured on a nickel column, exploiting the affinity of the hexahistidine tag for this resin. The protein was eluted with a gradient of imidazole, and the tag cleaved with the corresponding histidine-tagged protease. A reverse nickel purification enabled the isolation of the cleaved protein from the tag and protease. A third step of purification was needed; for this both Ion Exchange and Gel Filtration Chromatography were tested. The former was optimised to enable the recovery of a native CBD with more than 90 % purity.

3.3.5. Characterisation

The expressed protein was characterised for its identity using Mass Spectrometry by MALDI-TOF, while the correct folding was investigated with an intact mass analysis with Electrospray Ionisation. The secondary structure was analysed by circular dichroism, while the folding into a tertiary structure was assessed by 2D NMR. The number of disulfide bridges in the protein was quantified with a DTNB assay. Finally, the functionality of the protein was tested on a gelatin sepharose resin, and its affinity for type II gelatin quantified using a plate binding assay.

3.4. Results

3.4.1. Expression of the CBD in *E.coli* BL21 and Rosetta cells

3.4.1.1. Small scale screening for optimal expression in Lysogeny Broth (LB)

E.coli Rosetta pLysS (DE3) and *E.coli* BL21 Star were first transformed with either the pOPINM (carrying the gene encoding for His_6-MBP-CBD) or the pOPINS vector (carrying the gene for His_6-SUMO-CBD), respectively. Cells were cultured at 37°C before induction, induced with a range of concentrations in IPTG (200 – 400 µM), and then either left for a further 5 hours at 37°C or overnight at 18°C.

Figure 3-2 shows the results of expression of His_6-MBP-CBD in Rosetta pLysS cells at 18°C. Gels show that the CBD was expressed as a soluble protein with every condition tested, with a higher proportion of insoluble protein with increasing concentrations of IPTG. The same pattern was observed in these cells at 37°C, with an overall reduced level of expression (not shown), as well as in the BL21 Star cells expressing the His_6-SUMO-CBD (not shown).

It was, therefore, decided from this screening step to carry out the production on a larger scale with a concentration of 200 µM IPTG at 18°C overnight for both constructs.

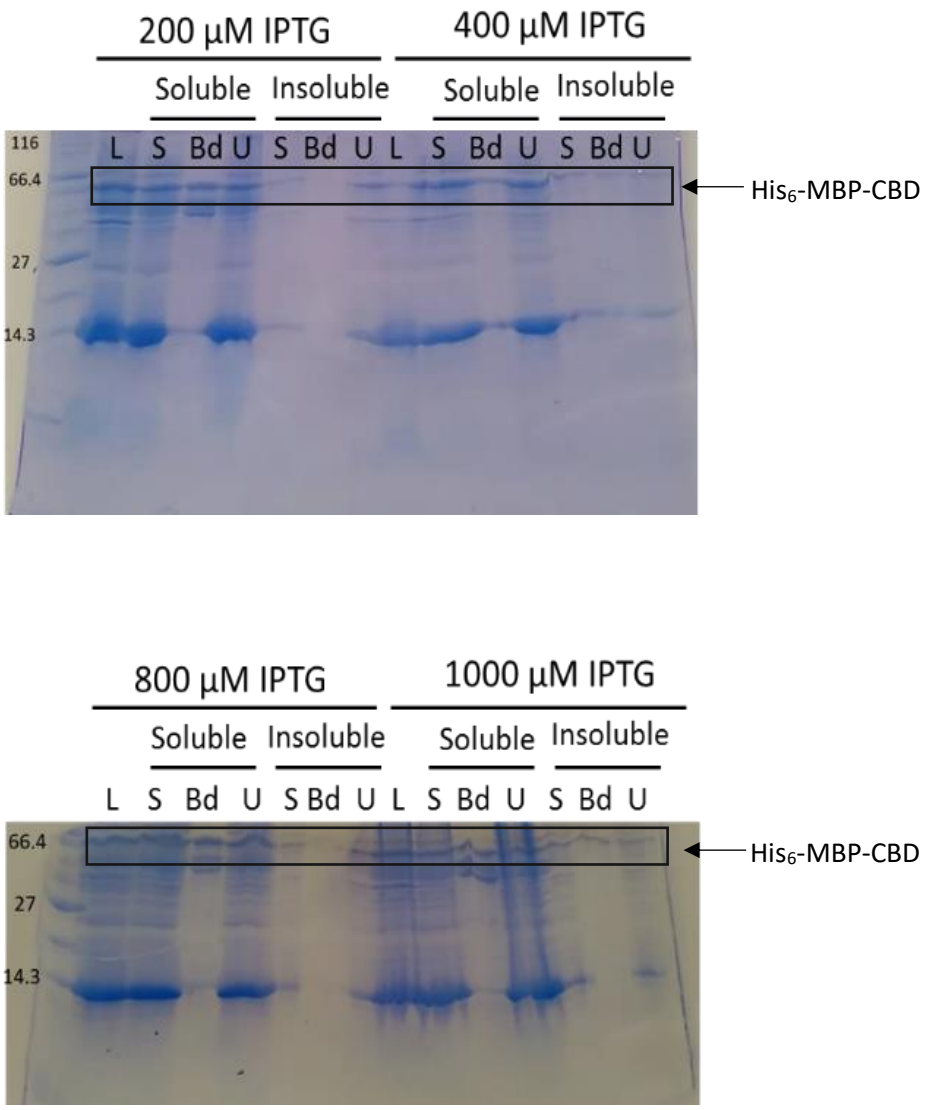


Figure 3-2: SDS PAGE showing the screening for expression of the His_6-MBP-CBD in LB in Rosetta pLysS (DE3) cells. Rosetta pLysS *E.coli* cells were transformed with a plasmid encoding for the CBD of MMP-2, carrying a His_6-MBP tag. Cells were grown in 10 mL LB until reaching an OD (600 nm) of 0.6. Cells were induced with various concentrations of IPTG overnight at 18°C. Cells were then lysed and both the soluble and insoluble fractions were subjected to a nickel purification. L: Lysate; S: Supernatant; Bd: Nickel bound; U: Unbound.

3.4.1.2. Large scale expression with the optimised conditions in LB

The protein expression was then carried out in 2L Flasks, as described in the Materials and Methods section. Results show a successful expression of both the His_6-MBP-CBD and His_6-SUMO-CBD in the soluble fraction (Figure 3-3). The cleavage of the tags was also performed successfully, although it seemed incomplete for the MBP tag (Gel A, Figure 3-3). The reverse purification enabled the recovery of the protein in the flowthrough.

The protein was further concentrated and the untagged CBD could be detected on a gel with both the CBD_M construct (initially carrying the MBP tag) and the CBD_S construct (initially carrying the SUMO tag) (Gel C, Figure 3-3).

The samples were further concentrated to 1mg/mL before being analysed by 1D NMR. The spectra showed that the proteins were folded like a molten globule (not shown). In order to further clarify the folding state of the protein, it was decided to express it in labelled minimal media to assess the foldedness of the protein by 2D NMR.

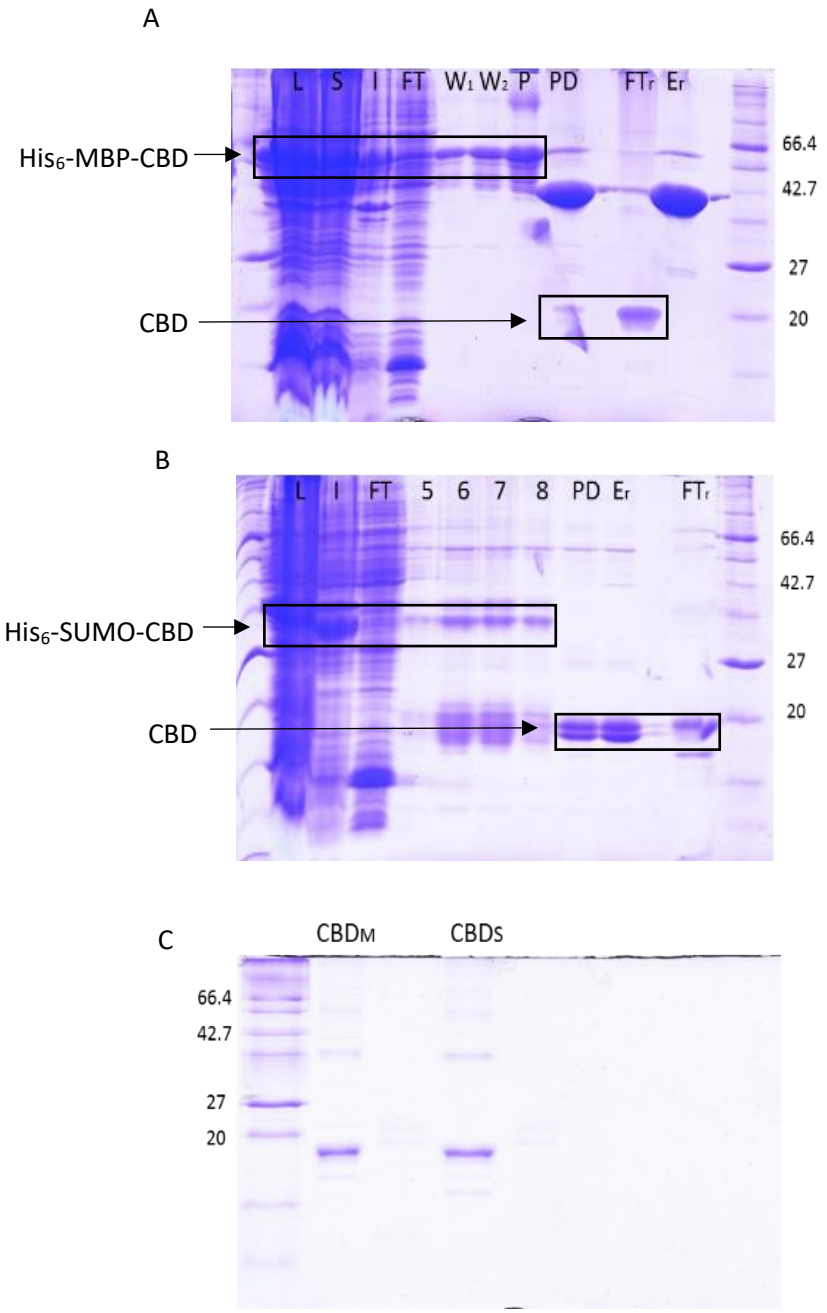


Figure 3-3: SDS PAGE showing the expression of the His_6-MBP-CBD (A) and His_6-SUMO-CBD (B) in LB and the level of purity of the CBD (C). Rosetta pLysS and BL21 (DE3*) E.coli* cells were transformed with a plasmid encoding for the CBD of MMP-2, carrying either a His_6-MBP tag or a His_6-SUMO tag, respectively. The cells were grown in LB at 37°C and induced with 200 µM IPTG overnight at 18°C. The cleared lysate was purified using a nickel column (His Trap). The eluted fractions were pooled and the tag was removed by cleaving with a histidine-tagged 3C protease or a histidine-tagged SUMO protease overnight at 4°C. Finally, the mixture was then passed back down the nickel column (reverse purification) and the protein collected in the flow through, while the protease and the tags were in the elution fraction. L: Lysate; S: Supernatant; I: Inclusion Bodies; FT: Flowthrough; W: washes; 5-8: His Trap fractions; P: Pool of His-Trap fractions; PD: Post Dialysis; FTr: Flowthrough reverse; Er: Elution reverse; CBD_M: CBD initially carrying the MBP tag; CBD_S: CBD initially carrying the SUMO tag

3.4.1.3. Small scale screening for optimal expression in minimal media

Considering the cost of the labelling, the expression was carried out in unlabelled media at first to ensure successful expression. Subcloning from the pOPINM to the pOPINB vector containing a MBP tag was also required for the MBP construct as Rosetta pLysS (DE3) cells were unable to grow in minimal media.

A small scale screening experiment was performed again to determine the optimal conditions for expression of a soluble CBD in minimal media.

The results below (Figure 3-4) enabled us to identify only two conditions where some soluble protein was obtained, namely a production phase of 5 hours at 25°C with 1 mM of IPTG and an overnight production at 18° C with 800 µM IPTG.

Those were only successful with the SUMO construct. They were both scaled up to 1L and subjected to the His Trap and reverse purification as described in the Materials and Methods section.

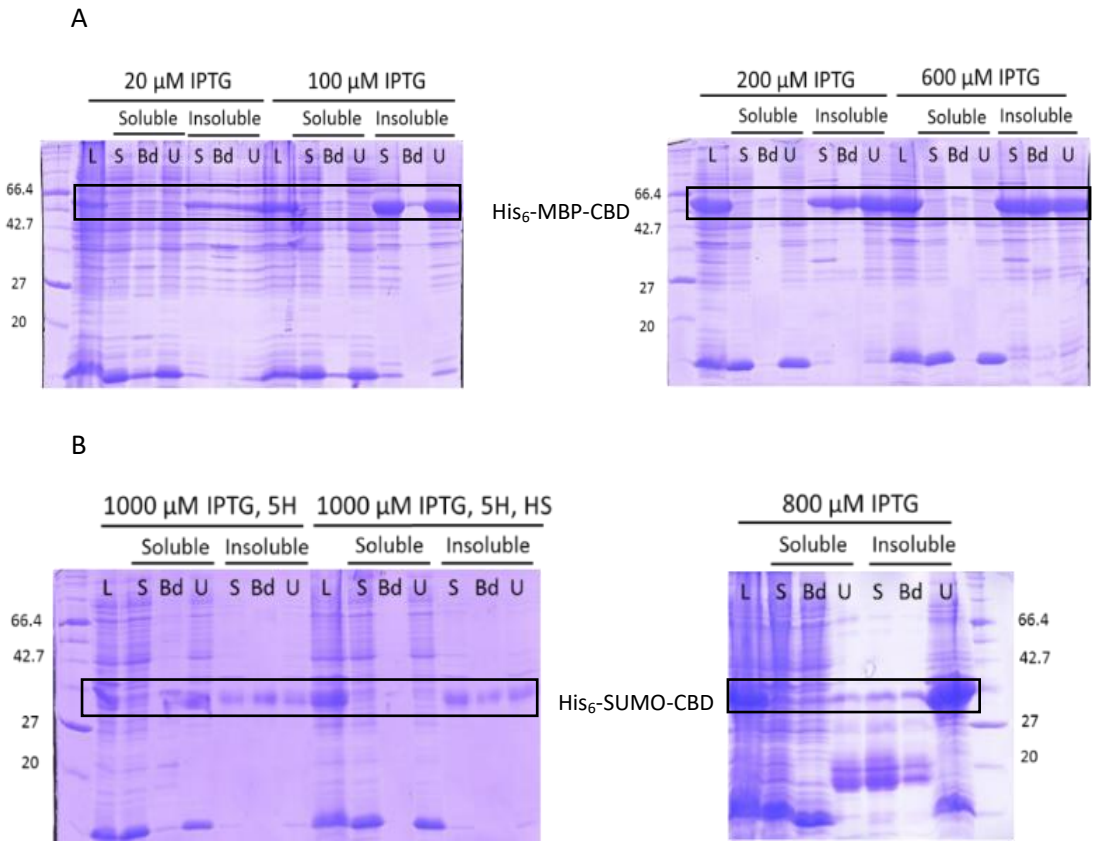


Figure 3-4: SDS PAGE showing the screening for expression of the His_6-MBP-CBD (A) and His_6-SUMO-CBD (B) in minimal media. *E.Coli* BL21 (DE3) cells were transformed with a plasmid encoding for the CBD of MMP-2, either carrying a His_6-MBP tag or a His_6-SUMO tag. Cells were grown in 10 mL minimal media until reaching an OD (600 nm) of 0.6. Cells were induced with various concentrations of IPTG overnight at 18°C, or for 5 hours at 25°C. Cells were then lysed and both the soluble and insoluble fractions were subjected to a nickel purification. The same conditions were tested for both constructs, but only a subset is presented in this figure. L: Lysate; S: Supernatant; Bd: Nickel bound; U: Unbound; HS: heat-shocked: growth at 30 °C, switch at 42°C when OD=0.1 (heat-shock), induction when OD=0.6, for 5 hours at 25 °C.

3.4.1.4. Large scale expression with the optimised conditions in minimal media

The scale-up led to an expression pattern where most of the protein was found in the inclusion bodies, although a small proportion was recovered in the soluble fraction (not shown). This process and the final yield were not suitable for further NMR experiments and an alternative method for studying the folding of the protein was required.

3.4.1.5. DTNB assay

Folded CBD contains 6 disulfide bonds; these are found in the native protein and are important to enable the CBD to bind gelatin with the optimal efficacy. An assay which quantifies the number of disulfide bridges within a protein was used to verify the presence of these bridges and enabled an early assessment of whether the expressed protein was likely to be folded or not.

The DTNB assay was therefore used to quantify the number of free cysteines within a molecule. It showed that only half of the disulfide bridges were formed in the protein originally carrying the MBP tag, and 3 free cysteines were found in the protein originally carrying the SUMO tag. This result correlated with our previous 1D NMR data highlighting the sub-optimal folding of the CBD in *E.coli.*

Alternative protocols and expression systems were therefore envisioned in order to obtain a correctly folded protein. The ability of mammalian cells to express a correctly folded CBD was therefore tested.

3.4.2. Expression of the CBD in HEK 293T cells

3.4.2.1. Transfection with Lipofectamine: proof of concept

Mammalian cells are thought to be better suited for the expression of proteins with post-translational modifications. HEK 293T cells were therefore tested for their ability to express the CBD. Transfection of the cells was carried out with the His_6-MBP-CBD construct, which vector pOPINM is suitable for mammalian expression. Results showed successful expression of the protein by Western Blot in the lysate and on the nickel resin (Figure 3-5). This was achieved using Lipofectamine 2000 as the transfection reagent, which would have been cost-prohibitive for a large scale expression of the protein. Other transfection methods were therefore tested in HEK 293T cells with cheaper transfection reagents, such as PolyethyleneImine and Calcium Chloride.

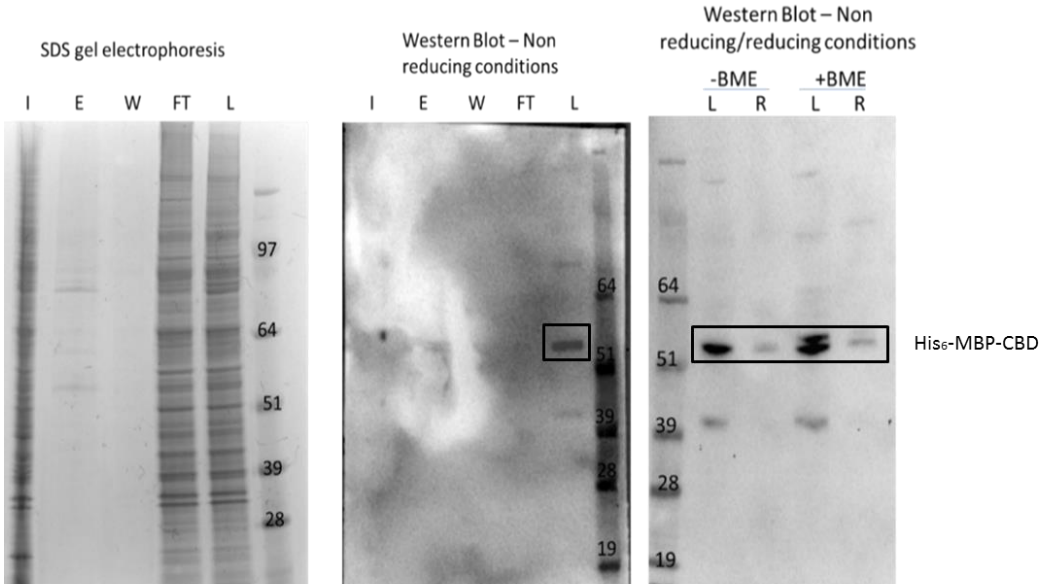


Figure 3-5: SDS PAGE and Western Blots showing successful expression of the His_6-MBP-CBD in HEK 293T cells transfected with Lipofectamine 2000. HEK 293T cells were transfected with a plasmid encoding for the CBD of MMP-2, carrying a His_6-MBP tag with Lipofectamine 2000. Cells were harvested and lysed 24 hours after being transfected. Both the soluble and insoluble fractions were subjected to a nickel purification. SDS PAGE was run and proteins were then transferred onto a nitrocellulose membrane under both reducing and non-reducing conditions. Proteins were detected with a mouse anti-Histidine antibody followed by an anti-mouse antibody. L: Cleared Lysate; FT: Flowthrough; W: Washes; E: Elution; I: Inclusion Bodies; R: Nickel resin

3.4.2.2. Transfection with Calcium Chloride and Polyethylenimine: Small scale screening for optimal expression

Both Polyethylenimine (PEI) and Calcium Chloride were used as transfection reagents. In the former case, the ratio of PEI:DNA (1:1, 3:1 and 5:1) was optimised, as well as the time of harvesting cells (24, 48 or 72 hours post-transfection). For transfection with Calcium Chloride, 3 amounts of DNA per well (1, 3 and 5 µg) were tested, as well as 3 incubation times of the complex (2, 6 and 20 hours). Cells were harvested 48 hours after transfection.

Cells were lysed and the cleared supernatants were loaded on a gel, transferred onto a membrane and blotted using a mouse anti-Histidine/anti-mouse antibody combination.

Results showed very low levels of expression in all samples, except for cells transfected with Calcium Chloride with either 3 or 5 µg of DNA and incubated for 20 hours with the DNA-Calcium Chloride complex (Figure 3-6). The GAPDH control confirmed an even lysis efficiency for all the samples transfected with Calcium Chloride. However, the molecular weight of the expressed protein seemed slightly lower than expected, suggesting that a cross-reaction of the antibody with a different protein could have happened. In spite of this observation, it was decided to scale up the expression in HEK 293T cells using the Calcium Chloride transfection method with 5 µg of DNA per well and an incubation of the complex for 20 hours.

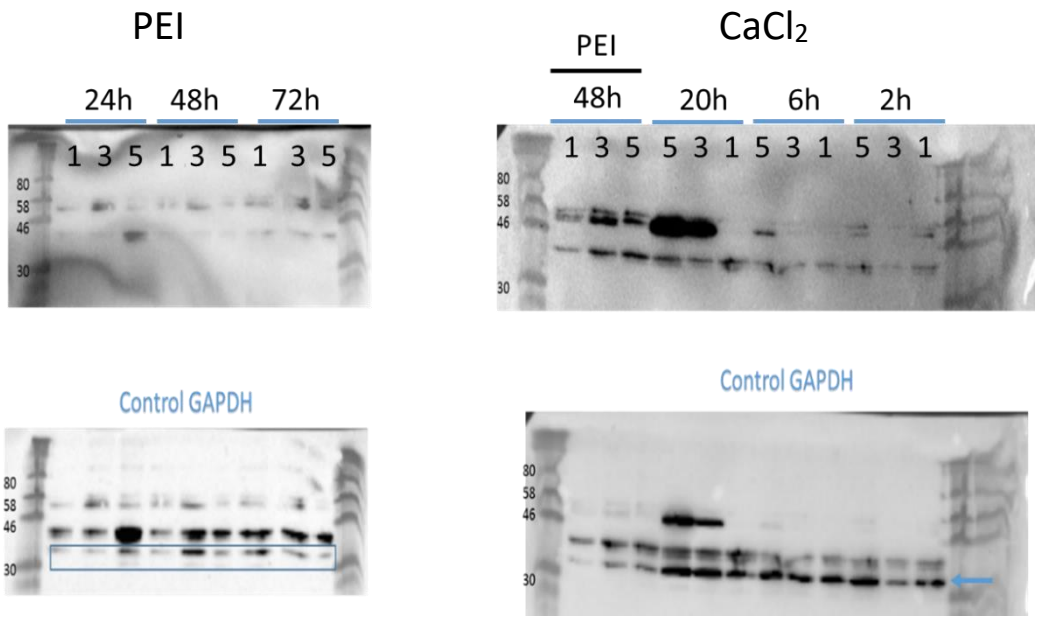


Figure 3-6 Western Blots showing screening for expression of the His_6-MBP-CBD in HEK 293T cells transfected with either PEI (left) or Calcium Chloride (right). HEK 293T cells were transfected with a plasmid encoding for the CBD of MMP-2, carrying a His_6-MBP tag with either PEI or Calcium Chloride ($CaCl_2$). SDS PAGE was run on the lysates and proteins were then transferred onto a nitrocellulose membrane. Proteins were detected with a mouse anti-Histidine antibody followed by an anti-mouse antibody. An anti GAPDH antibody was used as a control of the lysis efficiency in all samples. The numbers labelled on the gels either correspond to the PEI:DNA ratios (left, PEI transfection) or to the amount of DNA used (right, $CaCl_2$ transfection) .

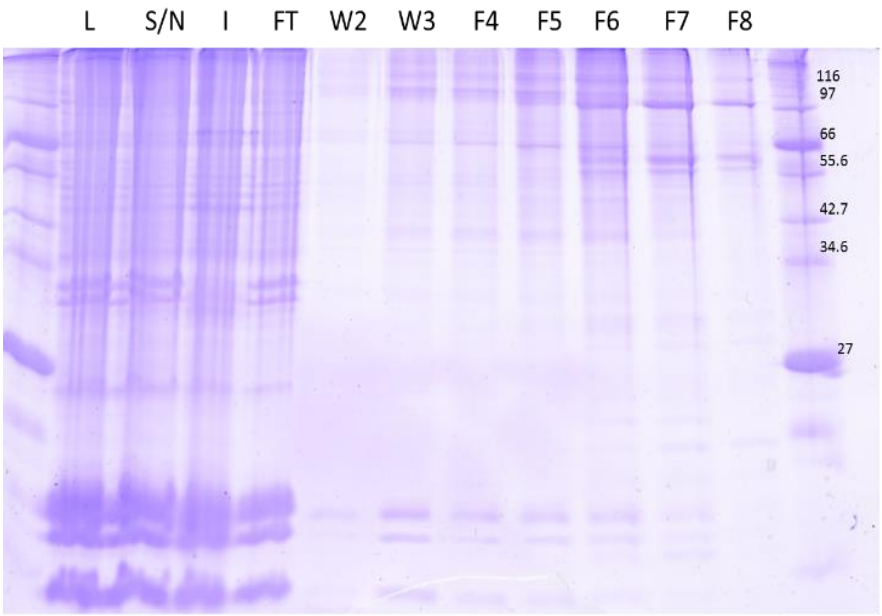
3.4.2.3. Large scale expression with the optimised conditions

Transfection of HEK 293T cells with 5 µg of DNA and Calcium Chloride, incubated for 20 hours, was chosen as the best condition and scaled-up to 6 15-cm dishes seeded at 9 500 000 cells per dish.

Cells were lysed and a His Trap purification was performed on the cleared lysate using a 1 mL column.

Unfortunately, the expression levels appeared to be very low as a very faint band was observed at the expected size (62.5 kDa) in only 3 fractions (Lanes W3, F4 and F5, Figure 3-7). The level of expression in these conditions in HEK 293T cells was therefore too low for follow-up studies, and this strategy was put on hold as a more promising one was being developed, using Shuffle *E.coli* cells.

A



B

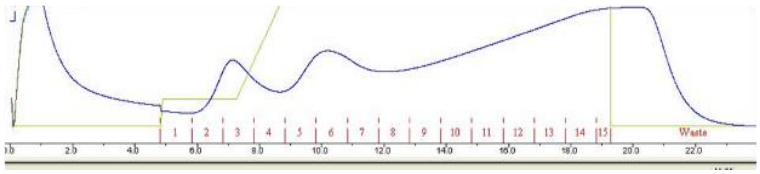


Figure 3-7: SDS PAGE (A) and chromatogram (B) showing the expression of the His_6-MBP-CBD in HEK 293T cells transfected with Calcium Chloride. HEK 293T cells were transfected with 5 µg of a plasmid encoding for the CBD of gelatinase, carrying a His_6-MBP tag, using Calcium Chloride. Cells were incubated with the plasmid for 20 hours and harvested and lysed 48 hours after being transfected. The cleared supernatant was subjected to an affinity chromatography using a 1mL His Trap column. A SDS PAGE was run on different fractions. L: Lysate; S/N: cleared supernatant; I: Inclusion bodies; FT: Flowthrough; W: Washes; F: Fractions (from corresponding chromatogram)

3.4.3. Expression of the CBD in *E.coli* Shuffle cells

3.4.3.1. Small scale screening for optimal expression

Small scale expression trials were set up with Shuffle cells, which were transformed with either the pOPINS or pOPINB construct, optimising the concentration of IPTG only. Proteins were expressed at 18°C overnight. From this screening, 200 µM and 600 µM of IPTG were selected as the optimal concentrations for induction for the MBP and SUMO constructs (Figure 3-8), respectively, and scaled up to 2L flasks.

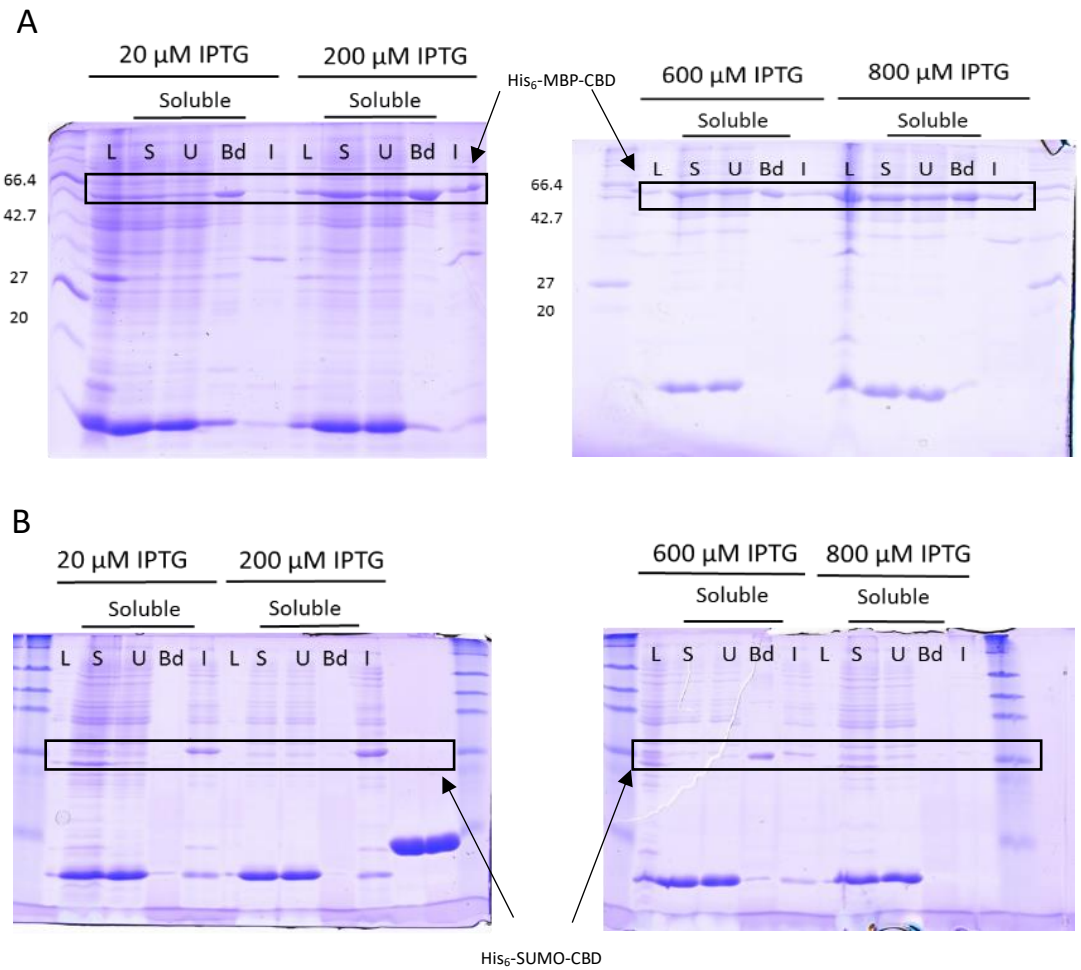


Figure 3-8: SDS PAGE showing the expression of the MBP-CBD (A) and SUMO-CBD (B) in *E.coli* Shuffle cells. *E.Coli* Shuffle (DE3) cells were transformed with a plasmid encoding for the CBD of MMP-2, carrying either a His_6-MBP tag or a His_6-SUMO tag, respectively. Cells were grown in 10 mL minimal media until reaching an OD (600 nm) of 0.6. Cells were induced with various concentrations of IPTG overnight at 18°C. Cells were then lysed and the soluble fraction was subjected to a nickel purification, while the insoluble fraction was solubilised in 8M urea. L: Lysate; S: Supernatant; U: Unbound; Bd: Nickel bound; I: Inclusion Bodies

3.4.3.2. Large scale expression with the optimised conditions in LB and minimal media

These conditions enabled soluble expression of proteins from both constructs, as shown in Figure 3-9, in both minimal media and LB (not shown). However, some high molecular weight contaminants were observed after the reverse purification, highlighting the need for a third purification step. Both Ion Exchange and Gel Filtration Chromatography were tested.

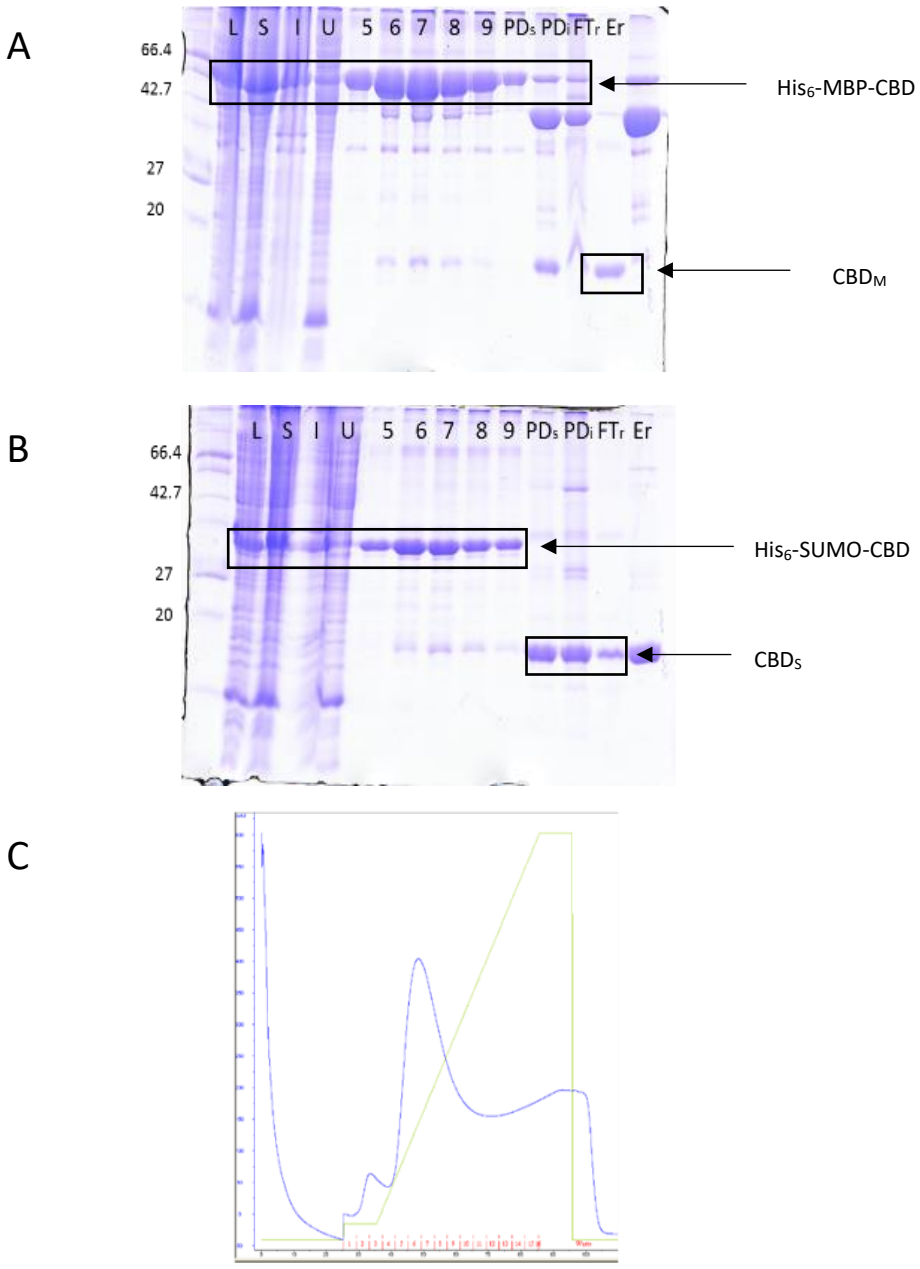


Figure 3-9: SDS PAGE gel showing the expression of the His_6-MBP-CBD (A) and His_6-SUMO-CBD (B) in minimal media and the corresponding chromatogram (C). A and B: *E.coli* Shuffle T7 Express (DE3) cells were transformed with a plasmid encoding for the CBD of MMP-2, carrying a His_6-MBP tag or a His_6-SUMO tag, respectively. The cells were grown in minimal media at 37°C and induced with 200 µM IPTG or 600 µM IPTG (for the MBP tagged and SUMO tagged constructs, respectively) overnight at 18°C. The cleared lysate was purified using a nickel column (His Trap). The eluted fractions were pooled and the tag was removed by cleaving with Histidine-tagged 3C or SUMO proteases overnight at 4°C. Finally, the mixture was then passed back down the nickel column (reverse purification) and the protein collected in the flow through, while the protease and the tags were in the elution fraction. L: Lysate; S: Supernatant; I: Inclusion Bodies; U: Unbound; W: washes; 5-9: His Trap fractions; PDs: Post Dialysis soluble; Pdi: Post dialysis insoluble FTr: Flowthrough reverse; Er: Elution reverse. C: Chromatogram corresponding to the His Trap purification of the His_6-SUMO-CBD. Fractions 5 to 9 were loaded on the gel. Blue line: UV trace; Green line: Concentration of imidazole

3.4.4. Optimisation of the purification method: polishing step

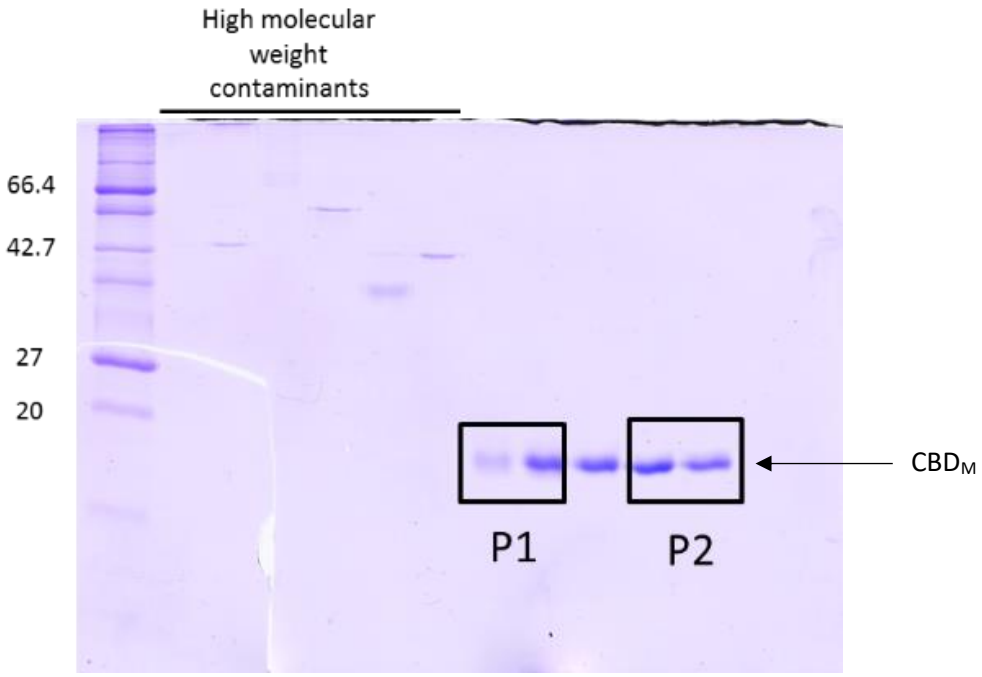
3.4.4.1. Gel Filtration

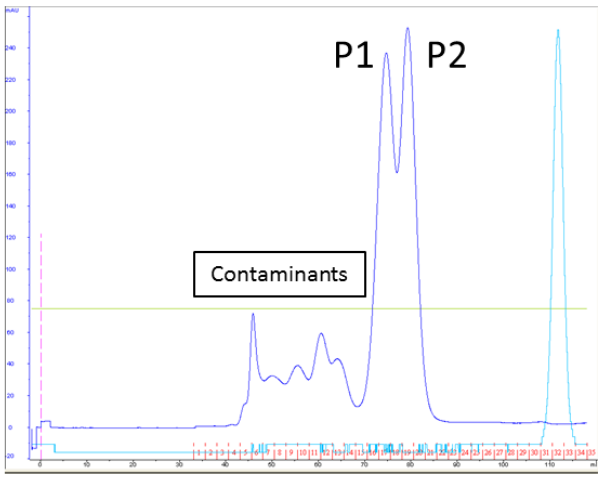
The size exclusion chromatography enabled pure proteins to be obtained, but most of it was lost on the HiLoad 16/60 column (Figure 3-10). This strategy also gave different elution profiles according to the construct which was used: the SUMO construct gave one single peak while the MBP one provided two consecutive peaks, suggesting the presence of two different species, which will be further investigated in the next section.

Because protein precipitation could explain the low recovery, a different column was tested (Superdex 75 10/300) and the concentration of salts was increased from 150 to 300 mM. None of these options significantly improved the amount of protein which was recovered (not shown).

It was concluded that the Gel Filtration was not a suitable polishing step to recover enough functional CBD, and it was decided to test an Ion Exchange Chromatography (IEC) instead.

A


High molecular weight contaminants
66.4
42.7
27
20
P1
P2
CBDM


P1
P2
Contaminants

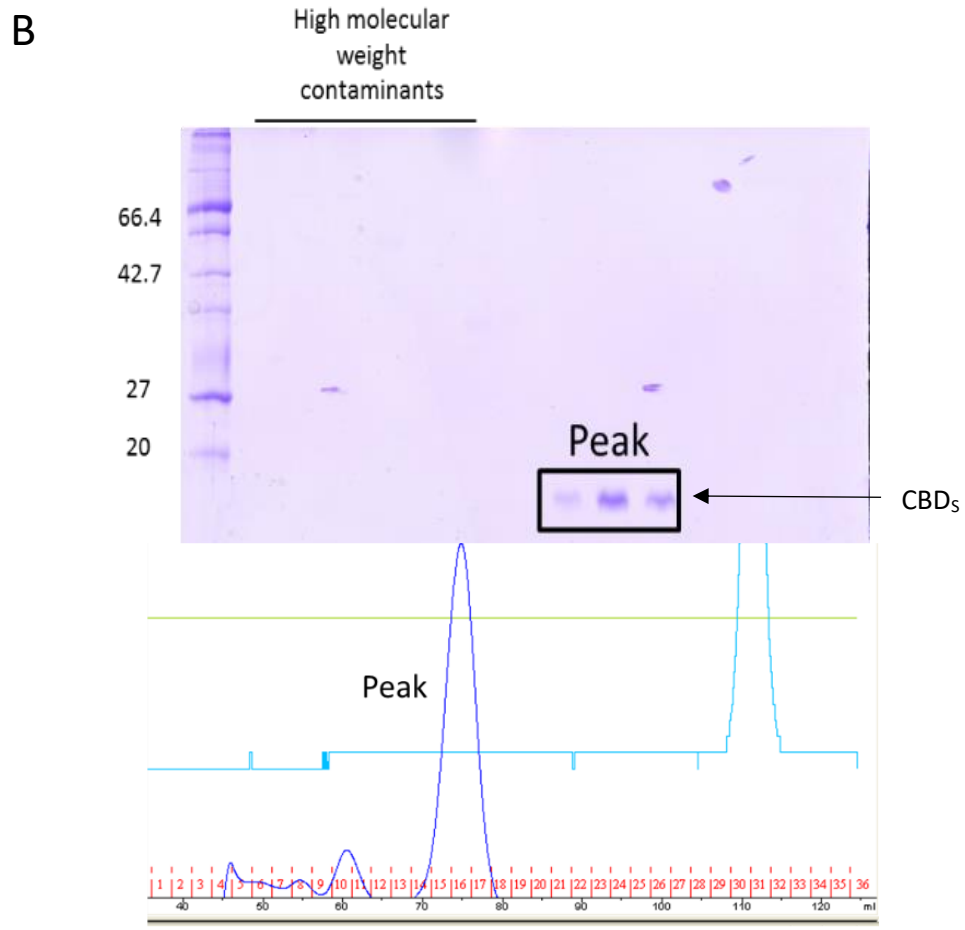


Figure 3-10: SDS PAGE and corresponding chromatograms showing the elution profile and purity of both the CBD_M (A) and CBD_S (B) construct on a Gel Filtration column. The CBD of MMP-2, originally carrying either a His_6-MBP tag or a His_6-SUMO tag, was purified by nickel column and reverse nickel column purification, as previously described. The protein samples were then concentrated to up to 2 mL and loaded on an equilibrated HiLoad 16/60 s75 Superdex column and eluted isocratically.

3.4.4.2. Ion Exchange Chromatography

The Ion Exchange strategy did not increase the purity of the originally MBP-tagged sample, as high molecular weight contaminants were still present in the eluted fractions (Figure 3-11, lanes 8 to 12). However, the construct originally carrying the SUMO tag gave a better outcome, as the CBD eluted specifically at a lower salts concentration than the contaminants (lanes 8 and 9, Figure 3-12).

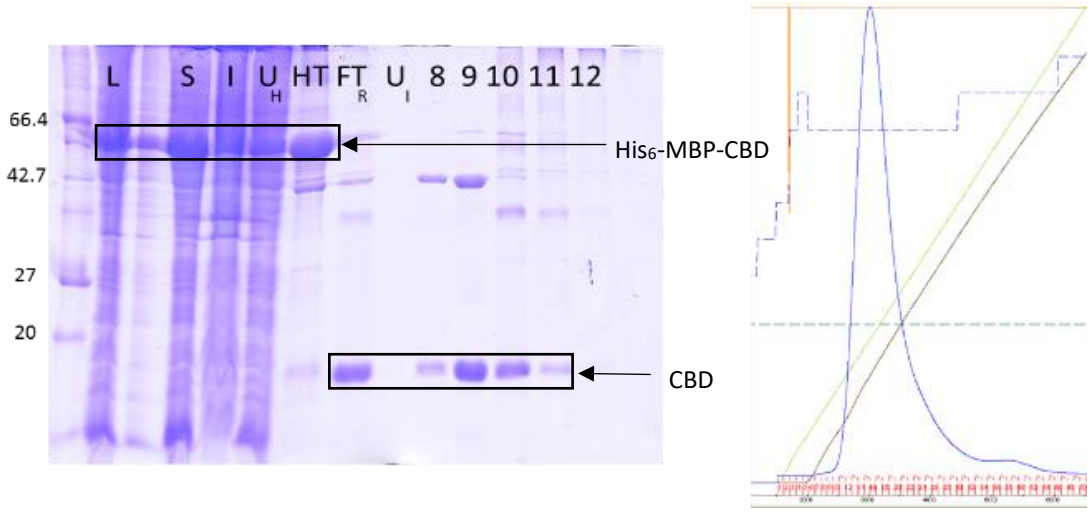


Figure 3-11: SDS PAGE and corresponding chromatogram showing the elution profile and purity of the CBD_M sample following linear elution from the Ion Exchange Q FF 5 mL column. The CBD of MMP-2, originally carrying a His_6-MBP tag, was purified by nickel column and reverse nickel column purification, as previously described. The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a linear gradient of sodium chloride. L: Lysate; S: Supernatant; I: Inclusion Bodies; U_H: unbound from His Trap; HT: Fraction from His Trap; FT_R: Flowthrough reverse; U_I: Unbound from IEC; 8-12: Fractions from IEC.

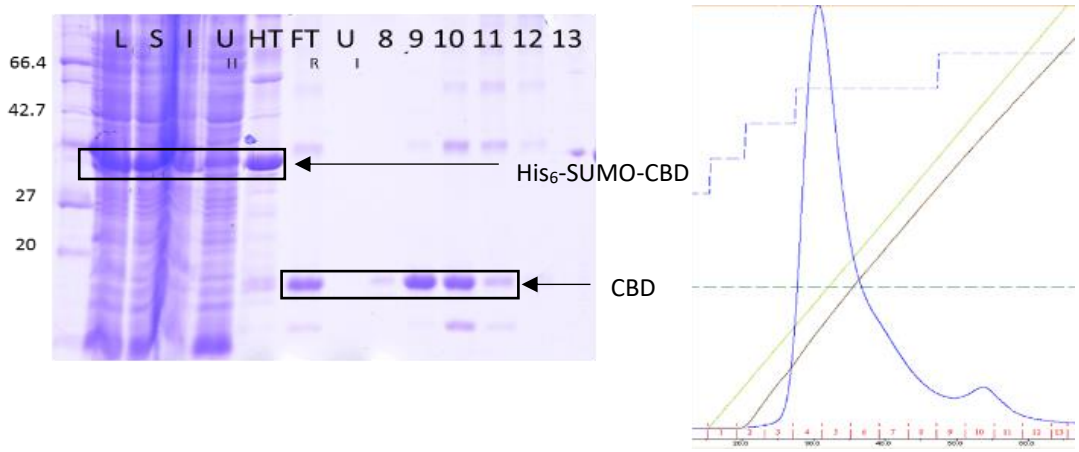


Figure 3-12: SDS PAGE and corresponding chromatogram showing the elution profile and purity of the CBD_S sample following linear elution from the Ion Exchange Q FF 5 mL column. The CBD of MMP-2, originally carrying a His_6-SUMO tag, was purified by nickel column and reverse nickel column purification, as previously described. The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a linear gradient of sodium chloride. L: Lysate; S: Supernatant; I: Inclusion Bodies; U_H: unbound from His Trap; HT: Fraction from His Trap; FT_R: Flowthrough reverse; U_I: Unbound from IEC; 8-12: Fractions from the IEC; 13: Fraction from the second peak of the IEC

It was therefore decided to test a gradient elution on the CBD_S construct, using a step elution at 24% of IEC Buffer B, which was the concentration at which the CBD started to get eluted in the previous experiment.

A sample free from high molecular weight impurities was obtained, although a low molecular weight contaminant could still be observed (Figure 3-12). This contaminant was later identified as being the cleaved histidine-tagged SUMO.

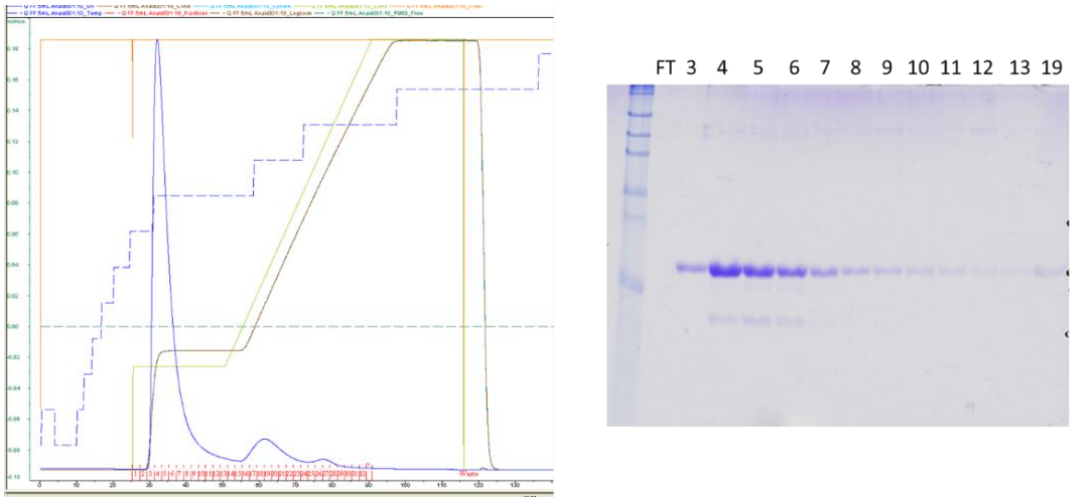


Figure 3-13: SDS PAGE and corresponding chromatogram showing the elution profile and purity of the CBD_S sample following a step elution from the Ion Exchange Q FF 5 mL column. The CBD of MMP-2, originally carrying a His_6-SUMO tag, was purified by nickel column and reverse nickel column purification, as previously described. The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a step elution of 24 % of IEC B. FT: Flowthrough from IEC; 3-19: Fractions from the IEC.

In order to get rid of this contaminant, a mono Q column was tested at pH 6.5 and 7.0, which are closer to the isoelectric point of the protein (*i.e.* 5.0). Unfortunately, this strategy did not allow enough recovery of pure protein to be considered as a suitable method (not shown).

Overall, it was decided that the step elution of the CBD_S construct from the Ion Exchange column, described in Figure 3-13, was the best strategy to recover enough protein with a high level of purity (equal to 90 %). The next step was to characterise the integrity of the protein fold and the function of this CBD_S sample.

3.4.5. Protein characterisation

3.4.5.1. Identity and structure

3.4.5.1.1. Identity

The protein identity was confirmed by a MALDI-TOF Mass Spectrometry analysis on a sample digested with trypsin. 67 ions were identified with more than 99 % of probability, each corresponding to a part of the sequence of the CBD of MMP-2. The identified peptides span the sequence highlighted in yellow in the figure below, covering almost the entirety of the CBD sequence (Figure 3-14).

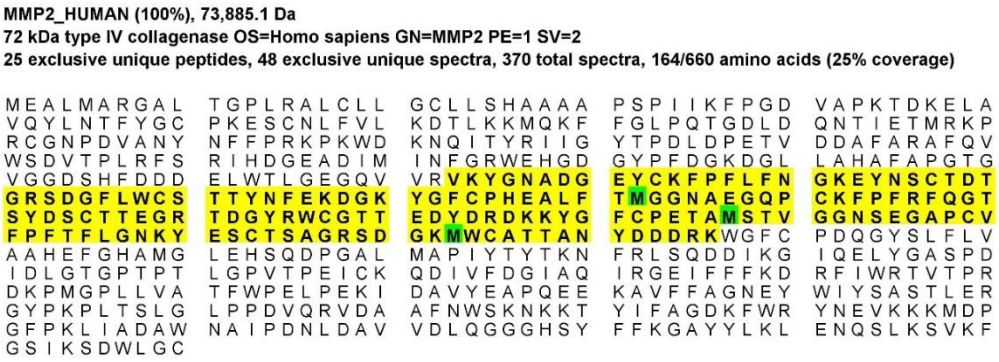


Figure 3-14: Mass Spectrometry results confirming the identity of the CBD. Ions obtained from tryptic digests of the CBD_S sample were analysed by MALDI-TOF. Their masses corresponded to peptides within the CBD sequence, which are indicated in yellow.

3.4.5.1.2. Disulfide bridges

One of the main structural features of the CBD is the presence of 6 disulfide bonds, which are also critical for the correct function of the protein. Their presence was first confirmed with a DTNB assay, which quantifies the number of disulfide bridges within a protein. The CBD_M samples from the Gel Filtration chromatography described in Figure 3-10 were also submitted to the DTNB assay. The results are summed up in Table 3-1 below. They show that the CBDs sample obtained after a step elution on the Q FF column was predominantly folded, with no free cysteines observed. The two species identified earlier with the CBD_M sample (*i.e.*

the CBD originally carrying a MBP tag) showed different folding states: the one corresponding to the first peak on the chromatogram in Figure 3-10 was unfolded, with 3 free cysteines, while that corresponding to the second peak was mostly folded, with no free cysteines. These results correlated with the 1D NMR spectra traces which were obtained for each sample (not shown). Obtaining mixtures of folded and partially folded proteins using the *E.coli* expression system is not uncommon since there is a fine balance between protein synthesis and protein folding *in vivo*.

Table 3-1: Characterisation of the CBD in terms of folding. The DTNB assay enables quantification of the number of free cysteines in each protein samples, as the reagent, namely, 5-dithio-bis-(2-nitrobenzoic acid), reacts with free sulphydryl, producing a coloured species, TNB2-. The characteristics of the 1D proton NMR spectra correlated with the data from the DTNB assay. The CBD_S sample was obtained after the Ion Exchange purification with a step gradient, as described in the previous section. The CBD_M samples come from the Gel Filtration chromatography.

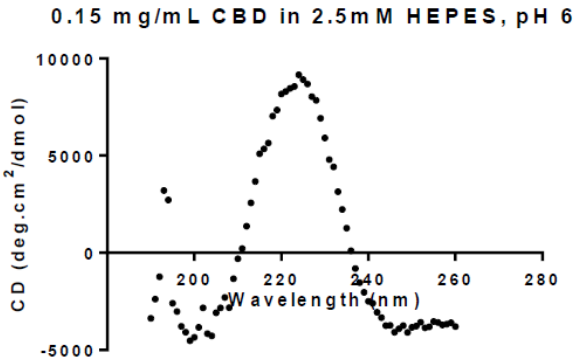
	CBD_S	**CBD_M peak 1**	**CBD_M peak 2**
DTNB assay	0.05 free cysteine/protein	3.3 free cysteine/protein	0.4 free cysteine/protein
1D NMR spectra	Predominantly folded	Unfolded	Predominantly folded

An intact mass analysis was used to further confirm this finding. The measured intact mass of the CBD_S sample was 19 628 Daltons, which corresponds to a 12-dalton deficit compared to the theoretical mass derived from the sequence (19 640 Da). This difference correlates with the formation of 6 disulfide bridges and the concomitant loss of 12 hydrogens.

3.4.5.1.3. Secondary structure

The secondary structure of the CBD_S protein was then studied by Circular Dichroism (CD). The CD spectrum obtained showed characteristics consistent with a folded Fibronectin domain-rich protein, with a maximum absorption at 224 nm and a minimum at 198 nm, with a mixture of alpha turn and beta sheet [143]. A spectrum of a thermodenatured CBD_S was also acquired at 90°C to show the disappearance of the maxima at 224 nm (Figure 3-15).

A



B

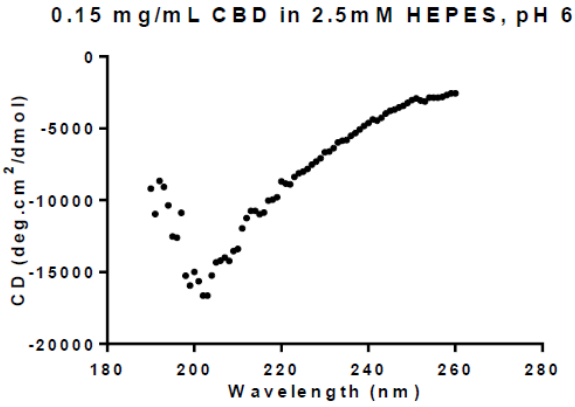


Figure 3-15: CD spectrum of the native (A) and thermodenatured (B) CBD_S sample. The CBD_S sample purified with a step elution from the Ion Exchange QFF column was buffer exchanged and submitted to CD analysis. Both spectra were acquired with a protein concentration of 0.15 mg/mL in 2.5 mM HEPES, pH 6.5. The CD signal was acquired from 260 to 190 nm with a JASCO instrument.

3.4.5.1.4. Tertiary structure: NMR characterisation

A 2D NMR spectrum of the free protein, expressed in ^{15}N labelled minimal media, was then acquired. The spectrum was similar to previously published data [140]; some precipitation or aggregation was observed. Some additional peaks of minor intensity were also observed (Figure 3-16).

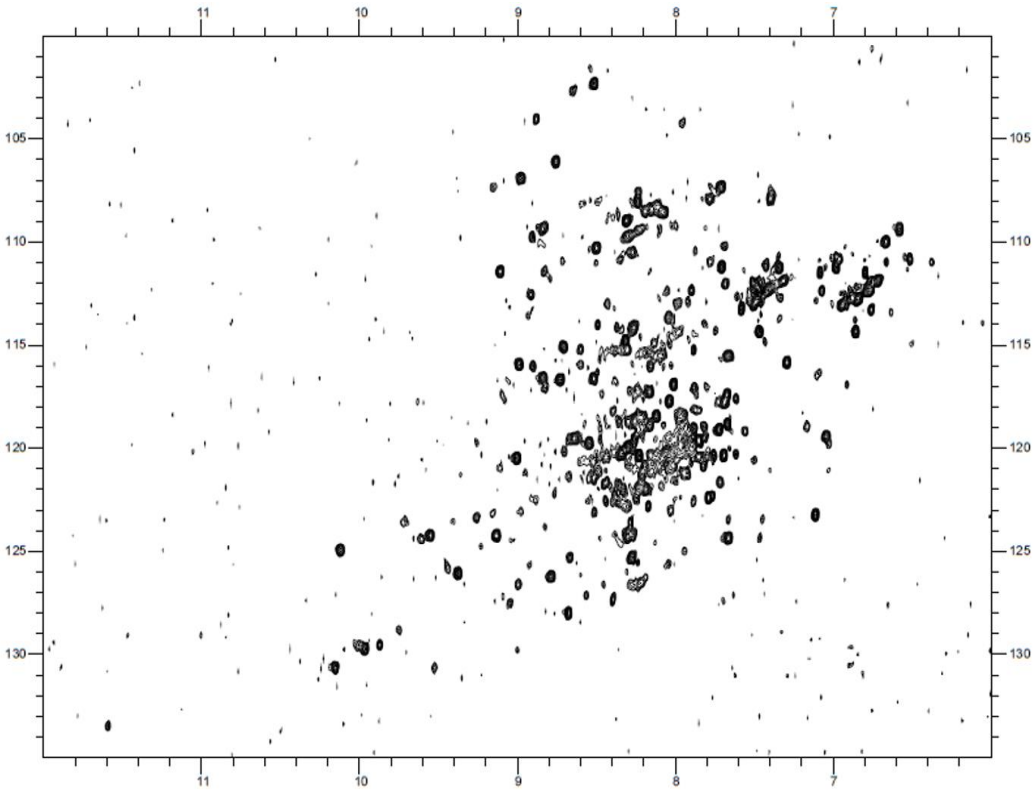


Figure 3-16: ^{1}H-^{15}N Heteronuclear Single Quantum Correlation (HSQC) spectrum of the CBD_S sample at 298 K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O, 115 µM. The spectra were acquired on a 800 Hz Bruker spectrometer instrument (Bruker Avance III fitted with [1H, 15N, 13C]-cryoprobes). Each peak on the spectrum corresponds to an amide proton in the CBD. The position of the peaks in the spectrum is determined by the chemical environment of the atoms of the protein.

When overlaying the spectrum of the CBD_S with that of the His_6-SUMO tag, it appeared that the observed additional peaks matched, indicating that the final protein sample was contaminated with the SUMO tag (Figure 3-17).

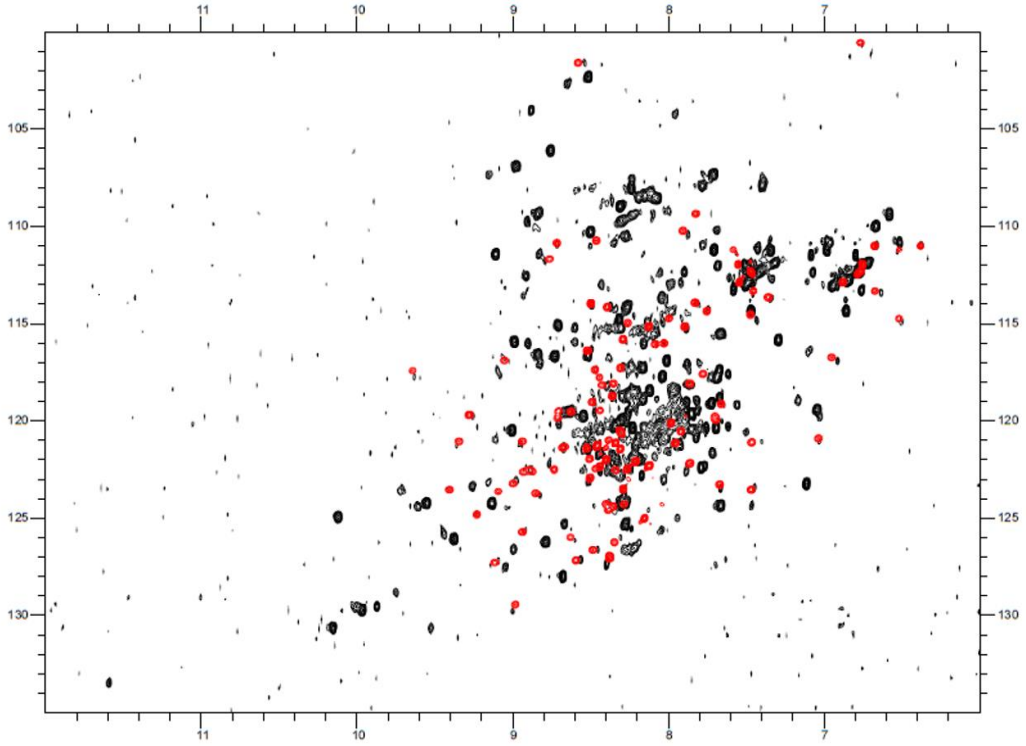


Figure 3-17: Overlay of the ^{1}H-^{15}N HSQC spectrum of the CBD_S sample at 298K in 25 mM NaP, pH 6.5, 10 % D_2O (black) with that of the His_6-SUMO tag (red). The purified CBD_S sample was buffer exchanged against 25 mM sodium phosphate (NaP), pH 6.5, concentrated to 115 µM and 128 spectra were acquired at 25°C on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the CBD. The position of the peaks in the spectrum is determined by the chemical environment of the atoms of the protein. The fact that peaks from the CBD and from the SUMO tag overlap indicates that a residual SUMO tag is found in the CBD sample.

Different purification strategies were therefore envisioned to try to purify the CBD further and isolate it from the tag.

As there is no cysteine in the sequence of the SUMO tag, we concluded that its interaction with the CBD was not covalent. A hydrophilic interaction would not have been maintained with the 500 mM NaCl already present in the purification buffers. It is, therefore, possible that the interactions between the CBD and contaminant protein(s) are mediated by hydrophobic contacts. Hence, 0.05 % Tween (v/v) was included in all purification buffers,

except those of the Ion Exchange step. Unfortunately, this led to more protein precipitation and no improvement in purity (not shown).

As this project does not require a degree of purity higher than 90 %, it was decided to keep the purification protocol as previously described, and to accept the presence of the SUMO tag as a contaminant.

3.4.5.2. Activity

After expressing the protein with an adequate level of purity and characterizing its folding, we went on to determine whether the expressed protein was able to bind to its target, *i.e.* gelatin.

3.4.5.2.1. Gelatin sepharose binding

The CBD_S protein sample purified with a step gradient elution from the Q FF column was first tested for its ability to bind to gelatin using a small scale gelatin sepharose affinity chromatography. Although some protein was lost in the flow through, most of it was recovered on the resin, demonstrating that the protein which has been obtained is both folded and functional as it was able to bind to the immobilised gelatin (Figure 3-18).

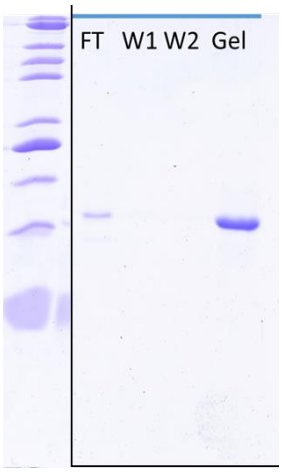
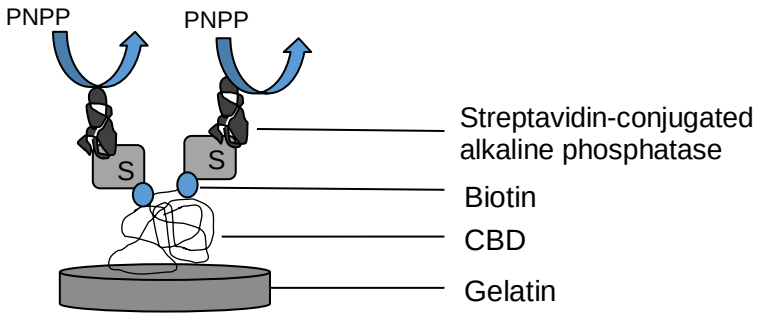


Figure 3-18: SDS PAGE showing the binding of the CBD_S to gelatin sepharose. The CBD_S protein sample from the Ion Exchange was tested for its ability to bind gelatin using a gelatin sepharose resin. 100 µL of resin was spun down, washed with ethanol and equilibrated before applying the samples. The mixture was spun again and the flowthrough removed (FT) before washing two times with the binding buffer (W: Washes). The bound protein was mixed with 2X SDS loading buffer and loaded on the gel (Gel: gelatin-bound).

3.4.5.2.2. Binding assay

A more quantitative gelatin-binding test was then carried out using type II gelatin coated plates. As described in the Materials and Methods section, the protein was first biotinylated before being added over a range of concentrations to gelatin-coated plates. The ability of alkaline phosphatase-streptavidin to convert its substrate, P-Nitrophenyl Phosphate (PNPP), into a coloured product was measured at 405 nm, and used as an indirect indication of the amount of protein bound to gelatin (Figure 3-19, panel A). The specificity of the signal was demonstrated by the fact that a reduced and alkylated version of the CBD did not show any binding (not shown), highlighting the need of the six disulfide bonds for the correct function of the protein. The CBD_S was shown to bind with a Kd of 20.4 ± 2.83 nM (Kd ± SEM, n = 3) (Figure 3-19, panel B, for a representative binding curve), in line with previous findings in an identical assay which yielded a Kd of 50 nM for a CBD construct containing a Histidine tag. [145]

A



B

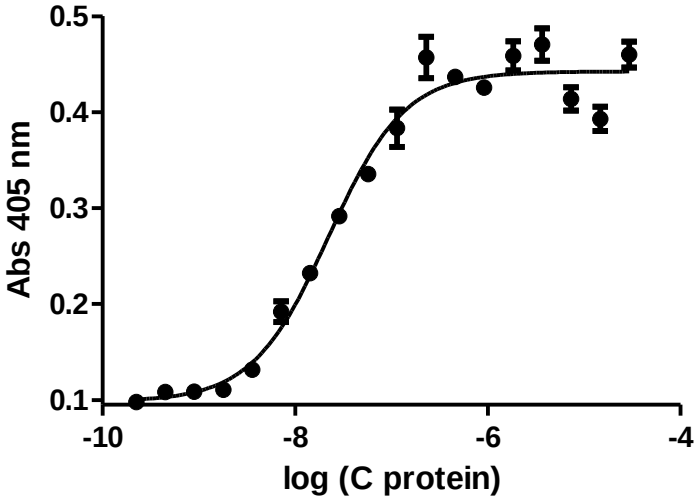


Figure 3-19: Binding of the CBD to type II gelatin plates. A: The binding assay relies on the use of plates coated with type II gelatin (thermodenatured collagen). The biotinylated protein was added over a concentration range before adding streptavidin-coupled alkaline phosphatase. Its substrate, namely PNPP, was added and converted into a coloured product which formation was measured by absorbance at 405 nm. B: An example of one of the binding curves is shown. Error bars represent standard deviations of three replicates.

3.5. Discussion

3.5.1. General conclusion

This chapter has described the successful expression and purification of a folded and functional CBD in the cytoplasm of the Shuffle *E.coli* strain. To the best of our knowledge, this is the first time that this protein has been expressed in the soluble fraction of *E.coli*. The protein was expressed in 2L Flasks, giving a consistent yield of 3.5 mg/L of culture. This yield is lower than that obtained in the original protocol for expressing the CBD, as the refolding strategy gave a yield of 15 mg/L. However, our strategy seems easier, quicker, and more reproducible, and it does not require elution of the protein from its substrate in the purification protocol. In addition, and contrary to the original refolding protocol, our strategy yields a native CBD, as its N-terminal is intact after cleavage of the SUMO tag by the SUMO protease.

This outcome was achieved after the more classical strains of *E.coli*, *i.e.* Rosetta and BL21 Star, had failed to enable the expression of a correctly folded protein in the soluble fraction, despite extensive optimisation of the expression conditions. The reducing environment of the cytoplasm of these strains is probably the reason why most of the protein accumulated in inclusion bodies. This might also explain why the soluble fraction contained misfolded versions of the CBD.

Mammalian expression was also tested in HEK 293T cells, whose folding machinery is more adequate for the expression of disulfide-bridged proteins. However, using this host led to a very low level of expression and this was not suitable for the subsequent work to be undertaken in this project.

As mentioned earlier, the purification protocol was specifically designed for the chosen construct. It exploits the affinity of the histidine tag for nickel, enabling the capture of the protein on a His Trap column. A polishing step was added with an anion Ion Exchange

Chromatography using a Q column, which enabled the purification of the protein to 90 %, even though some of the initial SUMO tag remained in the final sample.

This contaminant did not affect the folding nor the function of the protein, which was shown to be able to bind to type II gelatin with a Kd of 20.4 nM, a value which was similar to that previously described for a refolded His_6-CBD (50 nM), confirming the validity of our method of expression and purification.

3.5.2. Future work

Are there other ways to express the CBD with a higher yield?

This chapter has demonstrated that it is possible to express the CBD of MMP-2 as a folded and fully functional protein, in the cytoplasm of Shuffle *E.coli* strains with a consistent yield of 3.5 mg/L of culture.

Although fairy low compared to the common expression levels obtained in *E.coli*, this yield was nevertheless sufficient to carry out the first part of this project. However, when thinking about potential translation of this project into the clinic, a higher yield of protein will be needed to optimise time and resources.

The first and more immediate step to take would be to try to scale up the expression in the same host, using a bioreactor. The advantages of using a bioreactor, apart from the fact that higher volumes of culture are used, is that critical growth conditions such as pH and dissolved oxygen are controlled. The literature is limited on the use of the Shuffle cells in bioreactors, although one paper reported optimised conditions for growth for this specific strain [178]. It describes the use of glycerol as the carbon source and a temperature of growth of 30°C. These optimised conditions have led to yields as high as 3.2 grams of protein/10L of culture [178].

Another strategy to explore would be to target the protein to the periplasm of *E.coli*, where disulfide-bridged proteins are naturally expressed in this host. This would require the addition of a leader peptide in the cloning vector to target the protein to one of the sec or SRP pathways (chapter 2 of reference 176). The advantages of expressing a protein in this compartment is that it contains fewer proteases than the cytoplasm (chapter 5 of reference 176), and is generally less complex, leading to easier purification procedures (chapters 2 and 5 of reference 176). Chaperones are also naturally available in the periplasm to help with the folding of the recombinant protein (chapter 5 of reference 176).

An alternative option to periplasmic expression would be to co-express the CBD with chaperones in the cytoplasm of the host, which could favour its folding (chapter 3 of reference 176).

Alternatives to these options would be to express the protein in a different host with a more complex post-translational machinery, as that of the Shuffle cells still led to the accumulation of some of the CBD in the insoluble fraction. Yeasts might be an interesting host to test, as they combine the advantages of *E.coli* (fast growth, inexpensive) and of a eukaryotic system, enabling them to perform correct disulfide bridges formation [164]. They are easier to scale up than mammalian cells and present a secretion capacity which would facilitate downstream purification [164]. Interestingly, *Pichia Pastoris* was used to express one of the fibronectin domains of Fibronectin [179].

Why is the SUMO tag co-purified with the CBD and how to purify it further?

The NMR spectrum of the CBD showed that the final protein solution was contaminated with ~10 % of the SUMO tag. Even though it was decided that having the protein at a 90 % purity level was enough for the purpose of this project, it would be interesting to find out why the SUMO tag is still present after the last step of purification.

One simple explanation would be that the SUMO tag does not bind to the resin of the His Trap column during the reverse purification. There could be several reasons explaining the absence of binding:

1. The use of imidazole in the His Trap A buffer of the reverse purification. A low concentration of imidazole (*i.e.* 20 mM) is indeed usually used in the His Trap purification to prevent contaminants from binding to the column. However, the use of imidazole, even at a low concentration, in the reverse step could prevent the SUMO tag from binding onto the column.
2. The histidine tag present at the N-terminal of the SUMO tag sequence could also be subject to degradation, preventing the SUMO tag protein from binding onto the column.
3. Finally, the resin could also be saturated with the SUMO protease itself, preventing the tag from binding. Using stronger resins, such as the Cobalt ones, or even bigger columns, could be used to test this hypothesis and potentially solve the issue.

Another reason why the SUMO tag is found in the final CBD product could be that both proteins interact with each other. As outlined earlier, the proteins cannot interact covalently, as there is no cysteine in the sequence of the SUMO tag. The interaction could then happen through hydrophobic or hydrophilic interactions. However, neither the addition of Tween (for hydrophobic interactions) nor the presence of 0.5 M NaCl abolished these interactions.

Another underlying question is why the SUMO tag was not separated from the CBD during the Ion Exchange Chromatography. This may be due to the fact that both proteins share the same charge at the pH used for the purification*, i.e.* at pH = 8.0. As the theoretical pI of the CBD is actually much lower (5.0), using a lower pH could have solved the issue, although the pI of the His_6-SUMO tag is very close (5.71). This might explain why, when running Ion Exchange purifications on a mono Q column at pH 6.5 and 7.0, the CBD was either not recovered fully or none of the SUMO tag nor the CBD were binding to the column. A cation exchange column could have been tested to see if a different outcome was obtained.

Overall, running a Q FF purification at pH 8.0 gave the best balance between recovery of protein and purity out of the different methods that were tested, which justifies why this method of purification was adopted.

Is the SUMO tag affecting the binding of the CBD to gelatin?

The presence of the his-tagged SUMO also questions whether it affects the function of the CBD. Even though it is hard to answer this question without having a control CBD free of any contaminant, several clues indicate that the function of the CBD is probably not altered by the presence of the cleaved his-tagged SUMO.

First, the lane corresponding to the bound proteins on the gelatine sepharose resin (Figure 3-18) only contains the CBD, indicating that the his-tagged SUMO might not bind to this substrate. This would also mean that the his-tagged SUMO is probably washed away during the first washes in the binding assay. This assumption is further corroborated by the fact that the Kd obtained for the CBD on type II gelatin is in line with previously published papers (20.4 nM for our construct, 50 nM for a refolded His_6-CBD).

The successful expression and purification of a functional CBD was a prerequisite for testing our hypothesis that MSCs coated with this protein would adhere to type II gelatin. It was

then decided to study in finer details the mechanism of binding of the CBD to this substrate, in order to design mutant proteins with higher binding affinities.

4. Characterisation of the binding of the CBD and its modules to gelatin

The first chapter described the successful expression and purification of a fully functional CBD with a dissociation constant for type II gelatin of 20.4 nM. The next goal was to design mutant proteins with increased affinity for this substrate. It was, therefore, necessary to first characterize in finer details the mechanism of binding of the CBD and its individual modules to gelatin. Previously published results on the key residues involved in the binding of the CBD to different gelatin-like substrates will first be carefully reviewed, before applying a similar approach to study the binding of our proteins of interest to both type I and II gelatin.

4.1. Introduction: Binding of the CBD to gelatin

4.1.1. Role of the modules

The resolution of the crystal structure of the full-length MMP-2 has demonstrated that each fibronectin module within the CBD is oriented outward of each other in a 3-pronged fishhook configuration [142]. It was, therefore, hypothesized that this structure would enable each module to act as an independent binding site for gelatin. Building upon these results, it was shown that a recombinant CBD can bind at least two molecules of collagen at a time [141], and three collagen-like peptides [140].

Further investigations on the individual domains of the CBD have shown that each of them is indeed able to bind to immobilized gelatin sepharose (*i.e* type I gelatin), with different affinities [144]. The first module was shown to have the lowest affinity for this substrate, an observation which was confirmed by the fact that the protein made of the two modules 2+3 had a higher affinity than those made of the modules 1+2 and 1+3 [144]. In these experiments, module 2 was also identified as the most potent module [144]. These observations were confirmed by Xu et al., who measured the dissociation constants of each module for a type I collagen-like peptide by NMR [140]. Module 1 was again identified as the weakest module,

with a Kd of $6.0.10^{-4}$ M, while module 2 had a slightly lower Kd than module 3 ($2.8.10^{-4}$ M against $3.4.10^{-4}$ M for modules 2 and 3, respectively) [140]. The prevalence of modules 2 and 3 was further confirmed by some mutational work, where alanine mutations in either module 2 or 3 affected the binding of the CBD to type I gelatin more than mutations in module 1 [140]. However, other authors, through similar alanine mutagenesis work, have attributed most of the binding properties and activity of the CBD towards type I gelatin to module 3 [149].

The full-length CBD was also shown to have a higher affinity for gelatin sepharose than the single modules or any combination of two modules [144], suggesting a cooperative mechanism. Interestingly, the opposite was observed for the CBD of MMP-9, where the second module was shown to provide most of the type I gelatin binding properties of the protein. Its association with the first and third module was shown to give a trimodular CBD with a lower affinity than the module itself [161].

4.1.2. Key residues

Several authors have described the key residues forming the binding pocket of each module, using different gelatin-like peptides. A generic collagen-like peptide, namely PPG6 ((Pro-Pro-Gly)$_6$) was used by a group of authors, and its binding to the individual modules as well as to the full-length CBD was studied by NMR [158–160,180]. These data were compared to those obtained with a different peptide, p33-42, whose sequence, PIIKFPGDVA, corresponds to a segment of the propeptide of MMP-2 [159, 160].

Another group used a type I gelatin peptide with a high degree of sequence similarity with the residues 715 to 721 of the human alpha I chain of type I collagen. The peptide, named P713 has the following sequence: CGA(HYP)-GA(HYP)GSQGA, where HYP refers to hydroxyproline [140]. It was identified in a one peptide-one bead screening method and was shown to interact specifically with the CBD, as well as to abrogate the binding of both the

CBD and MMP-2 to gelatin in a concentration-dependent manner [181]. Following up on these early observations, authors also measured the chemical shifts caused by the binding of this peptide to the CBD of MMP-2 by NMR.

The NMR results from these different groups have been pooled, and residues which were identified in one or several of these studies and which belong to the binding pocket of each module were mapped onto the protein structure of the CBD (PDB Accession Number 1CK7). For clarity, each module will be described and discussed individually in the following paragraphs (Figures 4-1 to 4-3).

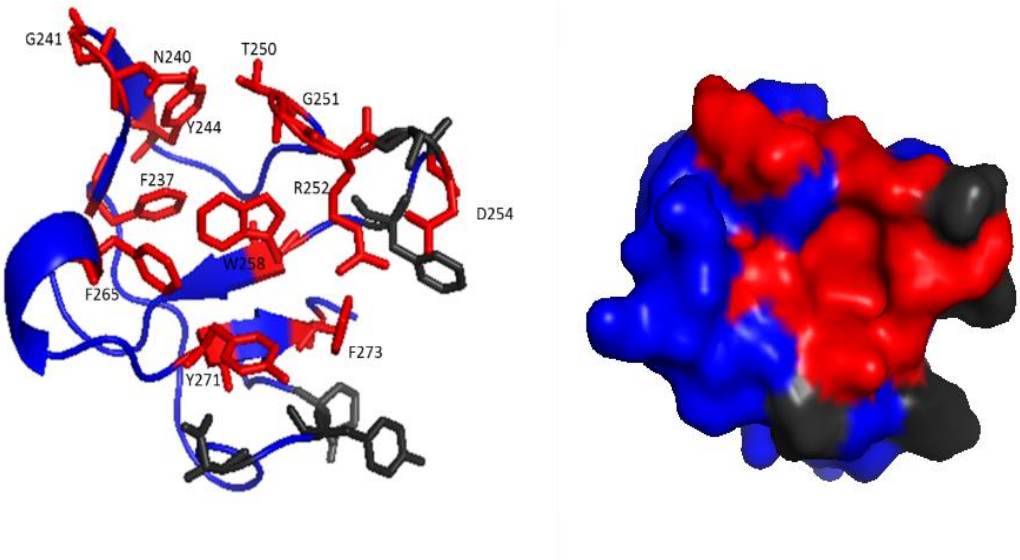


Figure 4-1: Three-dimensional surface representation of module 1 of the CBD of MMP-2 showing the position of the residues that underwent shift changes in the presence of PPG6, P713 or p33-42 and that belonged to the binding pocket. Residues in red form the hydrophobic binding pocket, while those in grey are further away from the binding site. The structure shown is from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7).

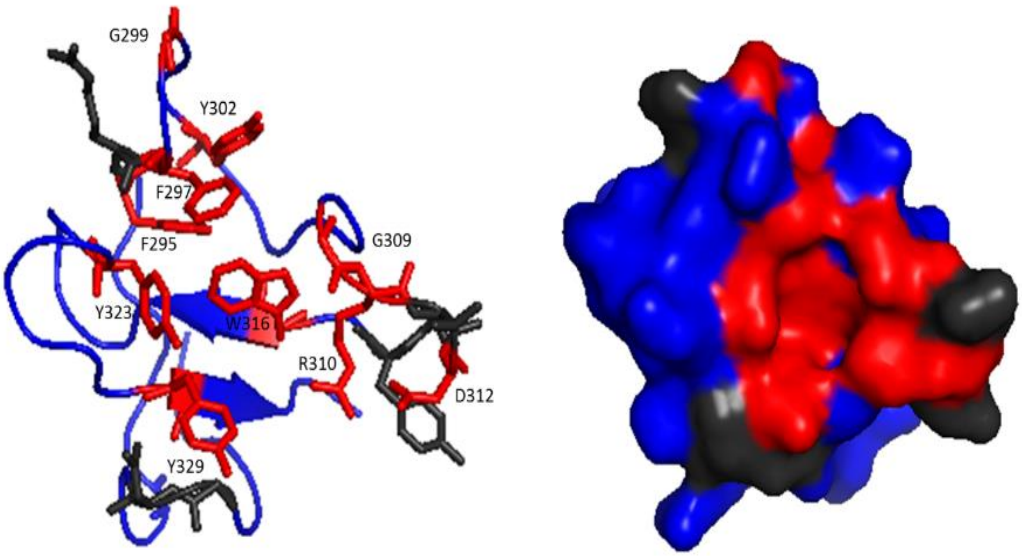


Figure 4-2: Three-dimensional surface representation of module 2 of the CBD of MMP-2 showing the position of the residues that underwent shift changes in the presence of PPG6, P713 or p33-42 and that belonged to the binding pocket. Residues in red form the hydrophobic binding pocket, while those in grey are further away from the binding site. The structure shown is from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7).

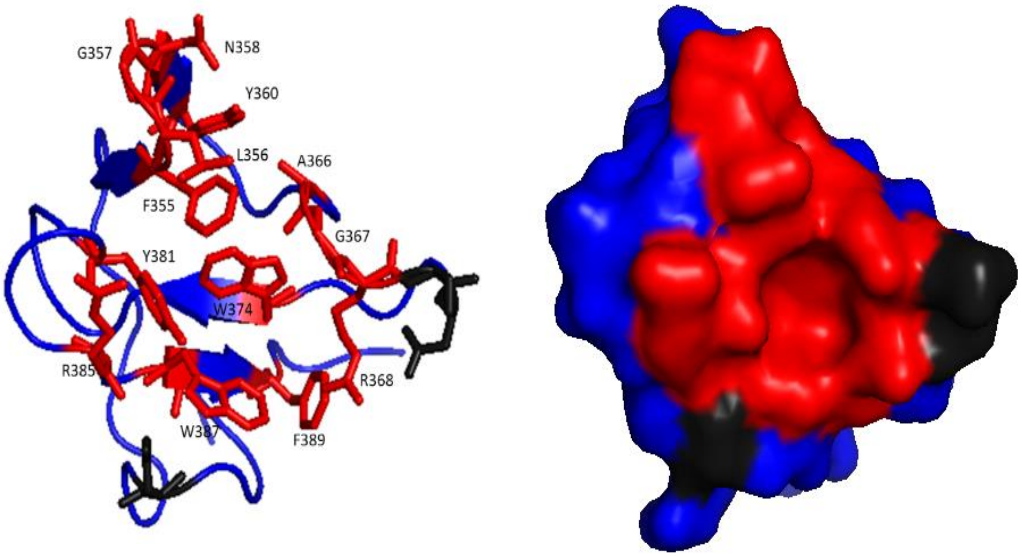


Figure 4-3: Three-dimensional surface representation of module 3 of the CBD of MMP-2 showing the position of the residues that underwent shift changes in the presence of PPG6, P713 or p33-42 and that belonged to the binding pocket. Residues in red form the hydrophobic binding pocket, while those in grey are further away from the binding site. The structure shown is from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7).

Overall, these studies enabled the identification of a cluster of aromatic residues forming the binding site in each module. This observation also confirms the prevalence of hydrophobic interactions for the binding of the proteins to gelatin.

In the study with P713, mutations were performed on residues involved in the binding in each of the three module, and the change in affinity towards native and denatured type I collagen was measured with a plate binding assay [140]. The results are summed up in Table 4-1 below. They correlate with the chemical shifts measurements on the peptide, except for E321, which is located outside of the hydrophobic binding pocket.

Table 4-1: Effect of the mutations of single residues for alanines on the affinity of the CBD for type I collagen and gelatin, as reported by Xu et al. [140]

Mutation	Binding on both native and denatured type I collagen	Module	Kd values increase for type I gelatin
$R_{252}A$	Moderate decrease	Module 1	2-fold
$E_{321}A$	No effect	Module 2	No change
R_{296} A	Moderate decrease		No change
$F_{297}A$, $Y_{302}A$	Strong reduction		5-fold
$Y_{323}A$,$Y_{329}A$			20-fold
$R_{368}A$, $W_{374}A$, $Y_{381}A$	Strong reduction	Module 3	20-fold

4.2. Aim of the chapter

As previously mentioned, all of these NMR data as well as the affinities studies by Bányai et al. were obtained on either gelatin-like peptides or type I gelatin. As our project focuses on the binding of the CBD to type II gelatin, we considered it to be crucial to acquire data on this substrate instead. Which CBD module binds most strongly to a particular target is unclear, highlighting ligand-driven specificity. Indeed previous studies had shown that the

relative affinity of the individual modules of the CBD was ligand-dependent [160,158,180]. Module 1, which was described as the weakest module for type I gelatin, was shown to have the highest affinity for PPG6, while module 3 was the weakest to bind to this peptide, but displayed the highest affinity towards p33-42. Finally, module 2, which was described by many authors as the most potent module for type I gelatin, had the lowest affinity for p33-42 [160,158,180].

Taking all the above results, work in this thesis will focus on understanding the binding of the CDB to both type I and II gelatin, in order to acquire data on more relevant substrates.

It was first decided to study and characterize the binding of the individual modules on their own, as each of them has a binding site for gelatin. The aim was not only to identify the key residues constituting the binding site in each of the modules, but also to determine if any of the modules showed a stronger affinity for type II gelatin compared to the others. The former approach could lead to the design of single-point mutants with increased affinity, while the latter could enable the design of chimeric proteins made of the most potent module.

Finally, we thought it was important to also test the binding of both substrates onto the full-length CBD, in order to determine if the mechanism of binding observed with single modules differed within the trimodular protein.

4.3. Materials and Methods

4.3.1. Cloning

The sequence of the individual modules were as follows:

Module 1: VKYGNADGEYCKFPFLFN GKEYNSCTDTGRSDGFLWCSTTYNFEKDGKYGFCPHEALFTM

Module 2: ALFTMGGNAEGQPCKFPFRFQGTSYDSCTTEGRTDGYRWCGTTEDYDRDKKYGFCPETA

Module 3: MSTVGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGKMWCATTANYDDDRKWGFCPDQG

The synthetic DNA constructs (GeneArt, ThermoFisher) encoding for these sequences were codon optimised and cloned into the pOPINS vector using the ligase independent InFusion cloning methodology. The constructs which were generated carried a His_6-SUMO tag, followed by a cleavage site for the SUMO protease.

4.3.2. Expression in *E.coli* strains

E.coli Shuffle cells were transformed with the sequenced expression plasmids. Conditions of expression such as the temperature and length of the production phase as well as the concentration of inducer were optimised on a small scale.

On a larger scale, protein expression was carried out in 2L flasks. Cells were grown either in LB or minimal media at 37°C, induced for protein expression with IPTG, harvested following the production phase and stored at -80°C.

4.3.3. Purification strategy

Following cell lysis, the released recombinant protein was captured on a nickel column, exploiting the affinity of the hexahistidine tag for this resin. The protein was eluted with a gradient of imidazole, and the tag cleaved with a histidine-tagged SUMO protease (produced in-house). A reverse nickel purification enabled to isolate the cleaved protein from the tag

and protease The polishing step was performed with an Ion Exchange column, and the mode of elution (either linear or step) was optimized for each protein.

4.3.4. Characterisation

The expressed proteins were characterised for their identity using Mass Spectrometry, as well as for their folding. The integrity of the three-dimensional tertiary structure was assessed by 2D NMR, while the number of disulfide bridges was quantified with a DTNB assay. Finally, their activity was tested on a gelatin sepharose resin.

4.3.5. NMR for ligand-binding studies

For NMR studies, proteins were expressed in either ^{15}N-labelled minimal media or, in the case of module 2, in ^{15}N-rich media purchased from Silantes, Germany. The conditions of expression and purification were kept the same as for production in LB. Proteins were then buffer exchanged into 25 mM sodium phosphate, pH 6.5 and 1D and 2D spectra of the free proteins were obtained. Stock solutions of type II gelatin (500 µM to 1 mM) or type I gelatin (485 µM) were obtained and added into the protein solutions at the following protein:gelatin ratios: 40:1, 20:1, 10:1, 5:1, 2:1, 1:1, 1:5 and 1:10. Spectra were acquired at each substrate concentration and the shift of peaks, indicative of binding, were studied. Data were analysed using the softwares TopSpin and CCPN.

4.3.6. Plate binding assay

The affinity of the single modules for both type I and type II gelatin was tested using the plate binding assay which was described in the previous chapter. Briefly, proteins were biotinylated and added over a concentration range onto gelatin-coated plates. Alkaline phosphatase coupled to streptavidin was added to the bound proteins, and its substrate was transformed into a coloured product. The absorbance of the product was used as an indirect indication of the amount of bound protein.

4.4. Results

4.4.1. Expression and purification of individual modules

The sequences of the individual modules were taken from previously published papers [158,180,159]. Cloning and amplification of the recombinant plasmids were performed as described in the Materials and Methods section, and all constructs were confirmed by sequencing.

Shuffle cells were used to screen for optimal expression conditions, which appeared to be an overnight expression phase at 18°C, with an IPTG concentration of 200 µM (not shown). All modules were expressed using the same conditions.

All three modules were successfully expressed in Shuffle cells on a large scale and purified following the same strategy as that used for the full-length CBD. Modules 2 and 3 were eluted from the Q FF Ion Exchange column with a linear gradient with no impurities (Figure 4-4). The first module, however, required a step elution at 12 % of IEC Buffer B (1 M NaCl, 20 mM Tris, pH 8) to remove contaminants. This strategy yielded two different peaks, the first one running at the expected size on the SDS PAGE gel (Figure 4-5). The second one contained a mixture of the same protein and another one with a higher molecular weight at around 15 kDa (Figure 4-5). Although the gel was ran in both denaturing and reducing conditions, the possibility of the second band being a covalently-bound dimer could not be ruled out at this stage. All subsequent work was carried out with the protein obtained from Peak 1 of this chromatogram.

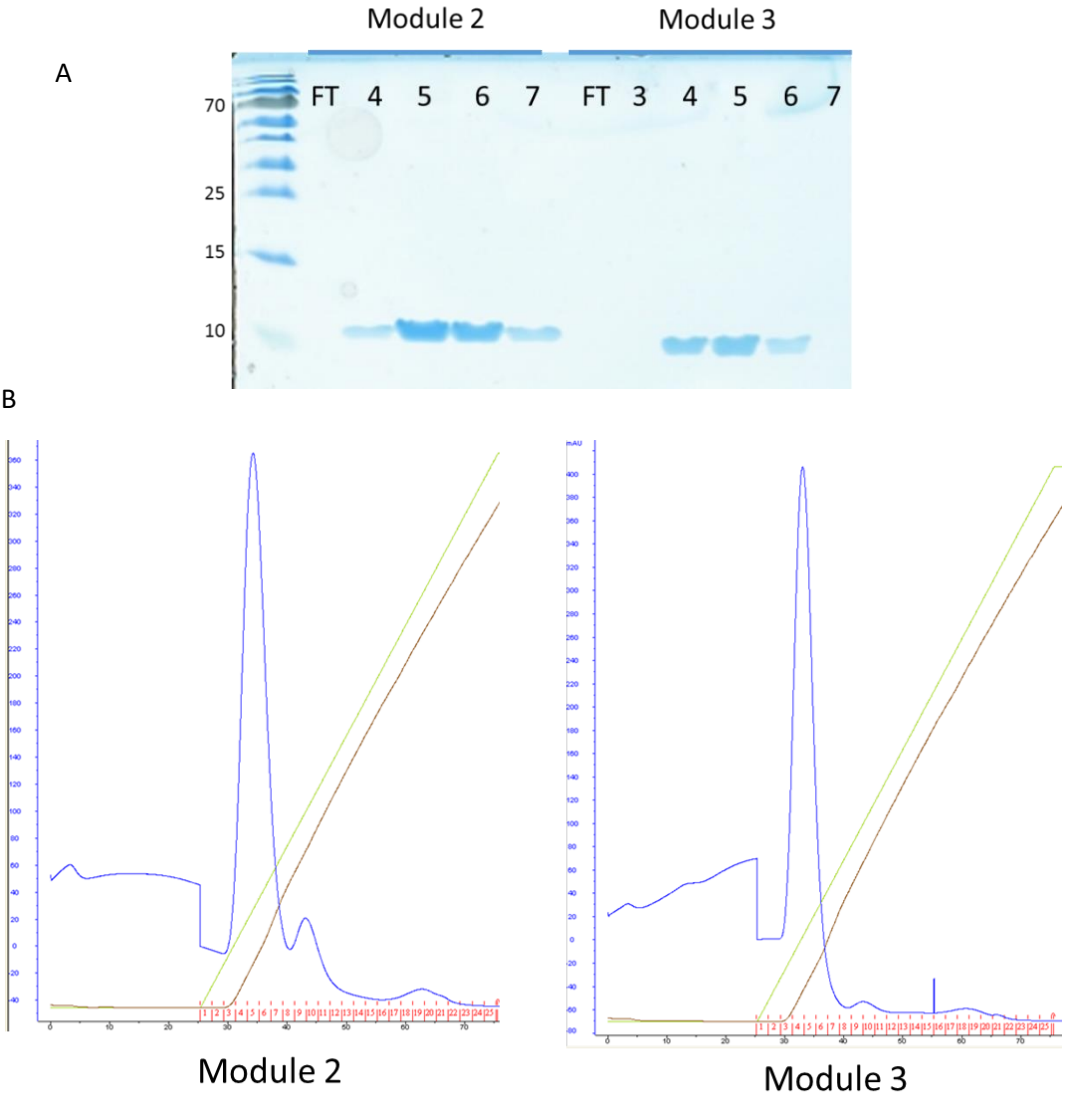
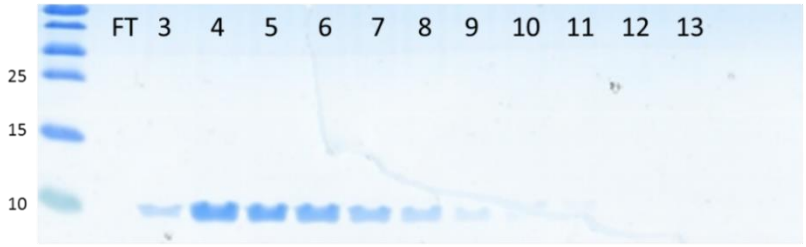
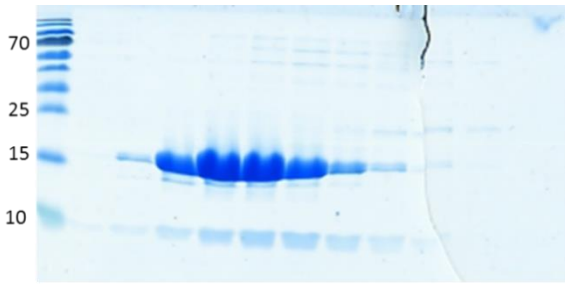


Figure 4-4: SDS PAGE gel showing the expression of the modules 2 and 3 in LB (A) and corresponding chromatograms (B). A - The modules 2 and 3 of MMP-2, originally carrying a His_6-SUMO tag, were expressed in Shuffle *E.coli* cells and purified by nickel column and reverse nickel column purification, as previously described. The protein samples were then desalted and loaded onto a 5 mL Q FF column, and eluted with a linear gradient of sodium chloride. FT: Flowthrough; 4-7: Ion Exchange fractions. B - Chromatogram corresponding to the Ion Exchange purification of the Module 2 (left) and Module 3 (right) proteins. Fractions 4 to 7 were loaded on the gel. Blue line: UV trace; Green line: Concentration of IEC B buffer; Brown line: conductivity.

A



Peak 1



Peak 2

B

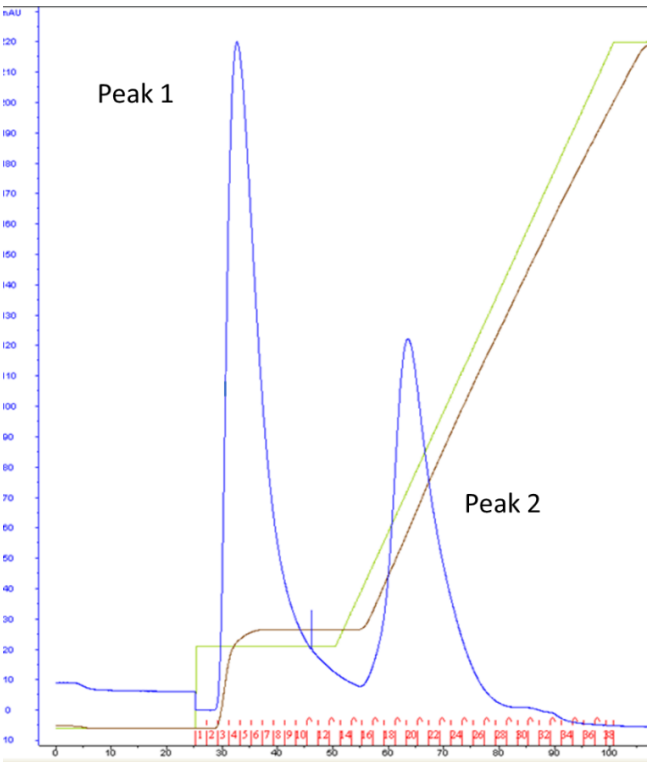


Figure 4-5: SDS PAGE gels showing the expression of module 1 in LB (A) and corresponding chromatogram (B). Module 1 of MMP-2, originally carrying a His_6-SUMO tag, was expressed in Shuffle *E.coli* cells and purified by nickel column and reverse nickel column purification, as previously described. The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a step gradient of sodium chloride. This step elution led to two peaks, each with a different protein content. FT: Flowthrough; 3-13: Ion Exchange fractions. B - Chromatogram corresponding to the Ion Exchange purification of Module 1. Fractions 3 to 13 were loaded on the first gel, and fractions 16 to 26 on the second. Blue line: UV trace; Green line: Concentration of IEC B buffer; Brown line: conductivity.

4.4.2. Characterisation

4.4.2.1. Identity and structure

All modules were confirmed by Mass Spectrometry. Protein samples were digested overnight with trypsin and ionised on a MALDI-TOF instrument. The peptides which were obtained had masses corresponding to those expected for the predicted cleavage for each module (not shown).

The number of disulfide bridges was also quantified with a DTNB assay, which indicated the presence of two disulfide bridges in all the modules. This result definitely ruled out the possibility for the first module to be forming a covalently-bound dimer resulting in the additional band and peak observed on Figure 4-5.

4.4.2.2. NMR on individual modules

^{1}H-^{15}N 2D HSQC spectra of all 3 free modules were obtained, and their folding verified as the peaks of the backbone amide protons are well dispersed with all peaks showing similar intensities (Figures 4-6 to 4-8). Module 1 showed a tendency to aggregate, especially for peaks in the central region of the spectrum (Figure 4-6), a phenomenon which was already observed in the original spectrum of the CBD (Figure 3-16). This might indicate that the module is naturally less stable, or less folded, than the other two, which do not show any region of overlapping peaks (Figures 4-7 and 4-8 for modules 2 and 3, respectively). Spectra were assigned using previously published results [159,158,180]. It was decided to leave overlapping peaks or peaks with more than one possible assignment unassigned.

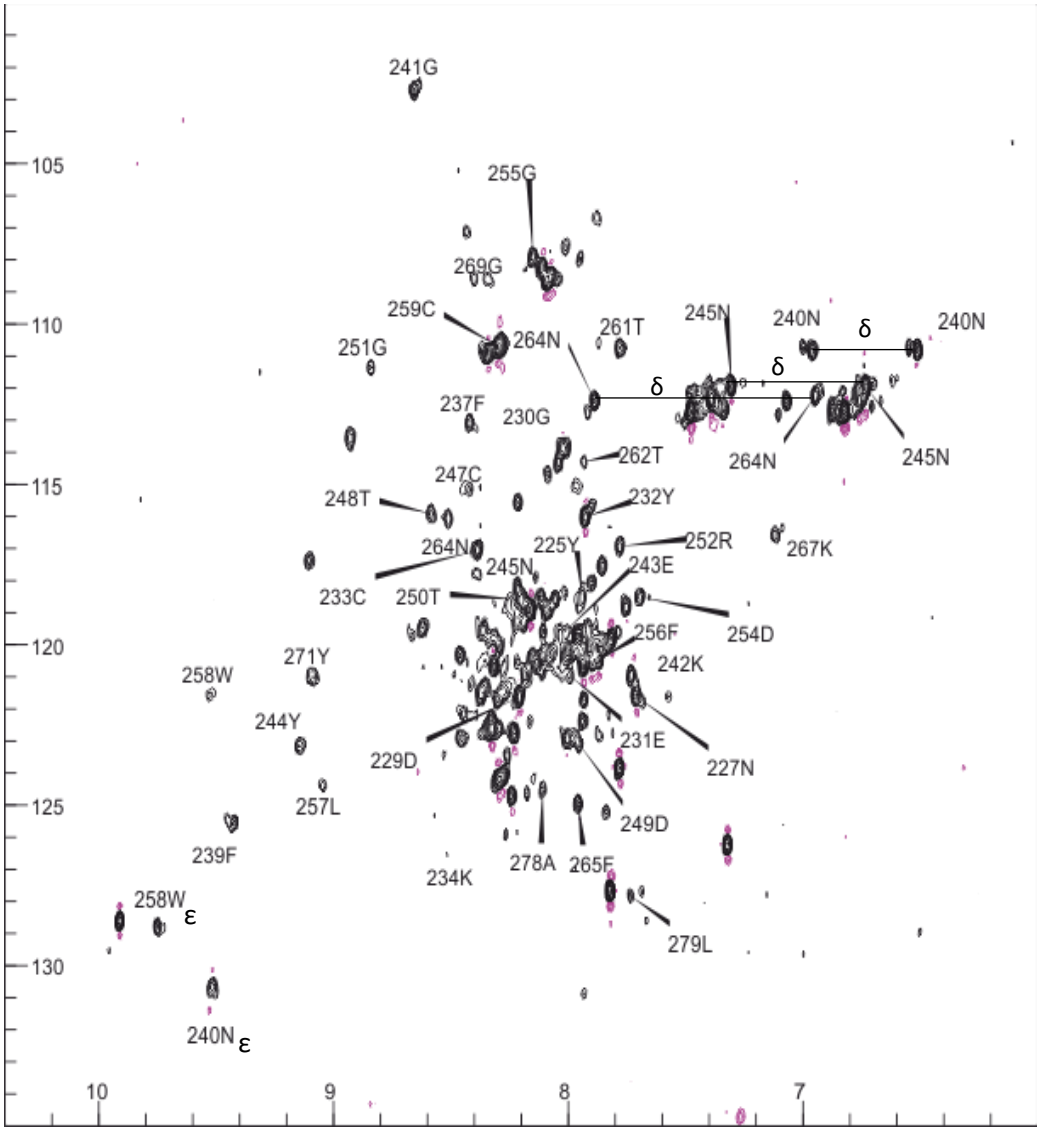


Figure 4-6: ^{1}H-^{15}N HSQC spectrum of the first module of the CBD at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O, 45 µM. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the first module. The peaks were assigned to their corresponding amino acid using previously published results and are labelled with their one-letter amino acid code and residue number.

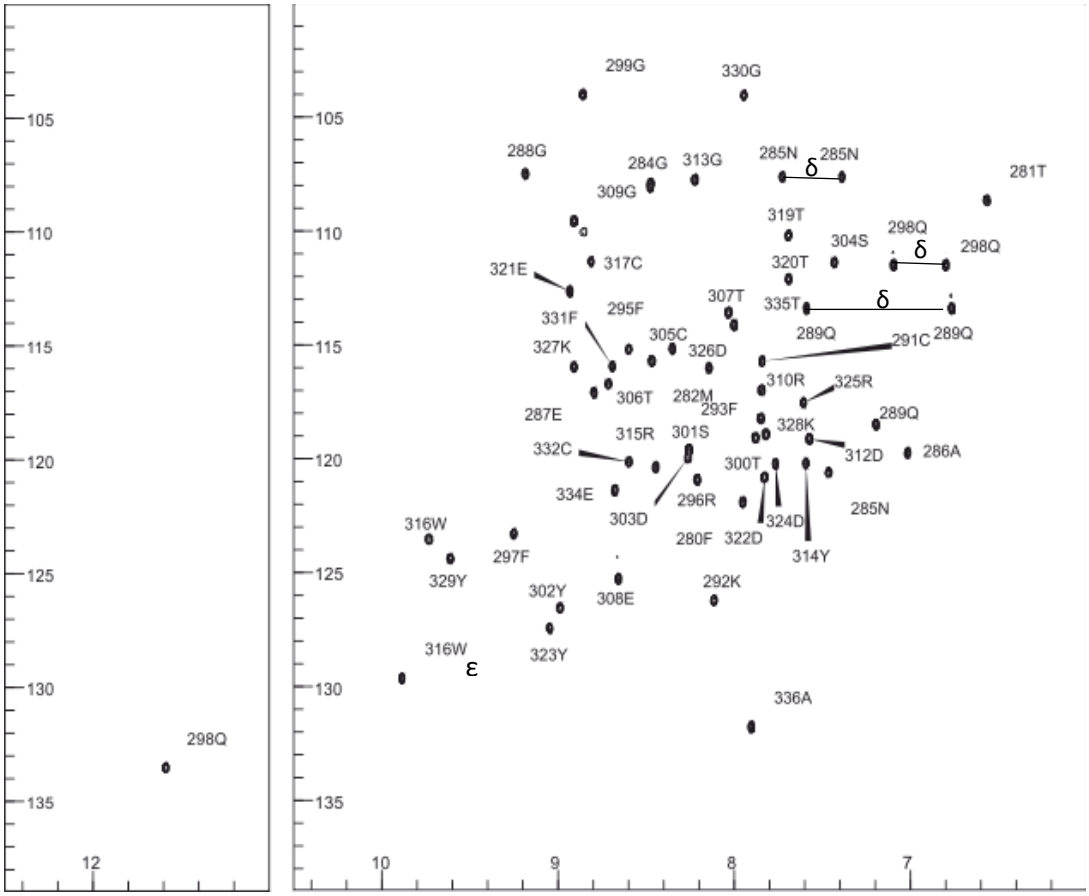


Figure 4-7: ^{1}H-^{15}N HSQC spectrum of the second module of the CBD at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O, 20 µM. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the second module. The peaks were assigned to their corresponding amino acid using previously published results and are labelled with their one-letter amino acid code and residue number.

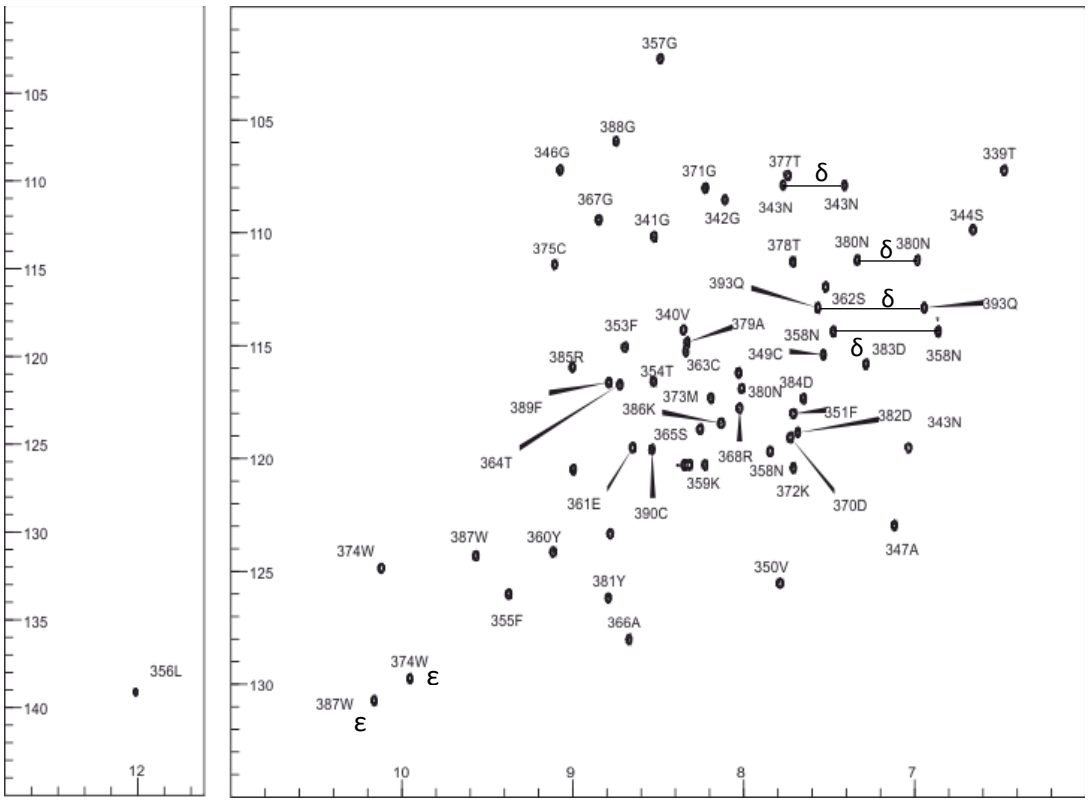


Figure 4-8: ^{1}H-^{15}N HSQC spectrum of the third module of the CBD at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O, 30 µM. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the third module. The peaks were assigned to their corresponding amino acid using previously published results and are labelled with their one-letter amino acid code and residue number.

4.4.2.3. Activity: gelatin sepharose

The binding of the modules to the gelatin sepharose resin was studied; both modules 2 and 3 were able to bind (Figure 4-9). However, module 1 was mostly recovered in the unbound fraction (FT lane, Module 1, Figure 4-9). This could be due to the very weak binding of the module to this substrate, which was previously described in various papers [140,144].

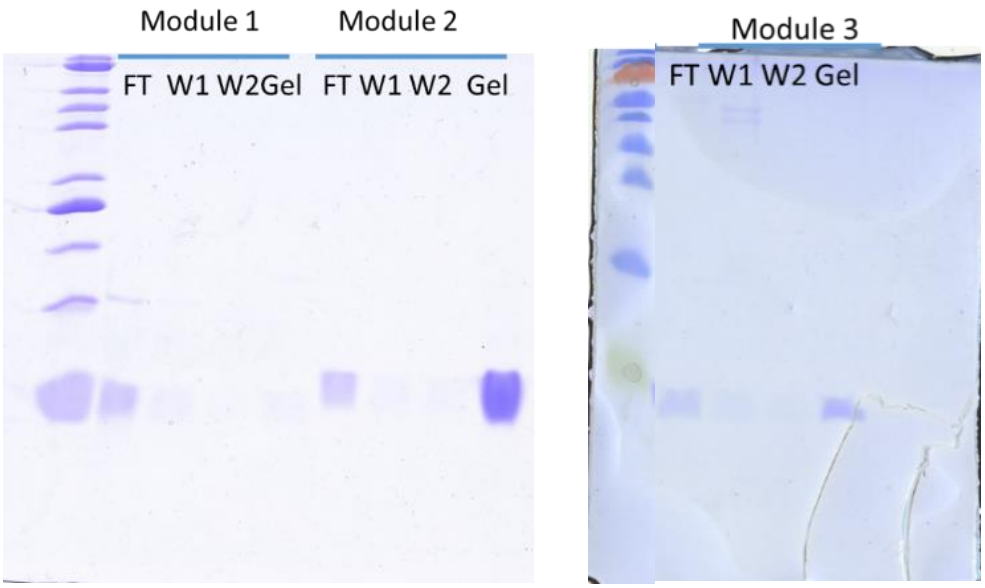


Figure 4-9: SDS PAGE showing the binding of the individual modules to gelatin sepharose. The individual modules from the Ion Exchange were tested for their ability to bind gelatin using a gelatin sepharose resin. 100 µL of resin was spun down, washed with ethanol and equilibrated before applying the samples. The mixture was spun again and the flow through removed (FT) before washing two times with the binding buffer (W: Washes). The bound proteins were mixed with 2X SDS loading buffer and loaded on the gel (Gel: gelatin-bound).

4.4.3. Characterisation of the binding of the single modules

4.4.3.1. NMR ligand studies

The binding of the single modules to both type I and type II gelatin was then tested in order to identify which residues form the binding pocket for both substrates.

4.4.3.1.1. Binding of module 1

Chemical shifts could be observed for module 1 when comparing the free protein to that mixed at a 2:1 (protein:gelatin) ratio, for both types of gelatin (Figures 4-10 and 4-11). Higher concentrations in gelatin gave more significant shifts with type II gelatin, but led to line-broadening for many peaks with type I gelatin. For clarity and comparison purposes, it was therefore decided to focus on the 2:1 ratio.

The spectra for the complexes formed using both substrates look very similar, with many of the same residues being affected identically (Figures 4-10 and 4-11). Residue N245 showed broadening in both spectra while residue K234 showed broadening following addition of type I gelatin only, although it was also highly affected by the binding to type II gelatin (Figures 4-10 and 4-11).

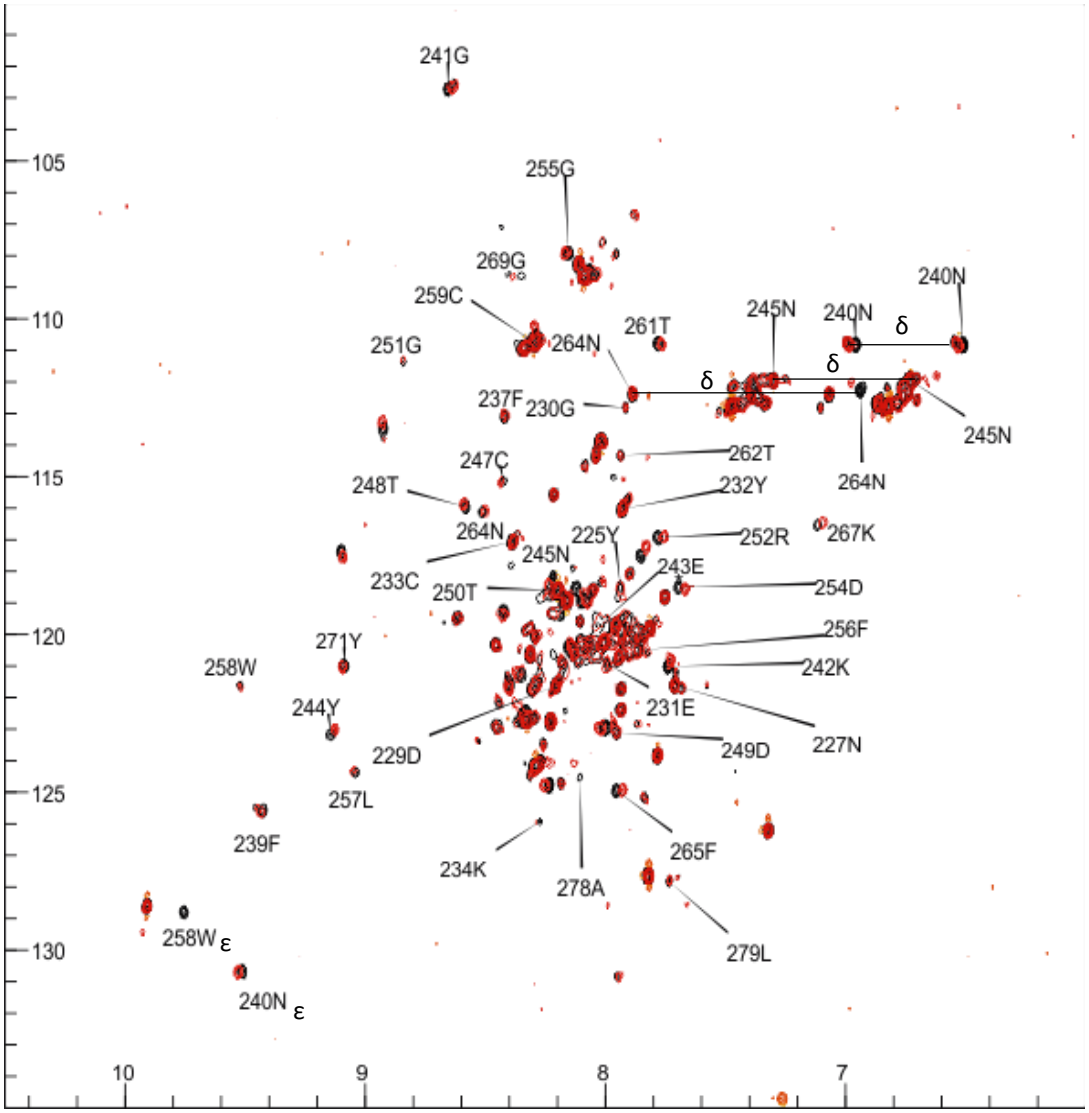


Figure 4-10: ^{1}H-^{15}N HSQC spectra of module 1 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free module 1 (black) and of module 1 in the presence of type II gelatin (red) are superimposed. A 2:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument Each peak on the spectrum corresponds to an amide proton in the first module. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

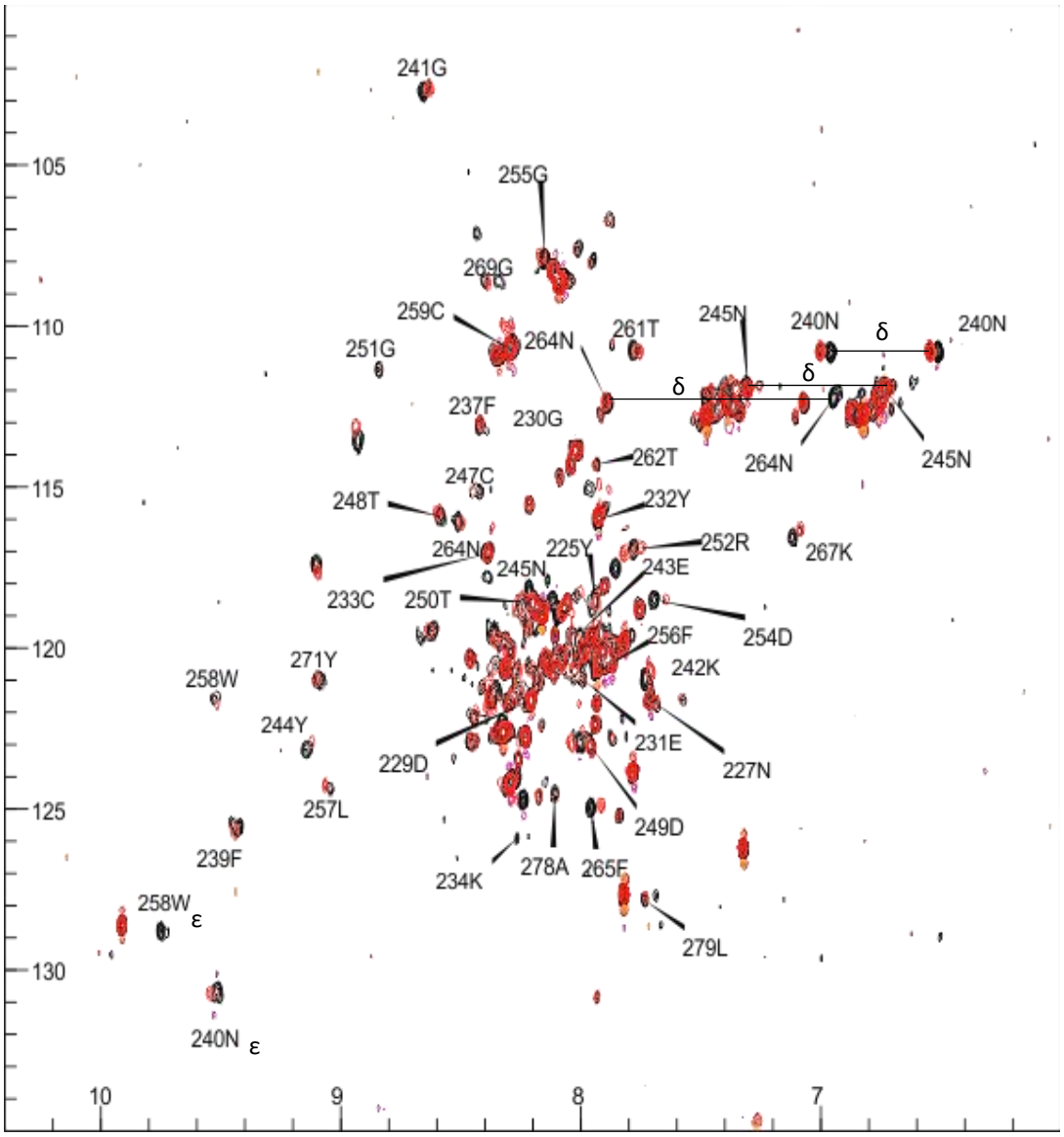


Figure 4-11: ^{1}H-^{15}N HSQC spectra of module 1 before and after addition of type I gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free module 1 (black) and of module 1 in the presence of type I gelatin (red) are superimposed. A 2:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the first module. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

In order to directly compare the results, the shift distance observed between the free and bound spectra was plotted for each residue, except those with broadened peaks, for both type I and type II gelatin (Figures 4-12 and 4-13). A threshold value had to be set for identifying the key residues involved in the binding. R252 was previously shown to be involved in the binding of module 1 on its own or within the CBD to all the different gelatin-like peptide substrates that were described in the introduction of this chapter [140,158,160]. It is therefore very likely that this residue is involved in the binding to type I gelatin. In our analysis, its chemical shift was 0.01935, which was the 15th biggest. It was therefore decided to set the threshold for type I gelatin at 0.019, in order to include this residue as a key one for the binding of the CBD to type I gelatin. The threshold for type II gelatin was then adjusted accordingly so that the ratio (maximum shift distance:threshold) was constant between both analyses, giving a threshold value of 0.00825.

From these figures (4-12 and 4-13), we could observe that many of the same residues were undergoing the most important shifts for both substrates: K234 (broadened for type I gelatin), F239, N240, G241, K242, E243, Y244, C247, T248, R252, D254, L257, W258, T261, F265 and K267. However, a few residues reached the set threshold for only type II gelatin: F237, G251, G255, N264 and G269. This could indicate a slightly different mode of binding for module 1 on type I and type II gelatin.

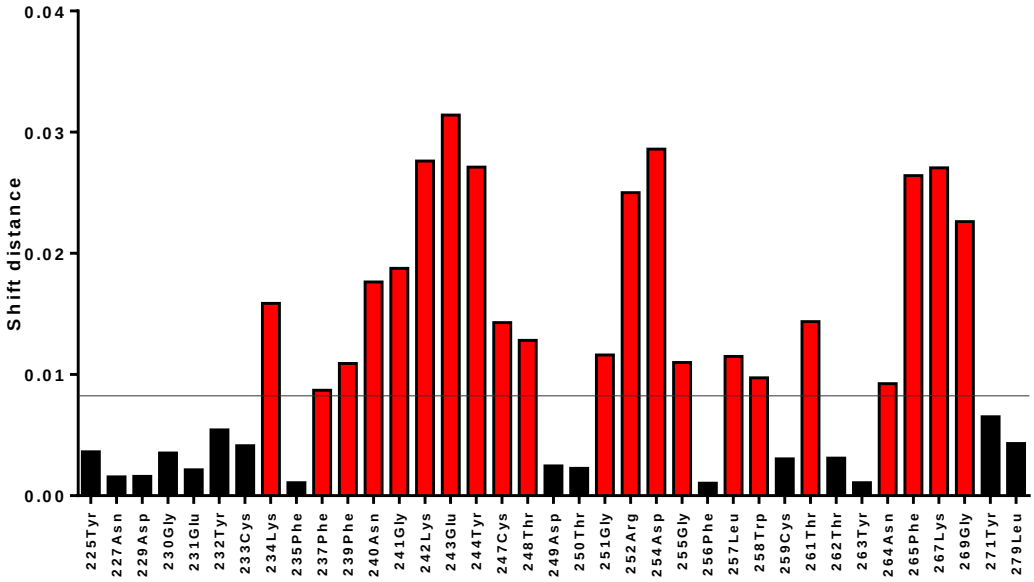


Figure 4-12: Magnitude of the chemical shift changes for the residues of the first module in the presence of type II gelatin at a protein:gelatin ratio of 2:1. Chemical shifts distances were calculated in the CCPN analysis software. A threshold for significant chemical shift distance was set at 0.00825.

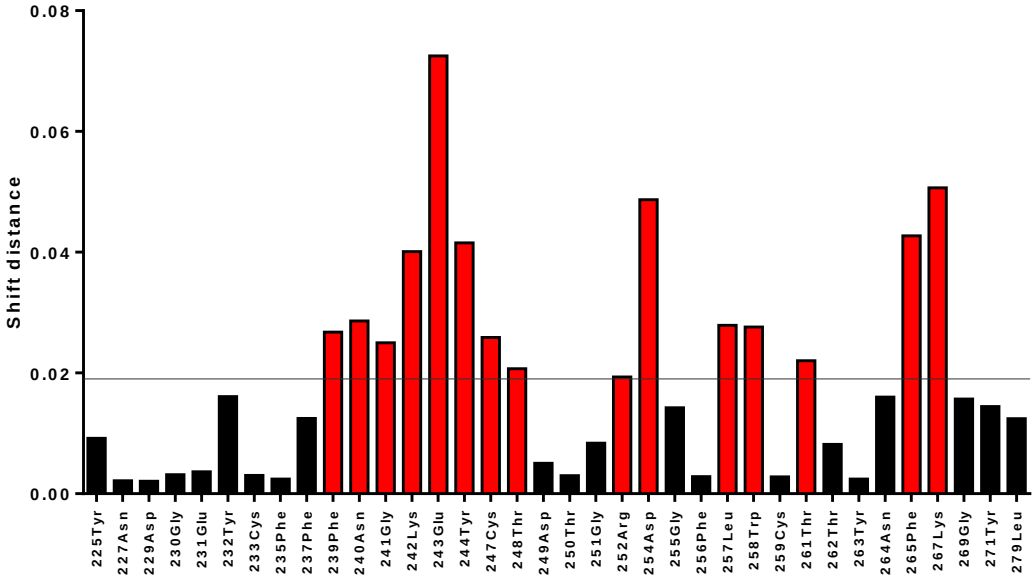


Figure 4-13: Magnitude of the chemical shift changes for the residues of the first module in the presence of type I gelatin at a protein:gelatin ratio of 2:1. Chemical shifts distances were calculated using the CCPN analysis software. A threshold for significant chemical shift distance was set at 0.019.

The residues identified in this experiment were then mapped onto the structure of the CBD in Pymol, in red for those common to both substrates, and in brown for those involved in the binding to type II gelatin only (Figure 4-14). Most of them were located on or around the binding pocket of the module, confirming the validity of the results.

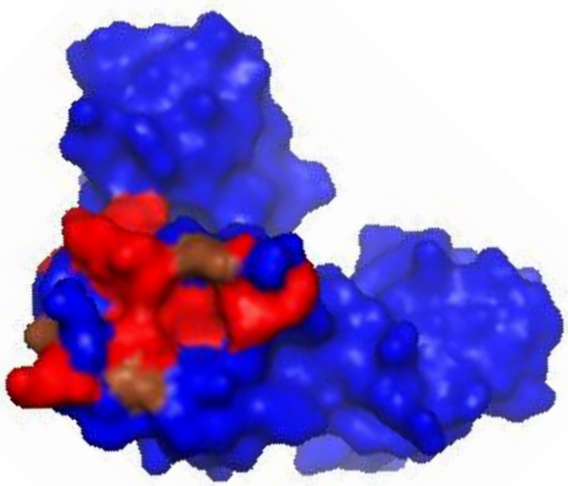


Figure 4-14: Three-dimensional surface representation of the CBD of MMP-2 showing the position of the residues in module 1 that underwent shift changes in the presence of type I and type II gelatin at a 2:1 protein:gelatin ratio. Residues undergoing changes with both substrates are indicated in red, while residues affected by the binding to type II gelatin only are indicated in brown. The structure shown is from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7).

4.4.3.1.2. Binding of module 2

Chemical shifts could be observed for module 2 when comparing the free protein to that mixed at a 10:1 (protein:gelatin) ratio, for both type I and II gelatin (Figures 4-15 and 4-16). More significant shifts were observed at higher gelatin concentrations, with many important residues showing line-broadening as a consequence of their involvement in the binding (not shown). For clarity and comparison purposes, it was therefore decided to focus on the 10:1 ratio.

The spectra for both substrates looked very similar, with many of the same residues being affected identically. Residues Y323, G309, G313 and T319 showed broadening in both spectra, while R310 and G299 showed broadening following type I and II gelatin binding respectively, although they were also highly affected by the binding of the other substrate (Figures 4-15 and 4-16).

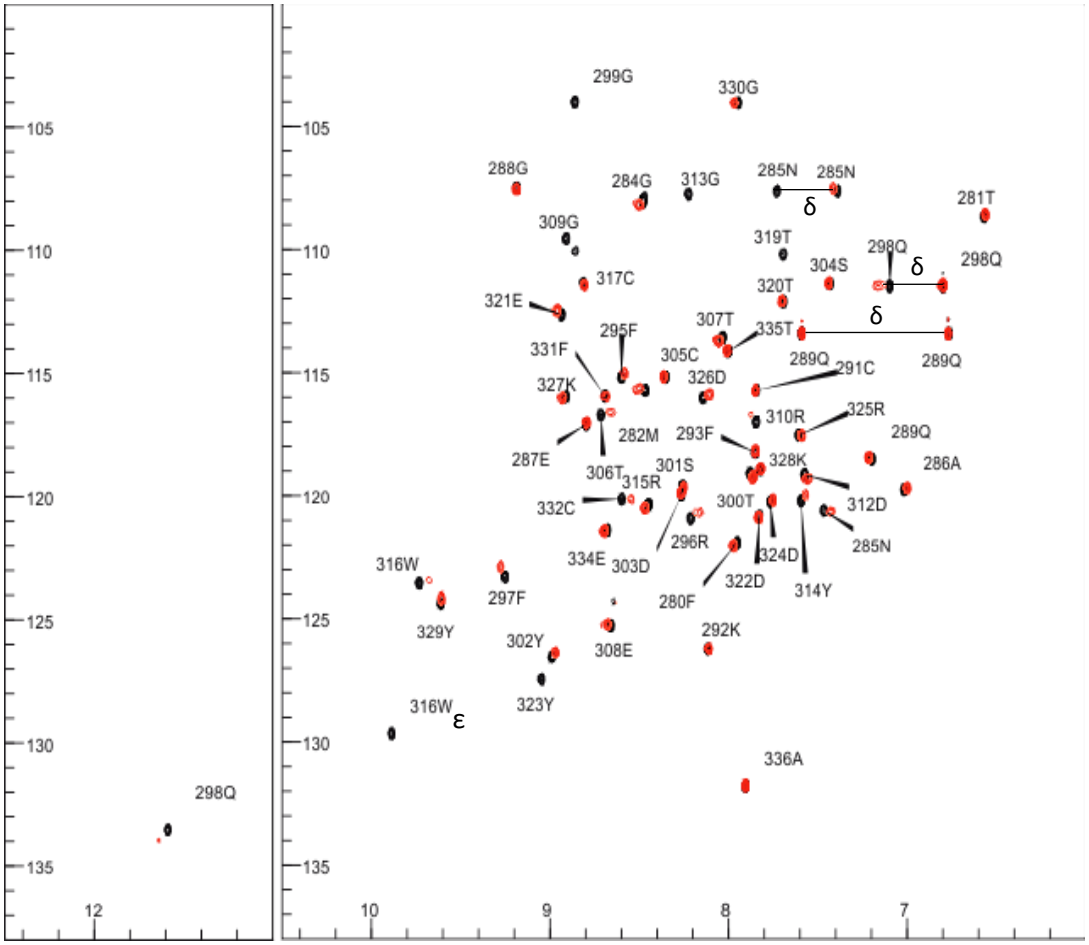


Figure 4-15: ^{1}H-^{15}N HSQC spectra of module 2 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free module 2 (black) and of module 2 in the presence of type II gelatin (red) are superimposed. A 10:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the second module. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

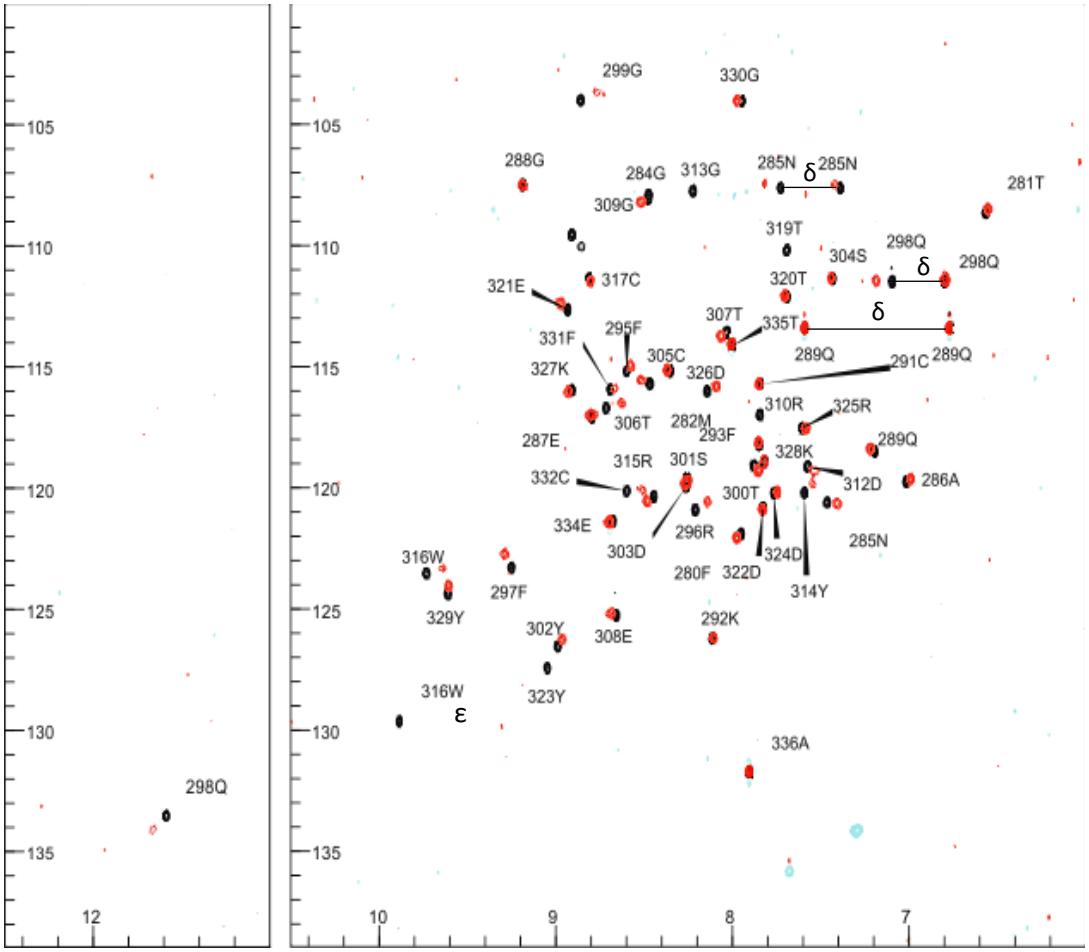


Figure 4-16: ^{1}H-^{15}N HSQC spectra of module 2 before and after addition of type I gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free module 2 (black) and of module 2 in the presence of type I gelatin (red) are superimposed. A 10:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the second module. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

In order to directly compare the results, the shift distance observed between the free and bound spectra were plotted for each residue, except those with broadened peaks (Figures 4-17 and 4-18). A threshold value had to be set for identifying the key residues involved in the binding. Y302 was previously shown to be involved in the binding of module 2 on its own and within the CBD to all but one of the different gelatin-like peptide substrates that were described in the introduction of this chapter [140,180,160]. It is therefore very likely that this residue is involved in the binding to type I gelatin. In our analysis, its chemical shift value was 0.04782, which was the 15th highest. It was therefore decided to set the threshold for type I gelatin at 0.047, in order to include this residue as a key one for the binding of the CBD to type I gelatin. The threshold for type II gelatin was adjusted accordingly so that the ratio (maximum shift distance:threshold) was constant between both analyses, giving a threshold value of 0.03268.

From these figures (Figures 4-17 and 4-18), we could observe that the exact same residues were undergoing the most important shifts for both types of gelatin: M282, G284, N285, R296, F297, Q298, G299 (broadened for type II gelatin binding), T306, R310 (broadened for type I gelatin), Y314, W316, E321, D326, Y329 and C332. Only residues R315 and Y302 reached the set threshold for only type I gelatin, but they were also affected by the binding of the other substrate. Overall, it was therefore shown that module 2 binds to type I and type II gelatin with the same mechanism, and through the same residues.

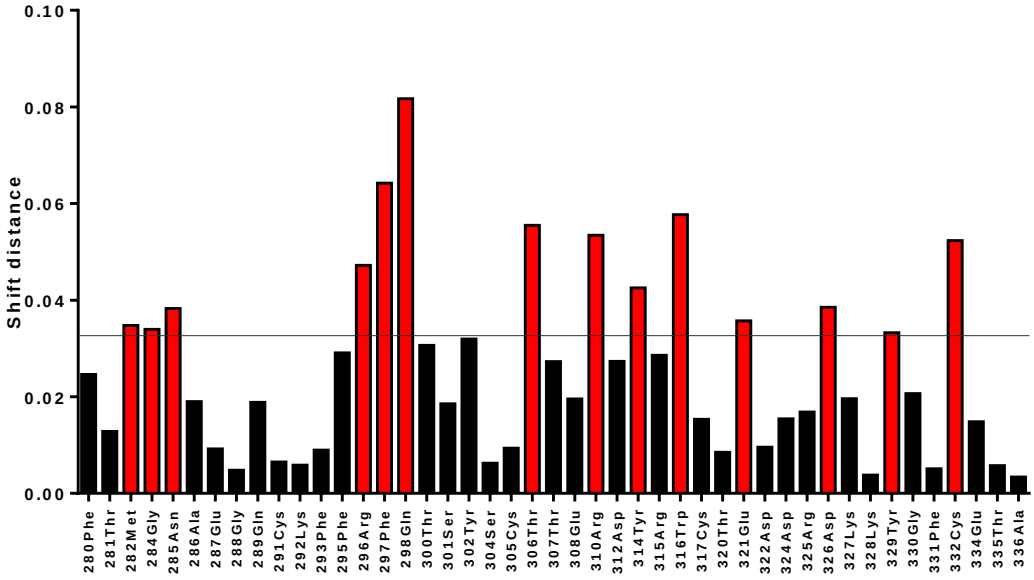


Figure 4-17: Magnitude of the chemical shift changes for the residues of the second module in the presence of type II gelatin at a protein:gelatin ratio of 10:1. Chemical shifts distances were calculated using the CCPN analysis software. A threshold for significant chemical shift distance was set at 0.03268

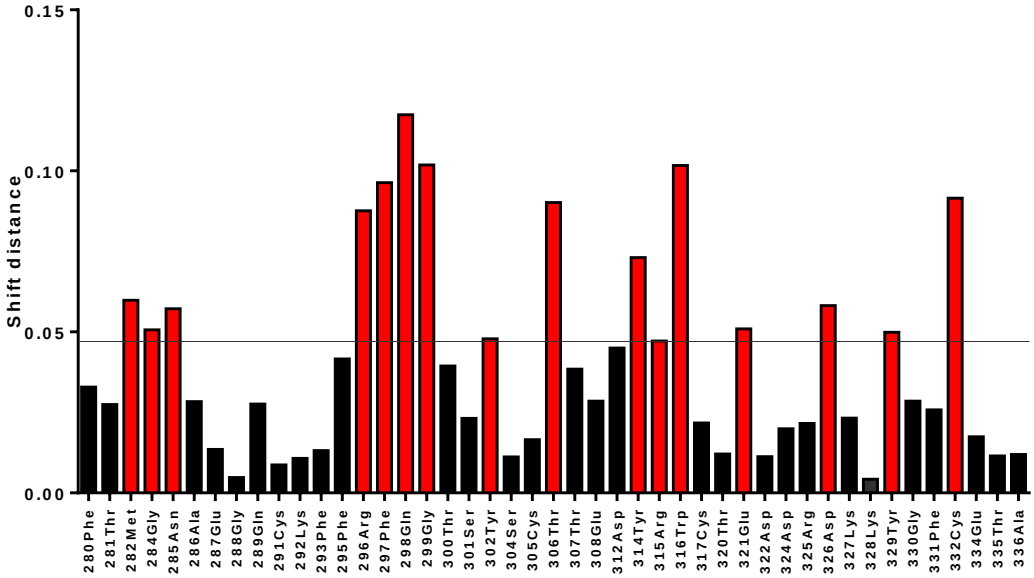


Figure 4-18: Magnitude of the chemical shift changes for the residues of the second module in the presence of type I gelatin at a at a protein:gelatin ratio of 10:1. Chemical shifts distances were calculated in the CCPN analysis software. A threshold for significant chemical shift distance was set at 0.047.

The residues identified in both experiments were then mapped onto the structure of the CBD in Pymol. Most of them were located on or around the binding pocket of the second module, confirming the validity of the results (Figure 4-19).

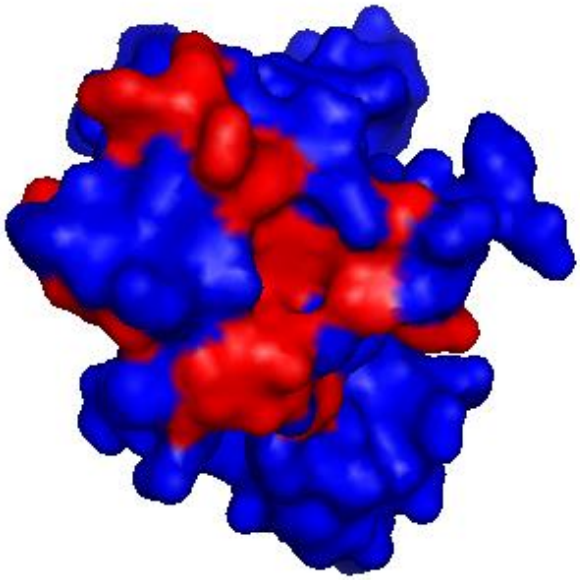


Figure 4-19: Three-dimensional surface representation of the CBD of MMP-2 showing the position of the residues in module 2 that underwent shift changes in the presence of both type I and type II gelatin at a 10:1 protein:gelatin ratio. Residues undergoing changes are indicated in red. The structure shown is from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7).

4.4.3.1.3. Binding of module 3

Chemical shifts could be observed for module 3 when comparing the free protein to that mixed at a 1:5 (protein:gelatin) ratio, for type II gelatin (Figure 4-20). As the concentration of the type I gelatin stock was lower, only a 1:4 (protein:gelatin) ratio could be reached, and this experiment was used to compare the results (Figure 4-21).

The spectra for both substrates looked very similar, with many of the same residues being affected identically. Residue G367 showed broadening in both spectra, while residues L356 (not shown) and R368 showed broadening following type I gelatin addition only, although they were also highly affected by the binding of type II gelatin (Figures 4-20 and 4-21).

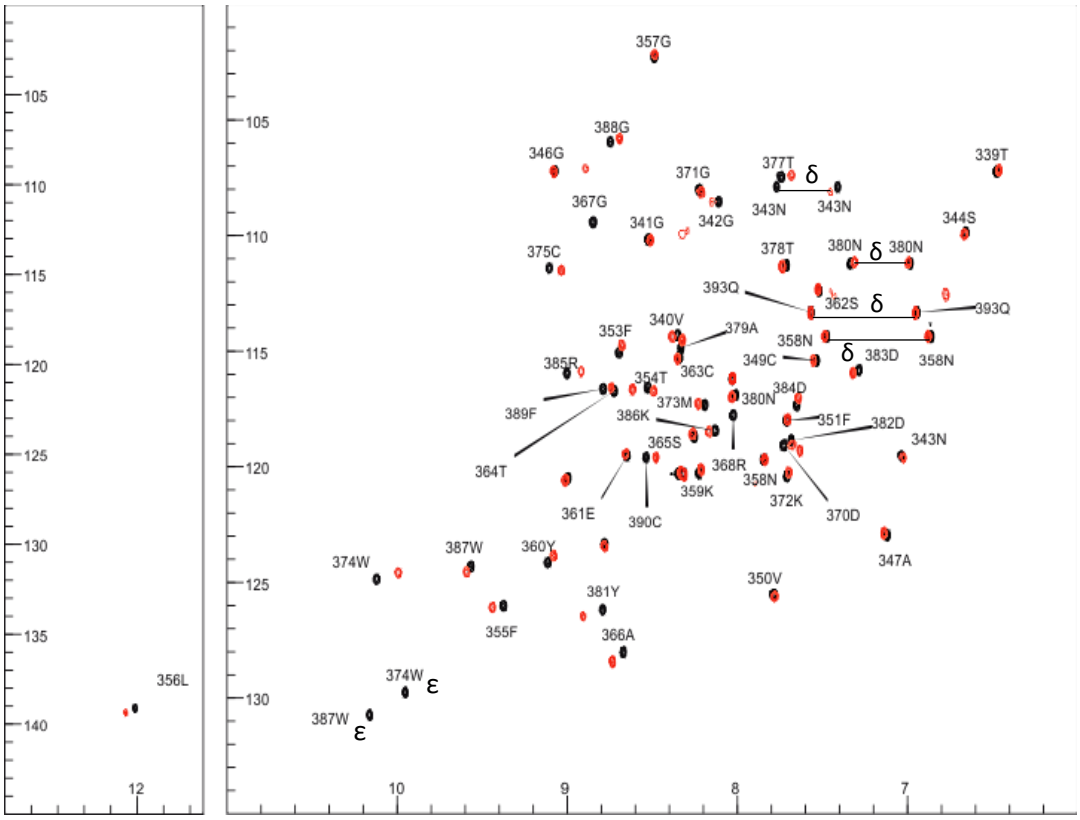


Figure 4-20: ^{1}H-^{15}N HSQC spectra of module 3 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free module 3 (black) and of module 3 in the presence of type II gelatin (red) are superimposed. A 1:5 protein:gelatin molar ratio was used. The spectra were acquired on a 800 Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the third module. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

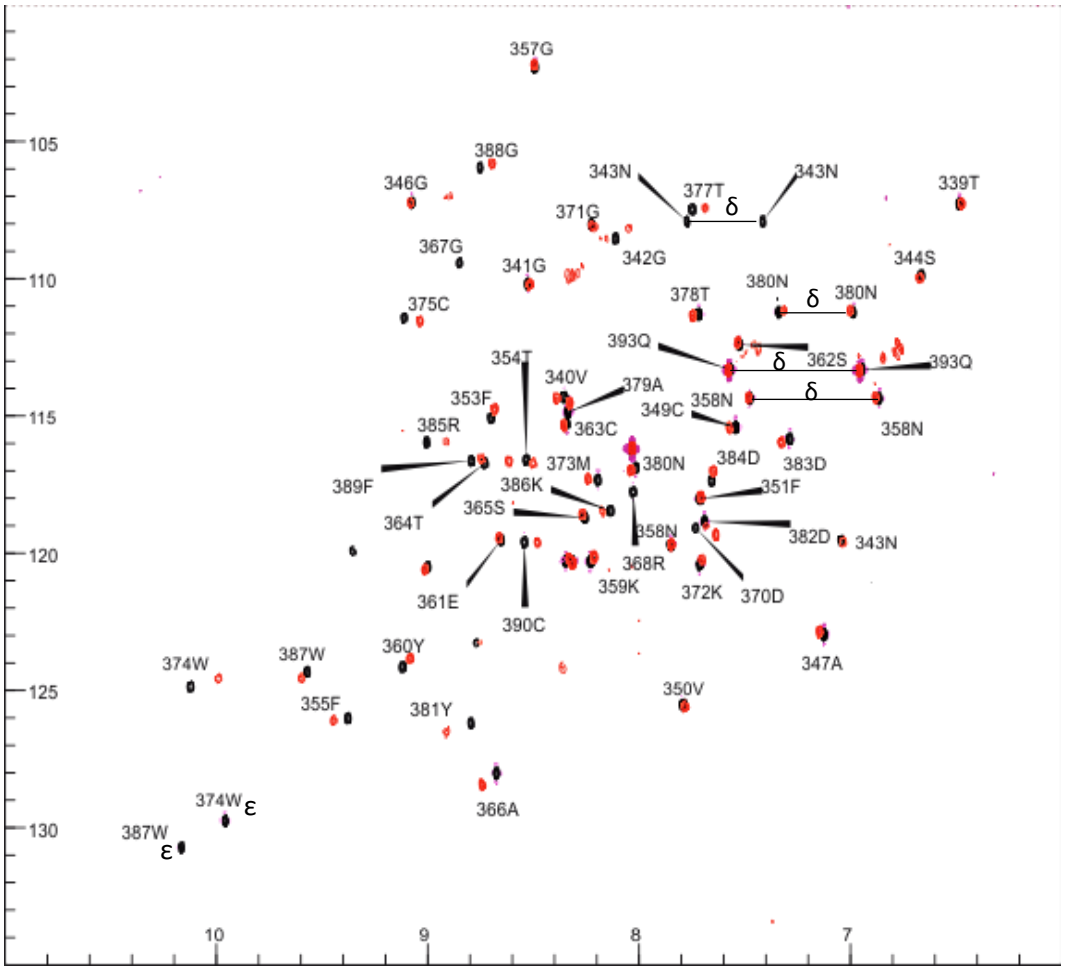


Figure 4-21: ^{1}H-^{15}N HSQC spectra of module 3 before and after addition of type I gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free module 3 (black) and of module 3 in the presence of type I gelatin (red) are superimposed. A 1:4 protein:gelatin molar ratio was used. The spectra were acquired on a 800 Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the third module. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

In order to directly compare the results, the shift distance observed between the free and bound spectra was plotted for each residue, except those with broadened peaks (Figures 4-22 and 4-23).

A threshold value had to be set for identifying the key residues involved in the binding. W387 was previously shown to be involved in the binding of module 3 on its own and within the CBD to one gelatin-like peptide substrate, namely PPG6 [159,160]. In addition, the residue lies within the binding pocket of this module, as outlined in the introduction of this chapter. It is therefore very likely that this residue is involved in the binding to type I gelatin. In our analysis, its chemical shift value was 0.041, which was the 21st highest. It was therefore decided to set the threshold for type I gelatin at 0.040, in order to include this residue as a key one for the binding of the CBD to type I gelatin. The threshold for type II gelatin was adjusted accordingly so that the ratio (maximum shift distance:threshold) was constant between both analyses, giving a threshold of 0.03823.

From these figures (Figures 4-22 and 4-23) we could observe that the exact same residues were undergoing the most important shifts for both type I and type II gelatin: G342, F353, F355, L356 (broadened for type I gelatin), Y360, T364, A366, R368 (broadened for type I gelatin), D370, W374, C375, T377, A379, Y381, D382, D383, D384, R385, W387, G388, F389, C390. Only one residue reached the set threshold for only one of the two substrates: M373 for type I gelatin. However, it was also affected by the binding of the module to the other substrate. Overall, it was therefore shown that module 3 binds to type I and type II gelatin with the same mechanism, and through the same residues.

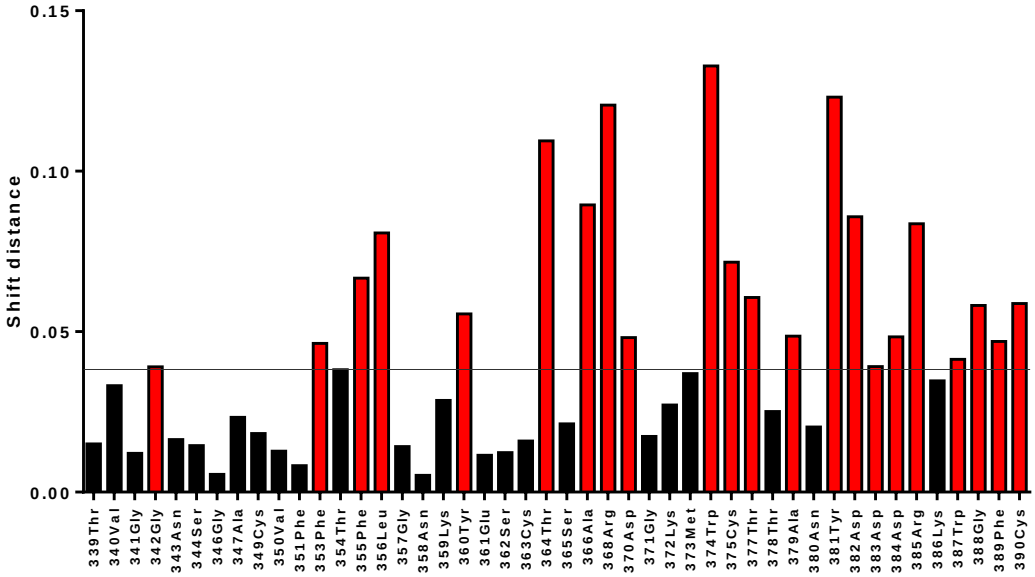


Figure 4-22: Magnitude of the chemical shift changes for the residues of the third module in the presence of type II gelatin at a protein:gelatin ratio of 1:5. Chemical shifts distances were calculated using the CCPN analysis software. A threshold for significant chemical shift distance was set at 0.03823.

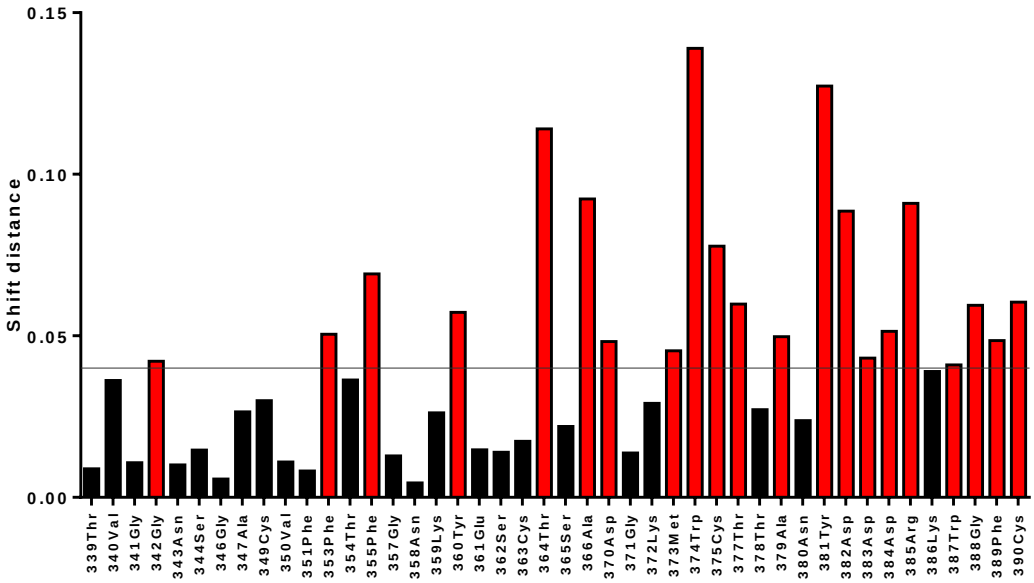


Figure 4-23: Magnitude of the chemical shift changes for the residues of the third module in the presence of type I gelatin at a protein:gelatin ratio of 1:4. Chemical shifts distances were calculated using the CCPN analysis software. A threshold for significant chemical shift distance was set at 0.04.

The residues identified in both experiments were then mapped onto the structure of the CBD in Pymol (Figure 4-24). Most of them were located on or around the binding pocket of the third module, confirming the validity of the results.

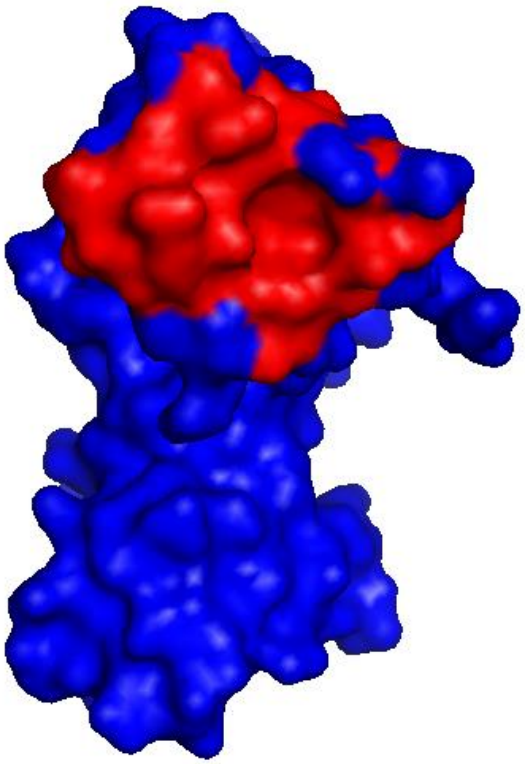


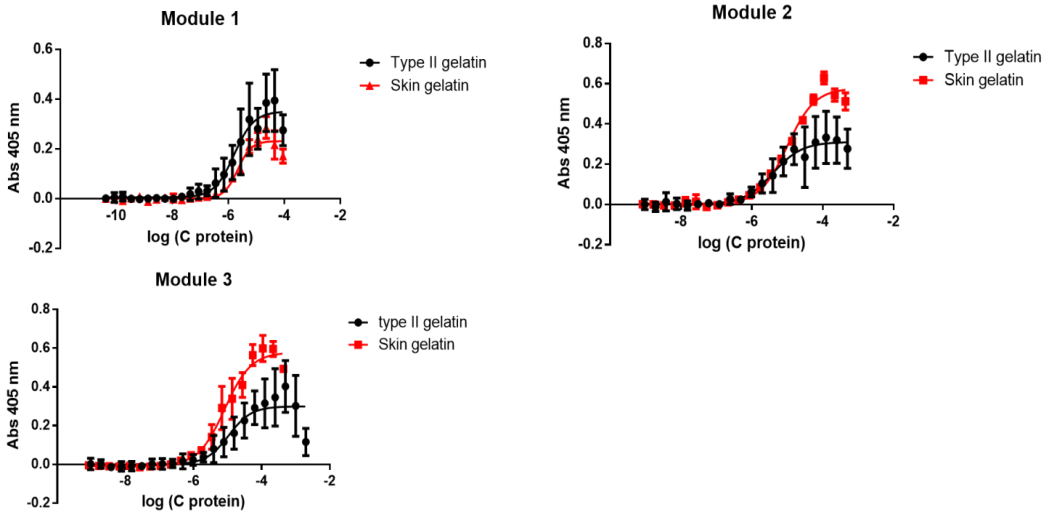
Figure 4-24: Three-dimensional surface representation of the CBD of MMP-2 showing the position of the residues in module 3 that underwent shift changes in the presence of both type I and type II gelatin at a 1:5 and 1:4 protein:gelatin ratio, for type II and type I gelatin, respectively. Residues undergoing changes are indicated in red. The structure shown is from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7).

4.4.3.1.4. Analysis of titration data

The initial aim of the NMR experiments with the single modules was not only to identify the residues forming the binding site, but also to determine a dissociation constant for each module. However, the binding constants for all three modules appeared to be too tight for the dissociation constants to be accurately determined using ligand-induced chemical shift changes in increasing ligand concentrations until saturation is reached. As the anticipated Kds are in the micromolar range, it would be necessary to use lineshape fitting analyses of the NMR line-broadening. It was, therefore, decided to test the binding of the modules to type I and II gelatin on the plate binding assay in order to get a quantitative measurement of their affinities for these substrates.

4.4.3.2. Binding assay

The plate binding assay described earlier was then developed in order to quantify the affinities of the individual modules for type II gelatin. Skin gelatin was also included as a control of the behaviour of the modules on a gelatin mostly composed of type I. The apparent Kd for all proteins and their binding curves are presented in Figure 4-25. It was shown that module 1 binds approximatively three times better to type II than to skin gelatin, while module 3 binds three times better to skin gelatin compared to type II gelatin, although these differences were not significant. Finally, module 2 was shown to bind equally well to both types of gelatin. These data, combined to those obtained on the gelatin sepharose column (*i.e.* type I gelatin), suggested that module 1 could be specific for binding to type II gelatin, while module 3 would bind more preferentially to type I gelatin.



N = 3	Type II gelatin – Kd	Skin gelatin - Kd
Module 1	1.66 µM ± 0.16 µM	4.77 µM ± 1.88 µM
Module 2	4.26 µM ± 2.53 µM	3.72 µM ± 1.21 µM
Module 3	12.0 µM ± 4.94 µM	4.09 µM ± 0.96 µM

Figure 4-25: Binding of the individual modules of the CBD to type II and type I gelatin plates. The binding assay relies on the use of plates coated with either type II or skin gelatin (thermodenatured collagen). The biotinylated proteins were added over a concentration range before adding streptavidin-coupled alkaline phosphatase. Its substrate, namely PNPP, was added and converted into a coloured product which formation was measured by absorbance at 405 nm. Examples of representative binding curves are shown for each module. Error bars represent standard deviations of three replicates. The apparent Kd values for each module on each substrate is indicated in the table (Kd ± Standard error). The differences in the affinity of each module for the two substrates were not significant (paired two-tailed t-test, p values of 0.1488, 0.8579 and 0.1073 for module 1, module 2 and module 3, respectively).

Overall, performing these experiments on the single modules has enabled us to identify the key residues belonging to the binding sites of each module, as well as to determine the affinities of each module for both type I and type II gelatin. However, it seemed important to confirm these results on the full-length CBD protein, as our end goal was to design a trimodular protein with a higher affinity for type II gelatin than the CBD itself. We therefore expressed and purified an ^{15}N labelled CBD and studied its binding to both type I and type II gelatin by NMR.

4.4.4. Characterisation of the binding of the full-length CBD by NMR

196 spectra of the free protein at 50 µM in 25 mM sodium phosphate buffer, pH 6.5 were acquired with a 800 kHz Bruker spectrometer before adding the ligand at a 33:1 (protein:gelatin) ratio (Figure 4-26).

Disappearing peaks were observed following the addition of the ligand, indicating their involvement in the binding to type II gelatin (Figure 4-26). For clarity, the spectra in Figure 4-26 was divided into four quadrants which are zoomed in in Figures 4-27 to 4-30.

Residues with broadened peaks included M282, N285, G288, Q289, F297, G299, Y302, C305, E308, G309, D312, G313, Y314, W316, T319, E321, Y323, D326, Y329 and G330. Residues F280, E287, F295, R310, R315, C317, D324, R325, K327, G330 and F331 were also shown to undergo chemical shifts upon the addition of the ligand (Figures 4-27 to 4-30). They were mapped onto the PDB structure of the CBD (Accession Number 1CK7) and belonged to the identified binding site of the second module (Figure 4-31).

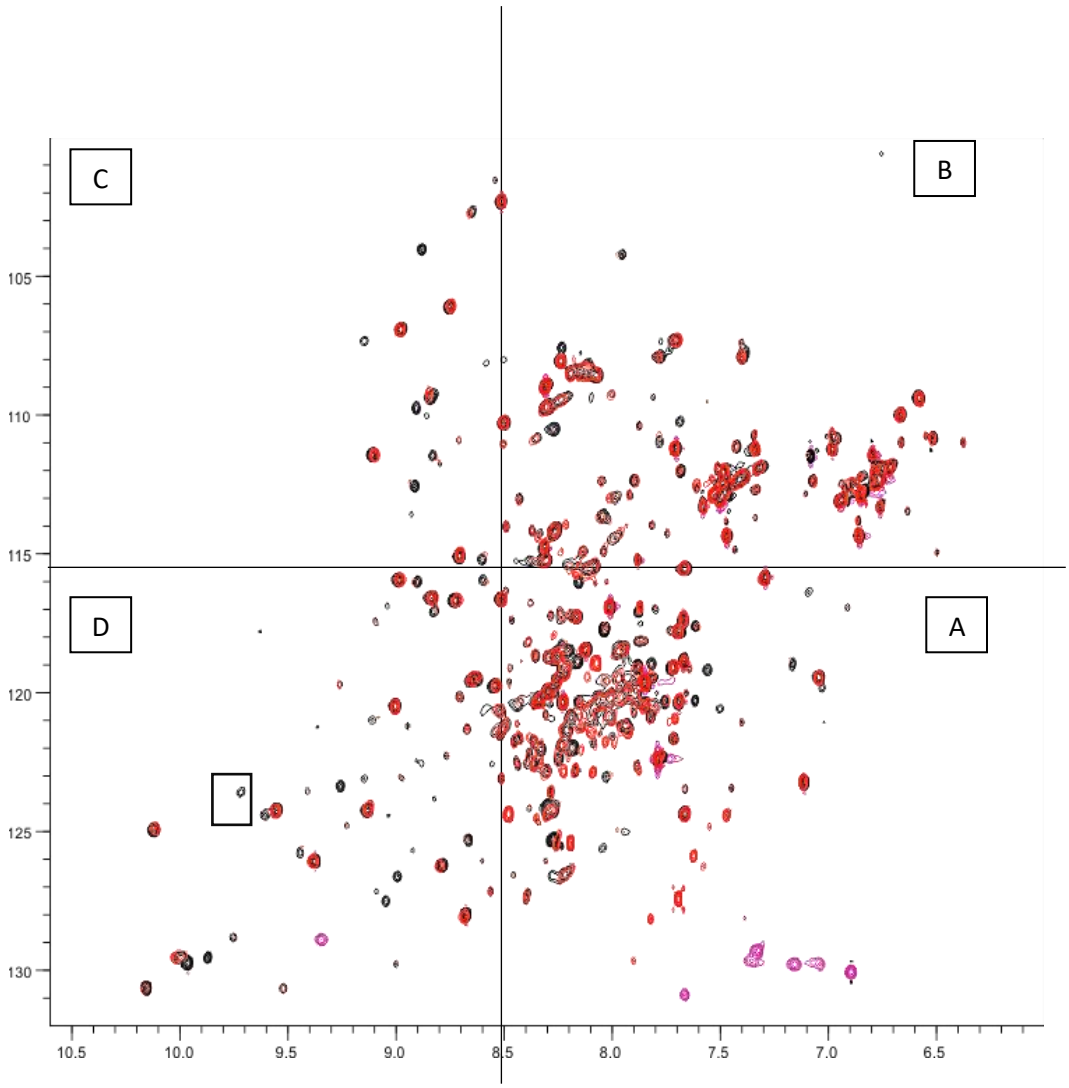


Figure 4-26: ^{1}H-^{15}N HSQC spectra of the CBD before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free CBD (black) and of the CBD in the presence of type II gelatin (red) are superimposed. A 33:1 (protein:gelatin) molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the CBD. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peak within the square is one example of peaks showing broadening: there is no corresponding red peak.

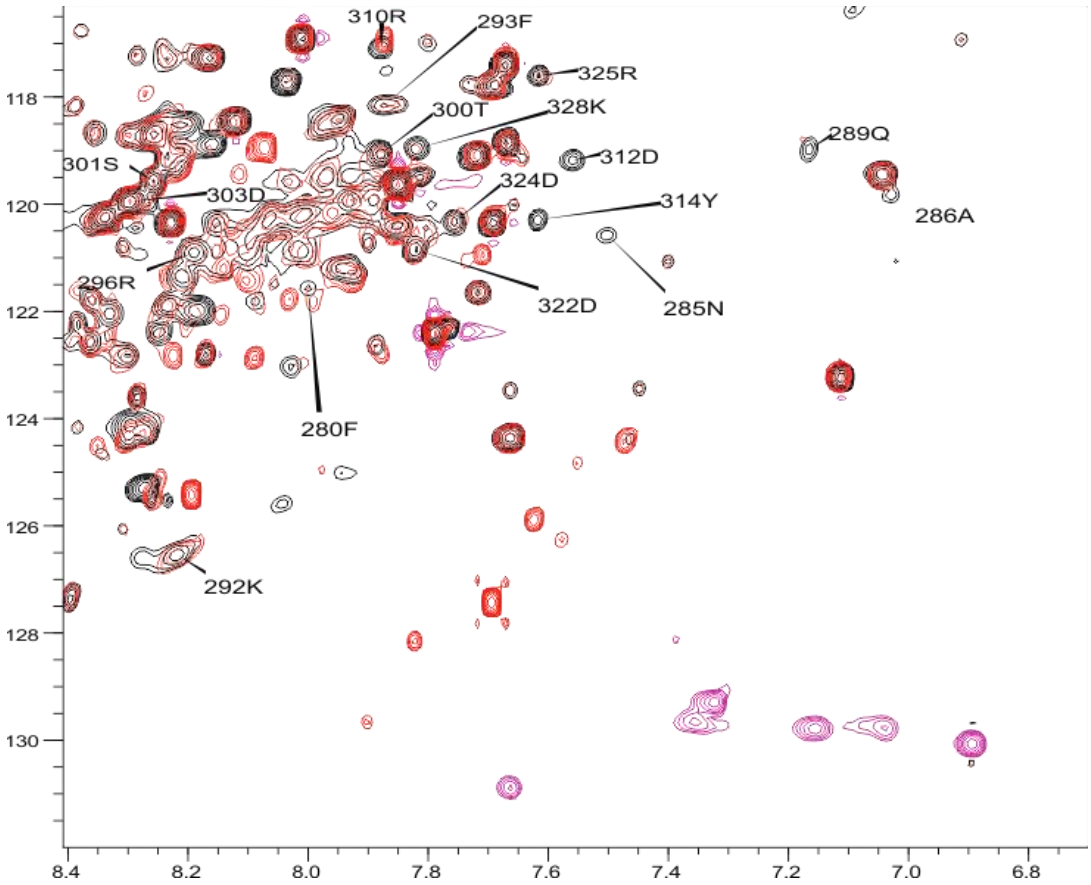


Figure 4-27. Part A of Figure 4-26. ^{1}H-^{15}N HSQC spectra of the CBD before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free CBD (black) and of the CBD in the presence of type II gelatin (red) are superimposed. A 33:1 (protein:gelatin) molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the CBD. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

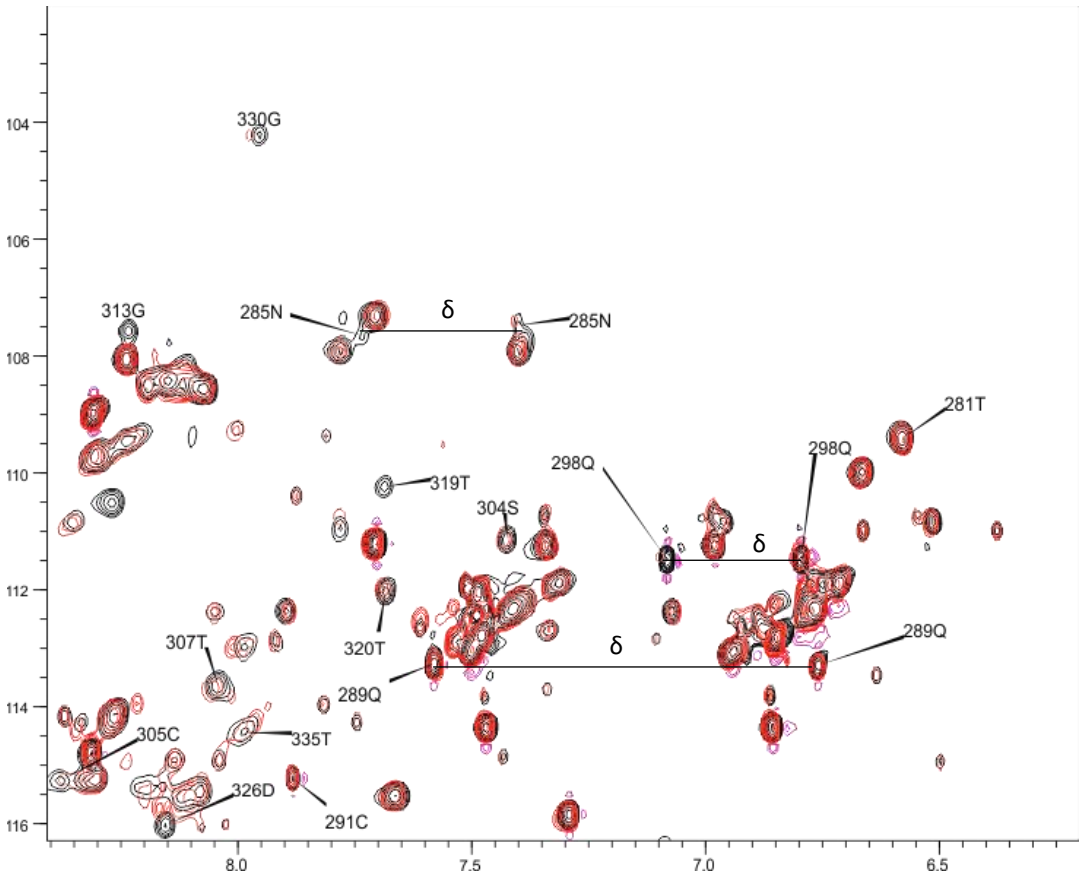


Figure 4-28. Part B of Figure 4-26. ^{1}H-^{15}N HSQC spectra of the CBD before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free CBD (black) and of the CBD in the presence of type II gelatin (red) are superimposed. A 33:1 (protein:gelatin) molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the CBD. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

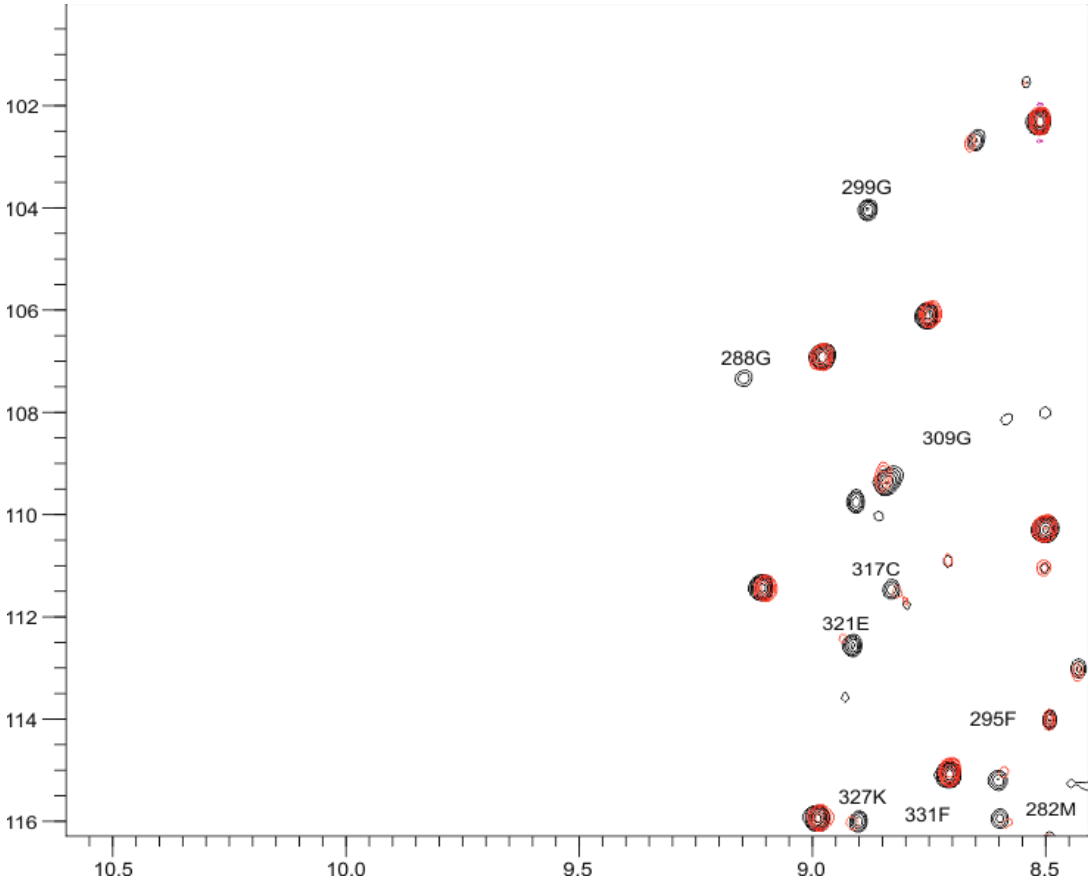


Figure 4-29: Part C of Figure 4-26. ^{1}H-^{15}N HSQC spectra of the CBD before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free CBD (black) and of the CBD in the presence of type II gelatin (red) are superimposed. A 33:1 (protein:gelatin) molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the CBD. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

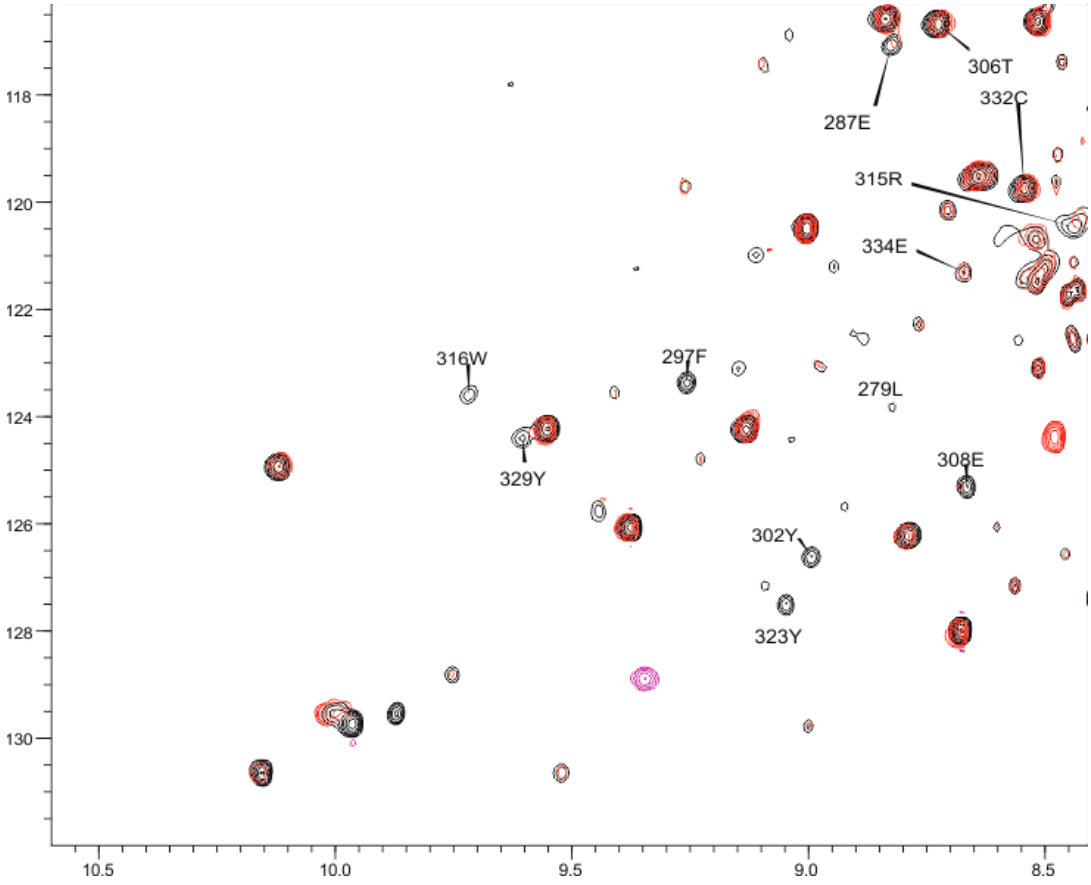


Figure 4-30: Part D of Figure 4-26. ^{1}H-^{15}N HSQC spectra of the CBD before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free CBD (black) and of the CBD in the presence of type II gelatin (red) are superimposed. A 33:1 (protein:gelatin) molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in the CBD. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using previously published results, and are labelled with their one-letter amino acid code and residue number.

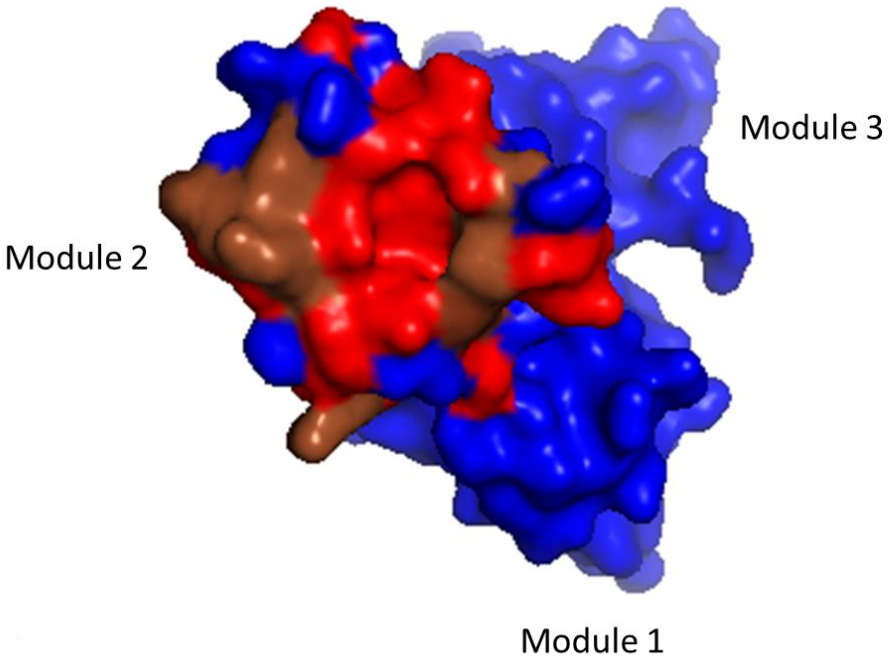


Figure 4-31: Three-dimensional surface representation of the CBD of MMP-2 showing the position of the residues that underwent peak broadening or shift changes in the presence of type II gelatin at a 33:1 protein:gelatin ratio. Residues undergoing peak broadening are indicated in red while residues undergoing shift changes are indicated in brown. The structure shown is from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7)

These data on the CBD showed that module 2 was the strongest module for binding to type II gelatin. When the same experiment was performed with type I gelatin, the exact same residues were affected upon the addition of this ligand (not shown). These results suggest that module 2 is the strongest, within the CBD, for binding to either type I or type II gelatin, providing a strong affinity, but no specificity for one type of gelatin over the other. Unfortunately, higher concentrations of either type I or type II gelatin did not enable to detect shifts in either module 3 or module 1 (not shown), meaning that the specificity of these modules for each type of gelatin could not be tested in the full-length protein.

4.5. Discussion

4.5.1. General conclusion

This chapter has described the expression, purification and the characterisation of the binding of each of the individual modules of the CBD to both type I and type II gelatin. All three modules were successfully expressed in Shuffle cells and purified following the method established for the CBD. Characterisation of the proteins confirmed their proper folding, with the DTNB assay showing the presence of two disulfide bridges in each module, and 2D NMR spectra confirming the correct tertiary structure of each module. Modules 2 and 3 were also shown to be able to bind to gelatin sepharose, although module 1 was mostly recovered in the flowthrough. This observation corroborates with the low affinity of this module for type I gelatin, which has been reported by others [140,144]. The functionality of this protein was later demonstrated when it was shown to bind to gelatin by NMR.

NMR ligand binding studies enabled to identify residues involved in the binding to gelatin for each module. Most of them were shown to belong to the hydrophobic binding pockets of the modules, and also correlated with residues identified with collagen-like peptides in previous publications. Overall, module 2 showed the exact same mechanism of binding for type I and type II gelatin, and similar dissociation constants on the plate binding assay. Slight differences in the identity of the residues involved in the binding to each substrate could be observed for module 3 and module 1: residue M373 in module 3 was slightly more affected by the binding to type I gelatin, while residues F237, G251, G255, N264 and G269 in module 1 were more affected by the binding onto type II gelatin. These differences could be due to experimental errors, but the fact that module 1 seemed to have a slightly higher affinity for type II gelatin, while module 3 seemed more specific to type I in the plate binding assay further reinforces the hypothesis that these modules could each be specific to one type of gelatin.

Following up on these observations on the individual modules, it was then decided to perform the same NMR ligand binding studies with the full-length CBD. These experiments showed that both type I and II gelatin bind the CBD primarily through its second module. More residues were affected by the binding in this configuration than when testing module 2 alone, which could be explained by the flexibility of the domains within the CBD.

To the best of our knowledge, this is the first time that NMR has been used to characterise the binding of the CBD of MMP-2 to type II gelatin. Data were available on type I gelatin like peptides, but the study on the full-length ligand had not been done before either.

4.5.2. Future work

How do the residues identified in these experiments compare to previously published results?

Several collagen-like peptides have been tested against either single modules or the full-length CBD in order to identify key residues involved in the binding. As outlined in the introduction of this chapter, it seemed crucial to test the binding of our proteins on more relevant substrates, which is why we carried out the above mentioned experiments on type I and type II gelatin.

However, it is interesting to compare the results obtained in all these different experiments. Below are the representations of the single modules used in the introduction of this chapter (Figure 4-32). All of the residues in red and grey, which were identified in previous experiments, were also shown to be involved in the binding of type I and type II gelatin, except for T250, F273 (not assigned), and residues in grey in module 1, and N358 in module 3. Residues F295 and D312 in module 2 did not reach the set threshold, but were affected by the binding of each ligand and their peaks broadened when looking at the full-length CBD.

Residues in cyan have been added to this figure (Figure 4-32). Those left unlabelled are residues which had been previously shown to be shifting upon the addition of some or all of the gelatin-like peptides, and which also showed significant shifts in our experiments. Most of them do not belong to the hydrophobic pocket, meaning that their shifts must be a consequence of a change in conformation instead of a direct involvement in the binding.

Finally, residues in cyan which are labelled are those which had never been reported to be involved in the binding with any of the peptides, but which showed some important shifts following the binding of either type of gelatin in our experiment. Again, most of them are located at the back of the binding pocket, which means that their shift must originate from a further change in conformation of the modules upon the binding of gelatin. However, it is interesting to note that other residues lie close to the hydrophobic pocket, such as G269 in module 1, Q298 in module 2 and G342 in module 3. It is particularly interesting to note that G269 was shown to be only involved in the binding to type II gelatin in our experiment, a substrate which had never been tested before.

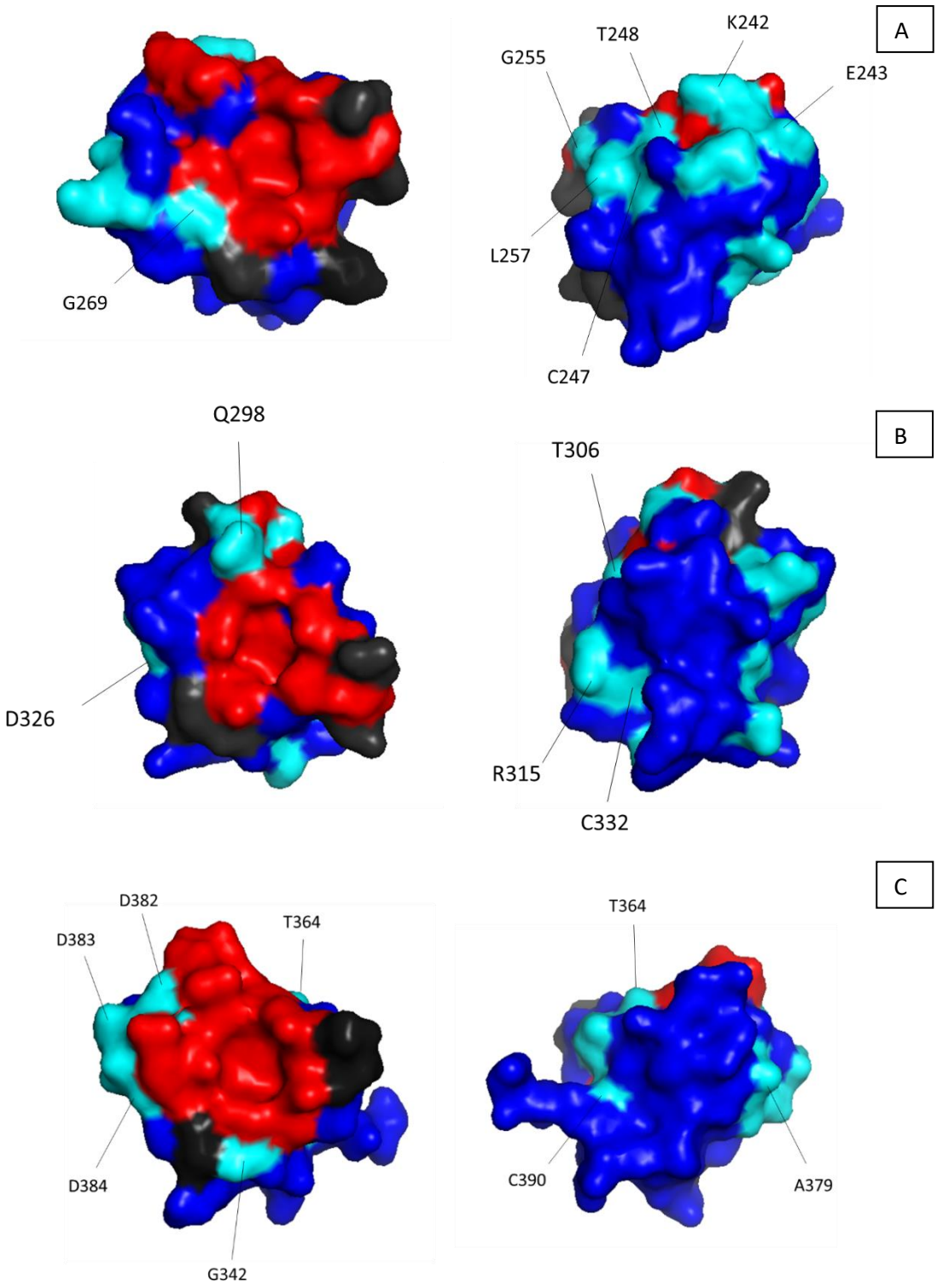


Figure 4-32: Three-dimensional surface representation of module 1 (A), module 2 (B) and module 3 (C) of the CBD of MMP-2 showing the position of the residues that underwent shift changes in the presence of collagen-like peptides and type I and type II gelatin. Residues undergoing changes in all situations are pictured in red and grey. Residues in cyan which are unlabelled were identified in previously published papers as well as in our experiment. Labelled residues in cyan were identified only in our experiments. The structures shown are from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7). Left and right structures are front and back, respectively.

Residues which showed some of the most important shifts following the addition of gelatin in our experiments were also identified as key residues by authors who mutated them for alanines. This includes residues Y302, Y323, D326 and Y329, which were shown to be critical for structure integrity [168] and Y314, which was shown to be involved in the binding [168]. Another paper reported that mutations of both W316 and W374 by alanines abrogated the binding of the CBD to both type I gelatin and collagen [147]. Finally, as outlined in the introduction of this chapter, the mutations R252A, R296A, F297A, Y302A, Y323A, Y329A, R368A, W374A and Y381A all affected the binding of the CBD to type I gelatin [140]. Again, these residues were all involved in the binding to both types of gelatin in our experiment.

Can the modules of the CBD be specific to different substrates?

As mentioned earlier, we observed that module 1 could bind better to type II gelatin in a plate binding assay, while module 3 seemed to bind with a slightly higher affinity to type I gelatin. Considering the high degree of sequence identity between the modules (58.5 % for modules 1 and 2, 55.7 % for modules 1 and 3, 58.9 % for modules 2 and 3), these differences could be surprising. However, previously published results have identified differences in the relative affinities of the modules for various substrates, suggesting that specificity could be possible.

As mentioned in the introduction of this chapter, the first module was often described as the weakest of the three for its binding to type I gelatin, but it had the highest reported Ka for binding to the PPG6 peptide [160,159]. Similarly, module 3 had the smallest affinity for the PPG6 peptide, but the highest for a peptide with a different sequence, namely p33-42 [160,159]. In a different context, the first module of the CBD was also shown to bind with the highest affinity to fatty acids, enabling them to inhibit the gelatinolysis activity of MMP-2 [182]. Taken together, these data confirm that the modules, in spite of their sequence similarity, can have different affinities towards the same substrate.

Another argument in favour of module specificity lies in the CBD of MMP-9. It was shown that the second module on its own binds better to gelatin than the full-length CBD [161]. The maintenance of the other two modules through natural selection must mean that they are required to bind to other substrates. Other matrix components, such as elastin, were hypothesised as potential targets for the first and third modules of the CBD of MMP-9 [183].

When mapping the different residues involved in the binding onto the alignment of the three modules (Figure 4-33), we could observe that residues belonging to the hydrophobic pockets were mostly conserved in each module. Residues F237, G251, G255 in module 1 (corresponding to residues F18, G32, G36 in Figure 4-33 below), which showed a preference for type II gelatin over type I in module 1, are also conserved in all three modules, while N264 in module 1 (N45 below) is conserved in the third one. It is therefore unlikely that these residues can confer any specificity for one substrate over the other in module 1. However, residues G269 (G50 in Figure 4-33) in module 1 and M373 (M38 in Figure 4-33) in module 3, which both showed preference for type II and type I gelatin, respectively, are not conserved in the other two modules, and could therefore contribute to gelatin specificity. However, M373 was shown to be present at the back of the binding pocket, and might be involved in a conformational change more than in a direct binding with the substrate.

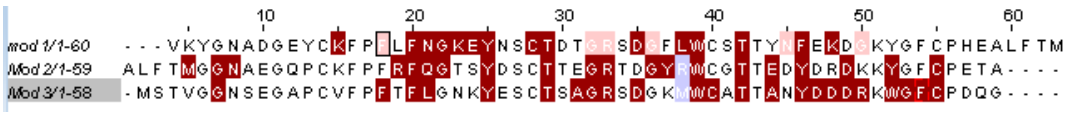


Figure 4-33: Alignment of modules 1, 2 and 3 of the CBD highlighting the residues involved in the binding to type I and type II gelatin. Residues common to both substrates are highlighted in red, while residues showing higher shifts for type II gelatin are in pink, and those showing higher shifts for type I gelatin in grey.

Considering all of these aspects, and based on our findings, we think it would be possible that the second module of the CBD of MMP-2 provides strong affinity towards both type I and type II gelatin, while module 1 would be more specific to type II gelatin, and module 3

to type I. In order to investigate this hypothesis further, and with the aim of generating more potent proteins than the CBD, it was therefore decided to design the chimeric proteins 111, 222 and 333 and to test their affinities for type I and type II gelatin.

5. Generation of more potent gelatin binding proteins

5.1. Introduction: chimeric proteins

Proteins are often subdivided into domains or modules, which are autonomous both functionally and structurally [184]. The study of these subdomains has been made easier with the development of genetic engineering techniques, which have enabled study of the function of the domains individually, as well as to generate chimeric proteins [184].

A common example is antibodies, which are Y-shaped proteins made of an effector domain (the Fc), and two antigen binding domains (the Fabs), both directed towards the same target [185]. Swapping these independent fragments has enabled the generation of new versions of antibodies with enhanced functions. One example is bispecific antibodies, where the Fab of two different antibodies are linked to the Fc of another, generating a new antibody with an unaltered structure and the ability to bind to two different antigens at the same time [185,186]. Other variants have also emerged with new combinations of the different domains, leading to constructs such as TriFabs, which contain three antigen-binding regions, fused at their C-terminus by a peptide linker [187]. The latest advance in the field has seen the fusion of an antigen-binding region of a monoclonal antibody to the transmembrane domain of a T-cell, enabling the elimination of cancerous cells through the recognition of their tumour-associated antigen by the chimeric T-cell receptor [188].

However, not all domains of all proteins are as interchangeable as those found in antibodies. Domain-domain or module-module interactions are often observed, and can influence both the structure and the function of the protein [184]. Care should therefore be taken when designing chimeric proteins in order to preserve the overall structure of the original protein [184].

As demonstrated in the previous chapter, we identified that module 2 is the strongest at binding to both type I and type II gelatin within the CBD. Other results showed that modules 1 and 3 could be specific for type II and type I gelatin, respectively. Finally, the CBD was previously shown to have a much tighter affinity than the single modules [144], which suggests a cooperative mechanism between the modules within the trimodular construct. We decided to exploit these observations by designing chimeric proteins made of new combinations of the individual modules, generating the chimeric proteins 111, 222 and 333. The goal was to design these proteins with the same fold as the original CBD in order to preserve the function of the protein as well as the cooperative mechanism between modules. Additionally, as each module constitutes a binding site, we wanted to investigate whether trimodular proteins made of the same module would conserve the specificity of the modules observed earlier. Our hypothesis was that 222 would provide a protein with a tight affinity for both types of gelatin, while 111 and 333 would have a high affinity for type II and type I gelatin, respectively.

5.2. Materials and Methods

5.2.1. Designing of chimeric proteins

All three proteins were designed in an attempt to maintain the module-module interactions present in the CBD. These interactions were studied using the PISA (Proteins, Interfaces, Structures and Assemblies) online tool. The residues involved in either hydrophobic, aromatic-aromatic, aromatic-sulphur, ionic or cation-pi interactions, as well as those forming hydrogen bonds, were conserved as in the original CBD sequence, as shown in blue in Figure 5-1. The modules are highlighted in red, and the linkers in black.

111:

EGQVVRVKYVNADGEYCKFPFLFNGKEYNSCTDTGRSDGFLWCSTTYNFEKDG
KYGFCPHEALFVMYVNADGEYCKFPFLFNGKEYNSCTDTGRSDGFRWCSTTYN
FEKDGKYGFCPETAVRVKYVNADGEYCKFPFLFNGKEYNSCTDTGRSDGFLWC
STTYNFEKDGKYGFCPDQGYSL

222:

EGQVVFTMYGNAEGQPCKFPFRFQGTSYDSCTTEGRTDGYRWCGTTEDYDRD
KKYGFCPHEALFTMGGNAEGQPCKFPFRFQGTSYDSCTTEGRTDGYRWCGTTE
DYDRDKKYGFCPETALFTMGGNAEGQPCKFPFRFQGTSYDSCTTEGRTDGYRW
CGTTEDYDRDKKYGFCPDQGYSL

333:

EGQVVSTVYGNSEGAPCVFPFTFLGNKYESCTSAGRSDGFLWCATTANYDDDR
KWGFCPHEALFTMGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGKRWCATTA
NYDDDRKWGFCPETAMSTVGGNSEGAPCVFPFTFLGNKYESCTSAGRSDGKM
WCATTANYDDDRKWGFCPDQGYSL

Figure 5-1: Sequences of the chimeric proteins 111, 222 and 333. The residues which might be involved in intramolecular interactions, and which were therefore left as in the native CBD, are shown in blue. The modules are highlighted in red, and the linkers in black.

5.2.2. Cloning

The synthetic DNA constructs (GeneArt, ThermoFisher) encoding for these sequences were codon optimised and cloned into the pOPINS vector using the ligase independent InFusion cloning methodology. The constructs which were generated carried a His_6-SUMO tag, followed by a cleavage site for the SUMO protease.

5.2.3. Expression in *E.coli* strains

E.coli Shuffle cells were transformed with the sequenced expression plasmids. Conditions of expression such as the temperature and length of the production phase as well as the concentration of inducer were optimised on a small scale.

On a larger scale, protein expression was carried out in 2L flasks. Cells were grown either in LB or minimal media at 37°C, induced for protein expression with IPTG, harvested following the production phase and stored at -80°C.

As 111 was shown to be highly insoluble, the refolding protocol described by Steffensen et al. [141] was tested. Briefly, the insoluble pellet was washed, solubilized in 6M Guanidine Hydrochloride and incubated for 1 hour at room temperature. The non-dissolved aggregates were removed by centrifugation while the supernatant was dialysed from 2 hours to overnight against aerated 0.1 M sodium borate buffer, pH 10.0. The sample was then dialysed against chromatography buffer and subjected to the same purification protocol as for the other two constructs.

5.2.4. Purification strategy

Following cell lysis, the released recombinant protein was captured on a nickel column, exploiting the affinity of the hexahistidine tag for this resin. The protein was eluted with a gradient of imidazole, and the tag cleaved with a histidine-tagged SUMO protease (produced in-house). A reverse nickel purification enabled the isolation of the cleaved protein from the tag and the protease. The polishing step was performed with an Ion Exchange column, and the mode of elution (either linear or step) was optimized for each protein.

5.2.5. Characterisation

The expressed proteins were characterised for their identity using intact mass analysis by Mass Spectrometry (ElectroSpray Ionisation), as well as for their folding, as the number of disulfide bridges was quantified with a DTNB assay. The activity of 222 and 333 was also tested on a gelatin sepharose resin. Finally, the integrity of the three-dimensional tertiary structure of 222 was assessed by 2D NMR.

5.2.6. Plate binding assay

The affinity of the constructs for both type I and type II gelatin was tested using the plate binding assay which was described in the previous chapter. Briefly, proteins were biotinylated and added over a concentration range onto gelatin-coated plates. Alkaline phosphatase coupled to streptavidin was added to the bound proteins, and its substrate was

transformed into a coloured product. The absorbance of the product was used as an indirect indication of the amount of bound protein.

5.2.7. Binding of 222 to type II gelatin by NMR

For NMR purposes, 222 was expressed in ^{15}N-labelled minimal media. The conditions of expression and purification were kept the same as for production in LB. The protein was then buffer exchanged into 25 mM NaP, pH 6.5 and 1D and 2D spectra of the free proteins were obtained. Type II gelatin was then added at protein:gelatin ratios of 40:1 and 20:1. Spectra were acquired at each substrate concentration and the shift or broadening of peaks, indicative of binding, were observed.

5.3. Results

5.3.1. Small scale screening for optimal conditions of expression

E.coli Shuffle cells were transformed with the sequence-checked pOPINS vector containing the inserts for 111, 222 or 333. The concentration of the inducer, the temperature of growth and the length of the production phase were screened to find optimal conditions for protein expression. Cells were cultured at 37°C before induction, induced with a range of concentrations in IPTG (20 – 1000 µM), and then either left for a further 5 hours at 25°C (Figure 5-2, B and D) or overnight at 18°C (Figure 5-2, A and C).

Results showed that 111 was poorly expressed, with both low levels of expression and a high degree of insolubility (Figure 5-2, A and B). Concentrations in IPTG of 200 and 1000 µM and an expression phase at 25°C for 5 hours seemed to give the highest ratio of soluble : insoluble protein (Figure 5-2, B), and both of these were selected for a scale-up.

The constructs 222 and 333 showed higher levels of expression and solubility (Figure 5-2, C and D, respectively), with around 50 % of 222 recovered in the pellet (Figure 5-2, C). A production at 18°C induced with 1 mM was selected for 222, while a production at 25°C induced with 800 µM of IPTG was scaled-up for 333, as these conditions gave the highest ratios of soluble over insoluble protein.

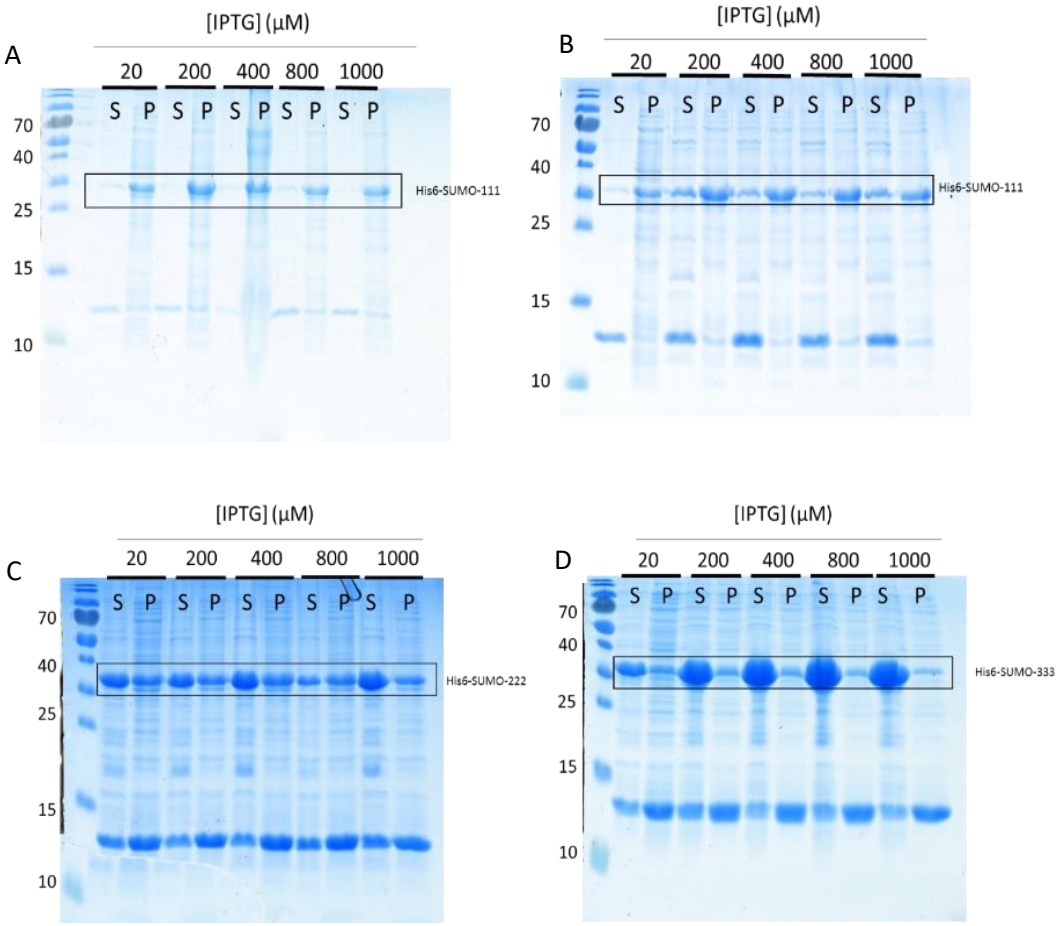


Figure 5-2: SDS PAGE showing the screening for expression of the His_6-SUMO-111 (A and B), His_6-SUMO-222 (C), and His_6-SUMO-333 (D) in LB. *E.coli* Shuffle cells were transformed with plasmids encoding for either 111, 222 or 333, each carrying a His_6-SUMO tag. Cells were grown in 10 mL LB until reaching an OD (600 nm) of 0.6. Cells were induced with various concentrations of IPTG overnight at 18°C (A and C), or for 5 hours at 25°C (B and D). Cells were then lysed and the soluble (S) fraction was subjected to a nickel purification. The insoluble fraction was solubilised in 8M urea and loaded on the gel (P: Pellet). The same conditions were tested for each construct, but only a subset is presented in this figure.

5.3.2. Large scale expression with the optimised conditions

The selected conditions were scaled-up to 2L flasks. 222 (Figure 5-3) and 333 (Figure 5-4) were successfully expressed, although 222 showed a high level of insolubility (Figure 5-3, gel A, lane I). Both proteins were purified as previously described for the CBD and its individual modules, with a linear elution from the Ion Exchange Q column (Figure 5-3, B and Figure 5-4, B).

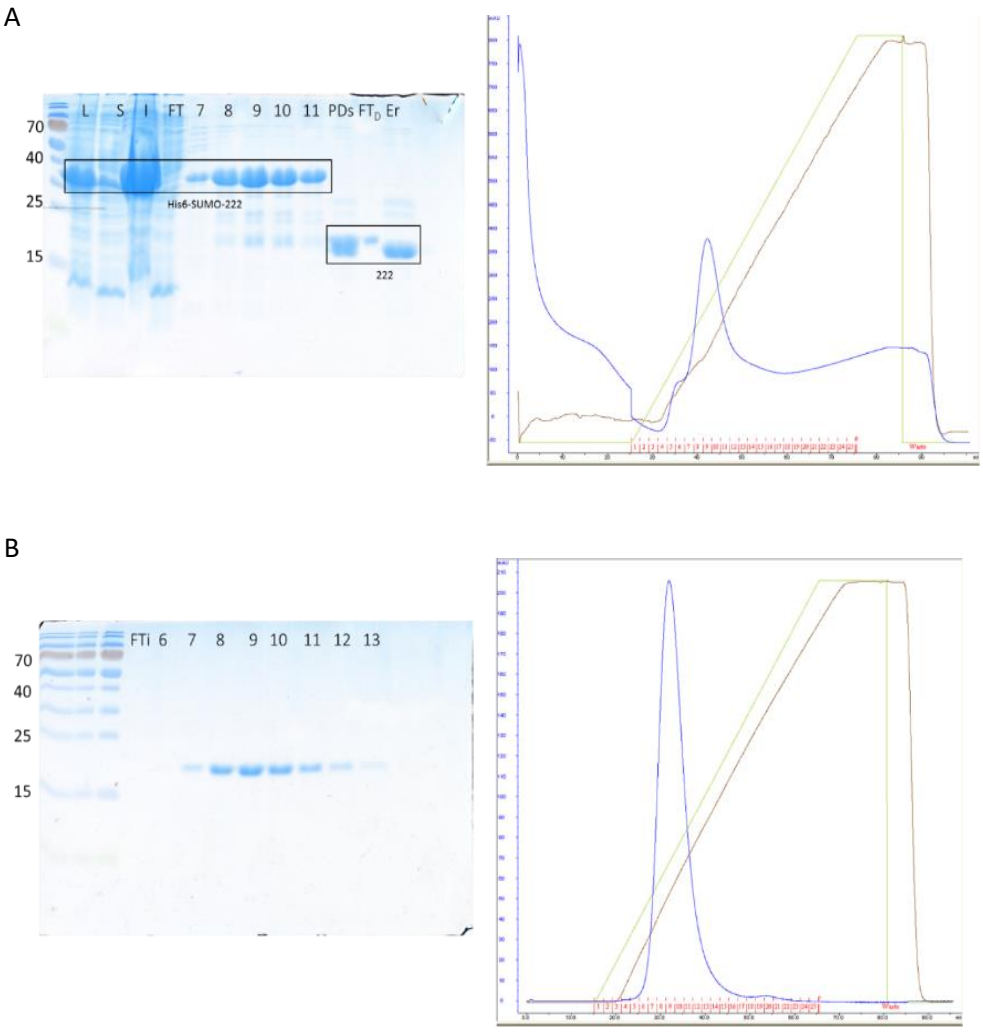


Figure 5-3: Expression and purification of 222. A-The chimeric protein 222, originally carrying a His_6-SUMO tag, was expressed in Shuffle *E.coli* cells and purified by nickel column and reverse nickel column purification, as previously described. B-The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a linear gradient of sodium chloride. L: Lysate; S: Soluble; I: Inclusion Bodies; FT: Flowthrough; PD_S: Post-Dialysis soluble; FT_D: Flowthrough Desalt; Er: Elution Reverse. FTi: Flowthrough Ion Exchange. The numbers labelled on the gels correspond to the fractions of the His Trap chromatogram (Part A) and of the Ion Exchange one (Part B). On the chromatograms: Blue line: UV trace; Green line: Concentration of IEC B buffer; Brown line: conductivity.

A B

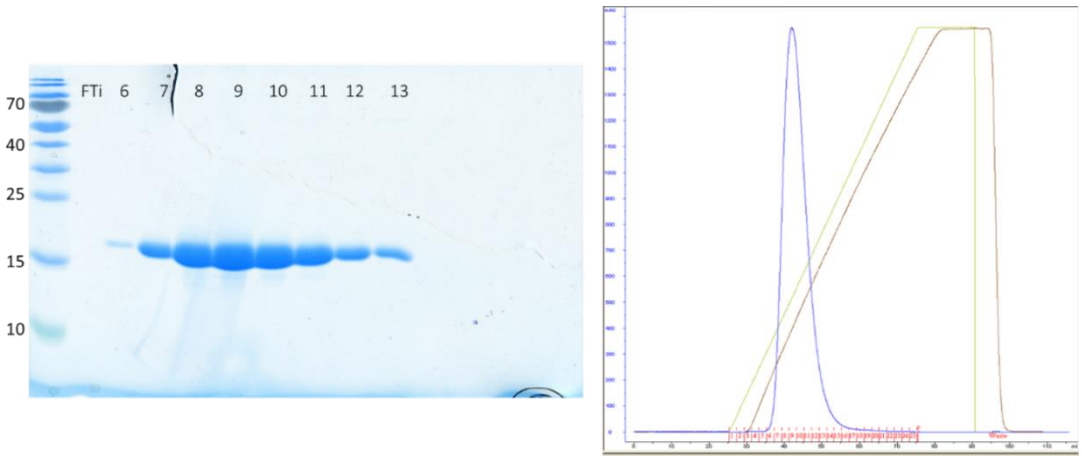


Figure 5-4: SDS PAGE gel showing the expression of the chimeric protein 333 in LB (A) and corresponding chromatogram (B). A-The 333 protein, originally carrying a His_6-SUMO tag, was expressed in Shuffle *E.coli* cells and purified by nickel column and reverse nickel column purification, as previously described. The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a linear gradient of sodium chloride. FT: Flowthrough; 6-13: Ion Exchange fractions. B - Chromatogram corresponding to the Ion Exchange purification of 333. Fractions 6 to 13 were loaded on the gel. Blue line: UV trace; Green line: Concentration of IEC B buffer; Brown line: conductivity.

However, 111 was mostly found in the insoluble fraction, as shown in Figure 5-5 (A), where the protein production was induced with 1000 µM IPTG. The same was observed when inducing with 200 µM (not shown). The protein recovered from the reverse purification (Figure 5-5, lane FTr, gel A) was submitted to desalting and Ion Exchange Chromatography. Only a very low amount of protein was obtained, and impurities could still be observed (Figure 5-5, lane IEC, gel B). No protein was lost in the flowthough of the His Trap (Figure 5-5, FT1, FT2 and FT3, gel B) nor in the flowthrough or washes of the Ion Exchange (Figure 5-5, lane FTi and Wi, gel B). Lowering the temperature to 18°C was tested, as well as shortening the production phase to 2.5 hours, but none of these conditions improved the level of solubility of 111 (not shown).

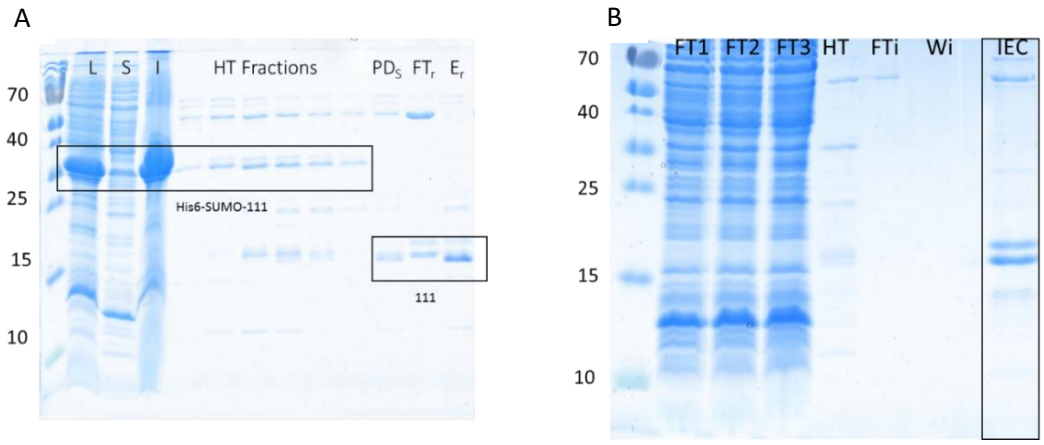


Figure 5-5: SDS PAGE gel showing the insolubility of 111. A-The 111 protein, originally carrying a His_6-SUMO tag, was expressed in Shuffle *E.coli* cells and purified by nickel column and reverse nickel column purification, as previously described. B-The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a linear gradient of sodium chloride (IEC lane, gel B). Gel A: L: Lysate; S: Soluble; I: Inclusion Bodies; HT fractions: fractions from the His Trap chromatography; PD_S: Post-Dialysis soluble; FT_r: Flowthrough Reverse; Er: Elution Reverse. Gel B: FT: Flowthrough from the His Trap; HT: Sample from a His Trap fraction; FTi: Flowthrough from the Ion Exchange; Wi: washes from the Ion Exchange; IEC: 111 sample recovered after the Ion Exchange Chromatography.

Considering the low amount of soluble protein which was recovered, as well as the large amount of 111 present in the pellet, it was decided to test the refolding protocol that was first described on the CBD [141]. Briefly, the insoluble pellets from the production phase at 25°C for both 2.5 and 5 hours were washed and dissolved in 6M guanidine hydrochloride. The non-dissolved aggregates were separated from the denatured protein by centrifugation, and the supernatant was allowed to refold through dialysis against 0.1 M aerated sodium borate buffer, pH 10.0, for 2 hours at 20°C, followed by an overnight dialysis against 100 mM sodium phosphate, 0.5 M NaCl, pH 8.0. The sample was then spun again, and the supernatant dialysed against His Trap A buffer, before being subjected to the same chromatography method as 222 and 333.

Results showed that, although some protein was lost in the flowthrough of the His Trap chromatography (Figure 5-6, lane FT, gel A), some could also be recovered on the resin (Figure 5-6, lanes 6-16, gel A). After desalting and Ion Exchange Chromatography, the recovered protein was pure (Figure 5-6, lanes 7 to 17, gel B). The next step was to confirm whether or not this protein, as well as 222 and 333, were correctly folded.

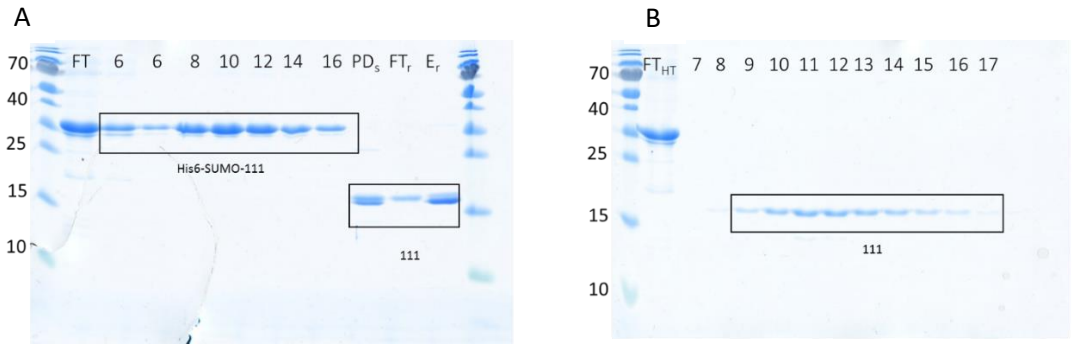


Figure 5-6: SDS PAGE gel showing the purification of the refolded 111. A-The 111 protein, originally carrying a His_6-SUMO tag, was refolded from the insoluble pellet from Shuffle *E.coli* cells and purified by nickel column and reverse nickel column purification, as previously described. B- The protein sample was then desalted and loaded onto a 5 mL Q FF column, and eluted with a linear gradient of sodium chloride. FT: Flowthrough; PD_S: Post-Dialysis soluble; FT_r: Flowthrough Reverse; Er: Elution Reverse. FT_{HT}: Flowthrough from the His trap chromatography; the numbers labelled on the gels correspond to the fractions of the His Trap chromatogram (Gel A) and of the Ion Exchange one (Gel B).

5.3.3. Characterisation

5.3.3.1. DTNB assay

The assay showed that no free cysteines were present in either of these proteins, suggesting that the refolding protocol on 111 had worked. An intact Mass Spectrometry analysis on 222 confirmed the presence of 6 disulfide bridges. The measured intact mass of the sample was 20 638 Daltons, which corresponds to a 12-dalton deficit compared to the theoretical mass derived from the sequence (20 650 Da). This difference correlates with the formation of 6 disulfide bridges and the concomitant loss of 12 hydrogens.

5.3.3.2. Gelatin sepharose

As low amounts of 111 were available, it was decided to keep this protein sample for the plate binding assay on both type I and type II gelatin. 222 and 333 were however used on the gelatin sepharose resin, and both showed binding to the column, although, in both cases, some protein was recovered in the flowthrough (Figure 5-7).

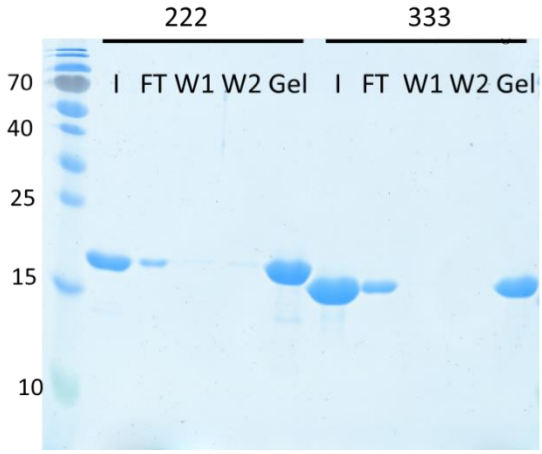


Figure 5-7: SDS PAGE showing the binding of 222 and 333 to gelatin sepharose. The 222 and 333 protein samples from the Ion Exchange were tested for their ability to bind gelatin using a gelatin sepharose resin. 100 µL of resin was spun down, washed with ethanol and equilibrated before applying the protein samples. The initial samples (I) were loaded and allowed to bind for 1 minute before collecting the flowthrough (FT). The resin was washed twice with the binding buffer (W1 and W2). The bound proteins were mixed with 2X SDS loading buffer and loaded on the gel (Gel, Gelatin-bound protein)

5.3.3.3. Plate binding assay

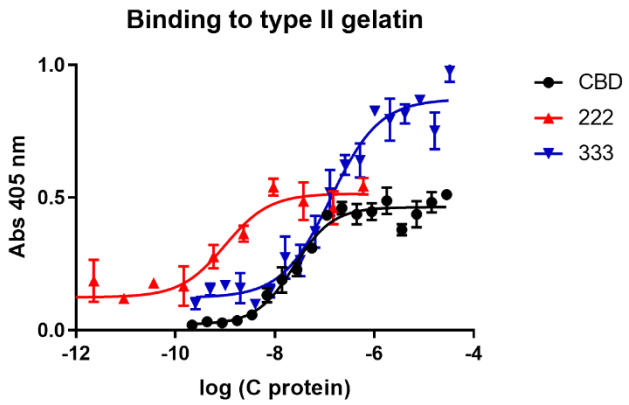
We then studied the binding of these chimeric proteins onto type I and II gelatin using the plate binding assay previously described (Figure 5-8, Part B and A, respectively). Results showed that 333 binds with a much tighter affinity to type I than type II gelatin, confirming the preference of module 3 for the former. Interestingly, 222 bound to type II gelatin with a 14-fold higher affinity than the CBD (Figure 5-8 and Table 5-1), as could be expected from the observation that module 2 was the strongest to bind to type II gelatin. However, 222 showed even tighter affinity towards type I gelatin, suggesting that module 2 could have a slight preference towards type I gelatin over type II (Figure 5-8 and Table 5-1).

Unfortunately, the refolded 111 did not show any strong binding to either type I or type II gelatin (Figure 5-9). The concentration in 111 that was used did not allow to reach saturation, meaning that the determination of a dissociation constant was not possible.

In order to investigate this further, the foldedness of the protein was checked by 1D NMR, which showed that the protein was folded like a molten-globule (not shown). A new refolding procedure was performed with longer dialysis times, in an attempt to allow more time for the protein to refold. Unfortunately, this did not improve the folding of the protein (not shown).

As 222 showed the highest affinity of all the proteins tested for type II gelatin, which we aimed to target in damaged cartilage, we then wanted to further confirm the binding of the protein to this substrate by NMR.

A



B

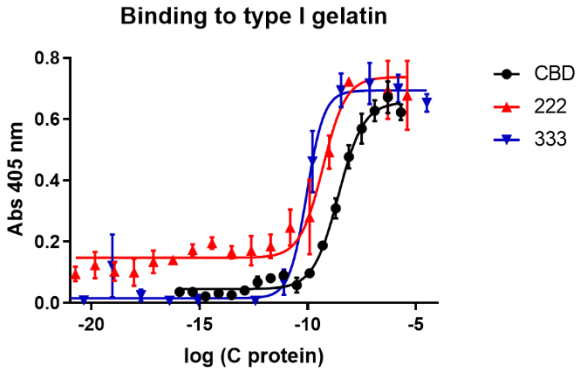


Figure 5-8: Binding of the CBD, 222 and 333 to type II (Part A) and type I (Part B) gelatin plates. The binding assay relies on the use of plates coated with gelatin (thermodenatured collagen). The biotinylated proteins were added over a concentration range before adding streptavidin-coupled alkaline phosphatase. Its substrate, namely PNPP, was added and converted into a coloured product which formation was measured by absorbance at 405 nm. Examples of representative binding curves are shown for each protein. Error bars represent standard deviations of three replicates. The apparent Kd values for each protein on each substrate are indicated in the table below (Table 5-1).

Table 5-1: Apparent dissociation constants of the CBD, 222 and 333 on both type I and type II gelatin. The dissociation constants were calculated from the binding assay described in Figure 5-8, using a 4PL regression in Prism. Numbers represent average ± standard error, and the number of replicates is indicated in brackets. Only the affinity of 333 to type II gelatin was shown to be significantly different to that of the CBD and 222 on the same substrate ($p < 0.005$, ANOVA for multiple comparisons with Tukey correction). No significant difference was observed in the affinity of each protein onto type I gelatin (p-values> 0.05, ANOVA for multiple comparisons with Tukey correction).

	Kd on Type II gelatin	**Kd on Type I gelatin**
FL CBD	20.4 ± 2.83 nM (n = 3)	1.56 ± 0.723 nM (n = 2)
222	1.46 ± 0.53 nM (n = 4)	0.315 ± 0.128 nM (n = 2)
333	117 ± 18.9 nM (n = 3)*	0.132 ± 0.039 nM (n = 3)

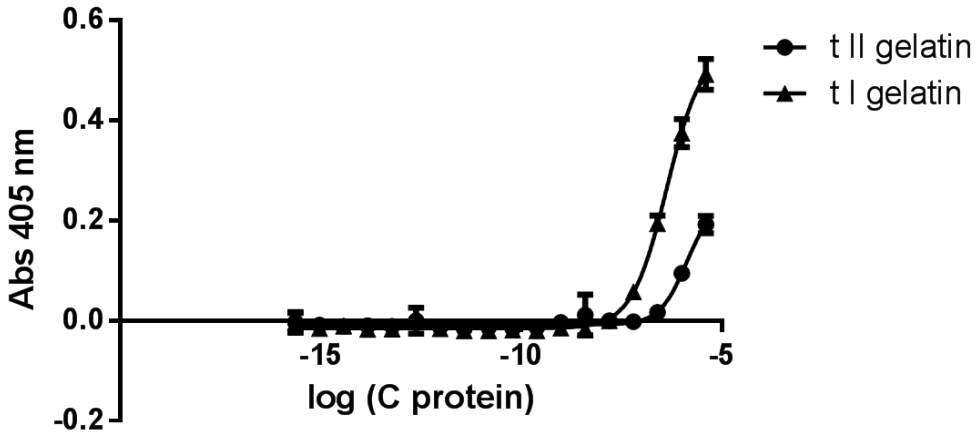


Figure 5-9: Binding of the refolded 111 to type I and type II gelatin plates. The binding assay relies on the use of plates coated with gelatin (thermodenatured collagen). The biotinylated protein was added over a concentration range before adding streptavidin-coupled alkaline phosphatase. Its substrate, namely PNPP, was added and converted into a coloured product which formation was measured by absorbance at 405 nm. Error bars represent standard deviations of three replicates.

5.3.3.4. NMR binding study of 222 on type II gelatin

^{1}H-^{15}N 2D HSQC spectra of the free 222 were first acquired (Figure 5-10 A). They confirmed the correct folding of the protein, even though some overlapping peaks were observed. This was due to the intrinsic nature of the protein, as it is made of three identical repeats. When overlapping this spectra with that of module 2 (Figure 5-10 B), we could observe that most of the peaks from the single module were found at the same position on the spectra of 222, although two to three peaks could be observed for some amino acids on the latter. For the ligand binding analysis, it was decided to transfer the assignments from the second module onto the spectra of 222. We therefore assumed that the cognate residues in all three modules of 222 were found at the same position on the spectra.

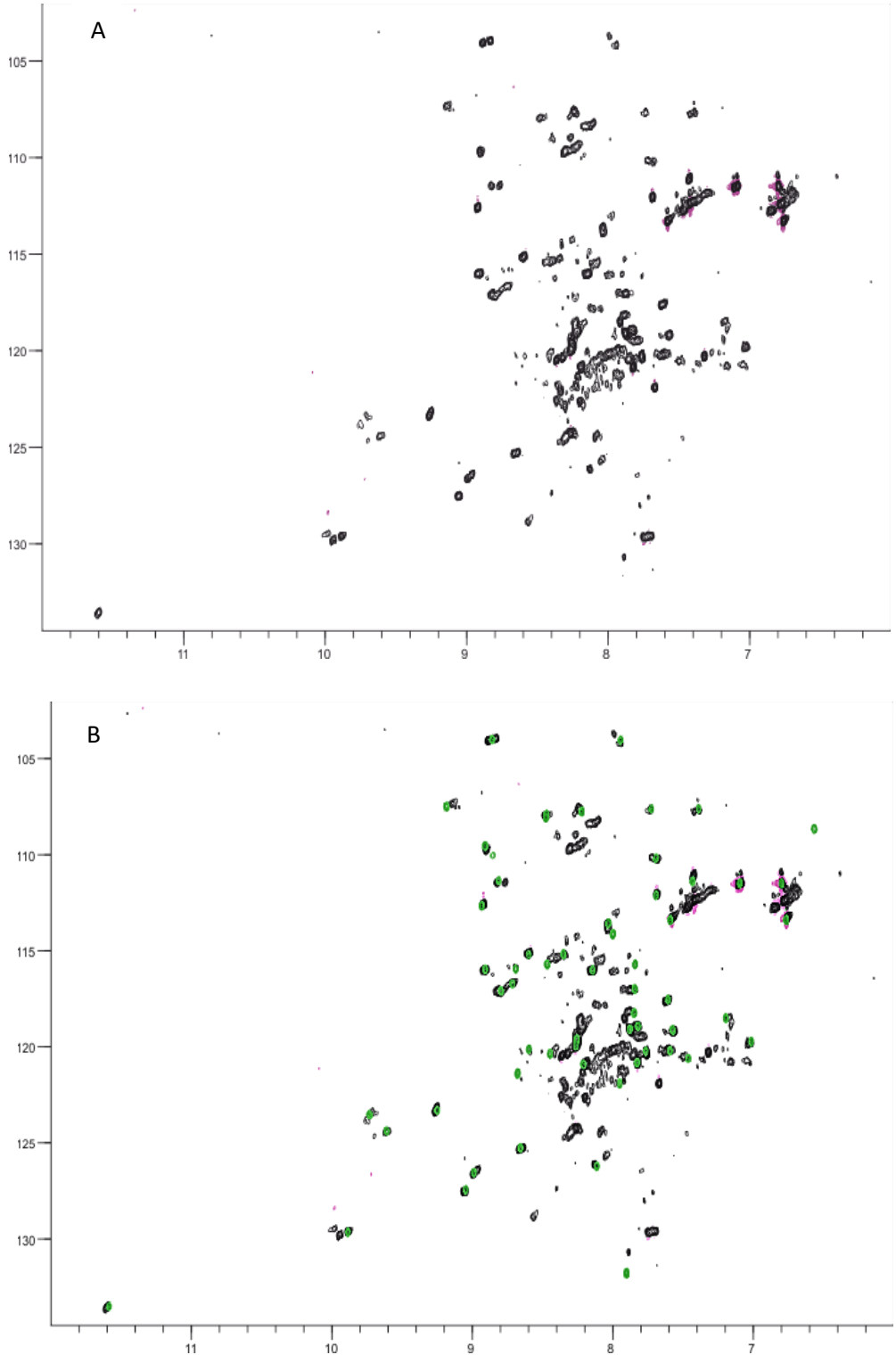


Figure 5-10: ^{1}H-^{15}N HSQC spectrum of the free 222 (A) and overlay of the spectra of the free 222 (black) and module 2 (green) (B) at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O, 50 µM. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton.

Type II gelatin was then added to 222 at a 40:1 and a 20:1 protein:gelatin ratio. Shifts could be seen in both cases but were clearer with the 20:1 ratio (Figure 5-11). For clarity, the spectra in this figure was divided into four quadrants which are zoomed in in Figures 5-12 to 5-15.

Residues which showed peak broadening and/or chemical shift changes corresponded to residues F280, M282, N285, A286, Q289, R296, F297, G299, G309, R310, G313, Y314, R315, W316, T319, E321, Y323, D326, G330, C332 and E334 in module 2. Residues G284, Y302, T306 and Y329, which were identified previously as being involved in the binding of module 2 to gelatin, also showed some small shifts in the spectra of 222. Overall, all of the residues identified to be involved in the binding of 222 to gelatin also showed similar chemical shift perturbations in the single module 2 bound to gelatin, or the CBD, except for A286 and E334.

It is, however, important to notice that it was impossible to determine if the residues corresponding to these amino acids were affected in all three modules of 222, nor to know which of the three modules was more affected than the other two. Residues have been, arbitrarily, mapped onto the second module of 222, as shown in Figure 5-16.

Overall, these NMR data confirm the specificity of the binding of 222 to our substrate of interest, and indicate that the binding occurs through the same mechanism as within the CBD.

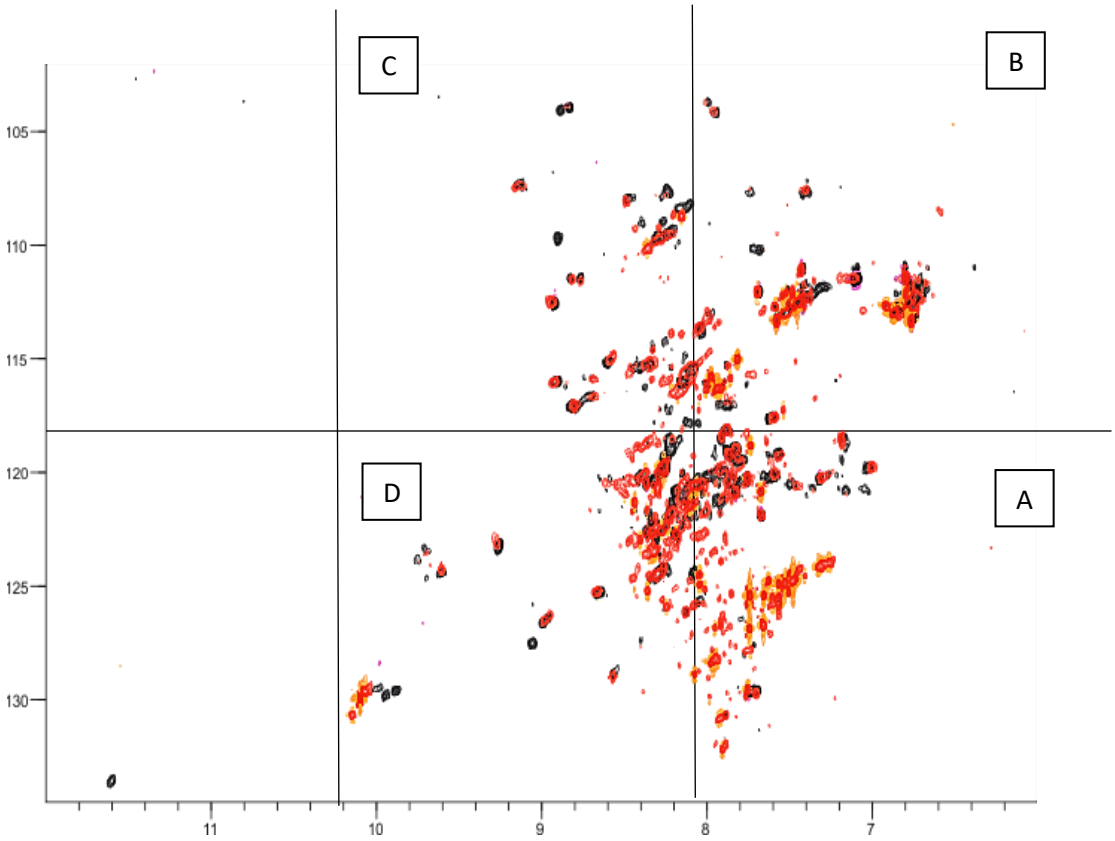


Figure 5-11: ^{1}H-^{15}N HSQC spectra of 222 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free 222 (black) and of 222 in the presence of type II gelatin (red) are superimposed. A 20:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in 222. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks clustered at around 8 ppm (1H dimension) are from the unfolded species of the protein.

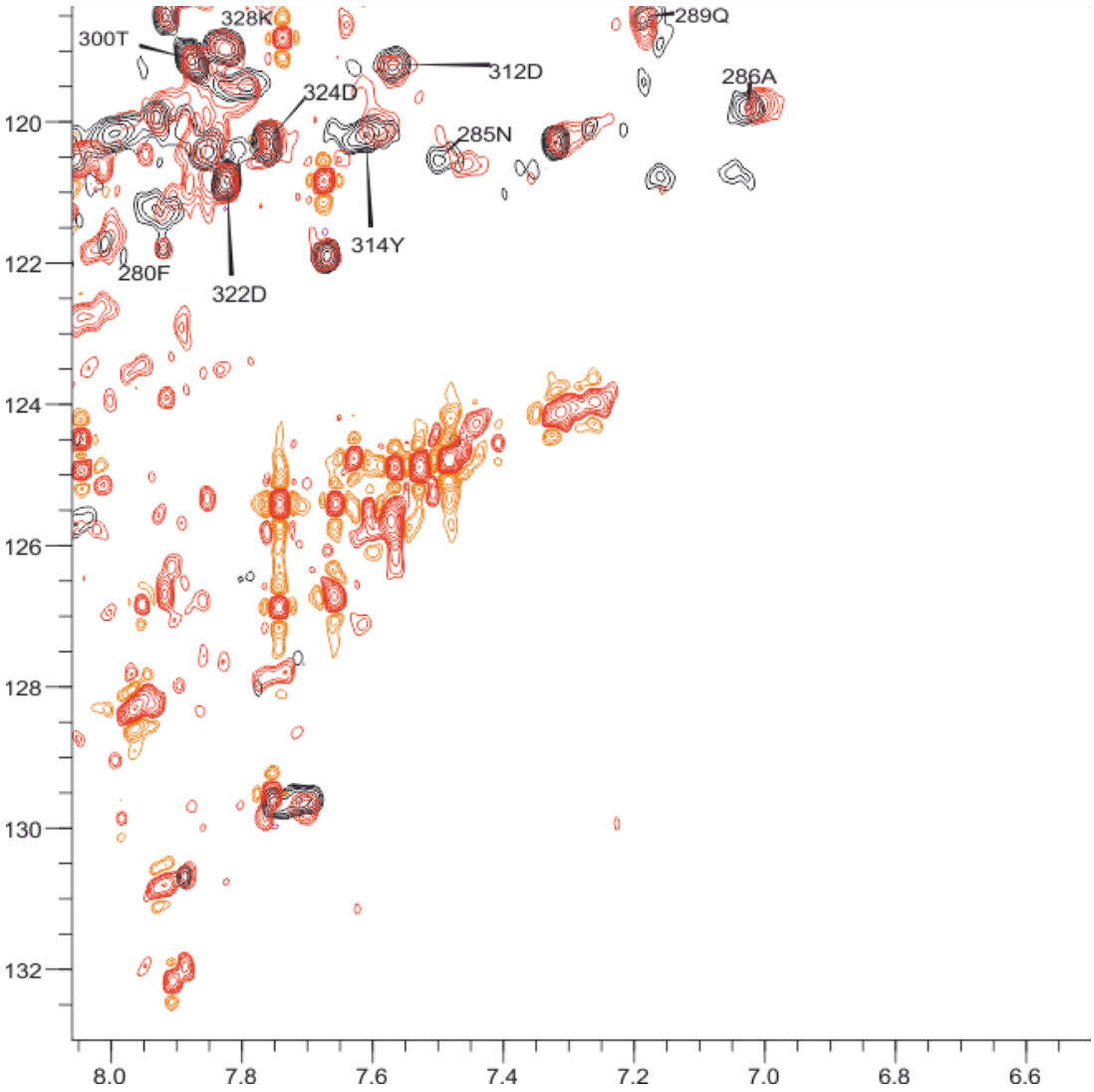


Figure 5-12: Part A of Figure 5-11. ^{1}H-^{15}N HSQC spectra of 222 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free 222 (black) and of 222 in the presence of type II gelatin (red) are superimposed. A 20:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in 222. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using the assigned spectra of module 2 and are labelled with their one-letter amino acid code and residue number.

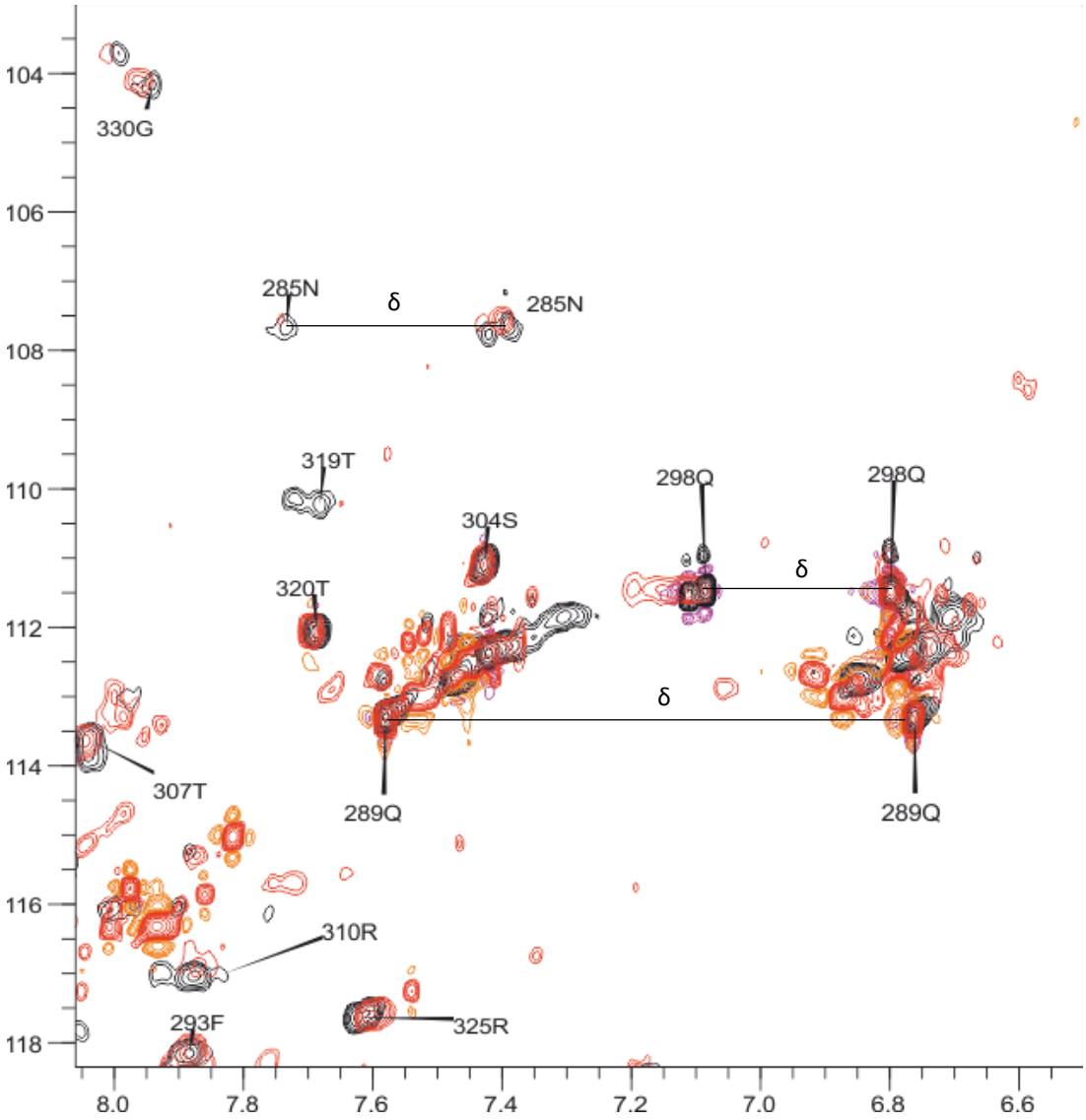


Figure 5-13: Part B of Figure 5-11. ^{1}H-^{15}N HSQC spectra of 222 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free 222 (black) and of 222 in the presence of type II gelatin (red) are superimposed. A 20:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in 222. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using the assigned spectra of module 2 and are labelled with their one-letter amino acid code and residue number.

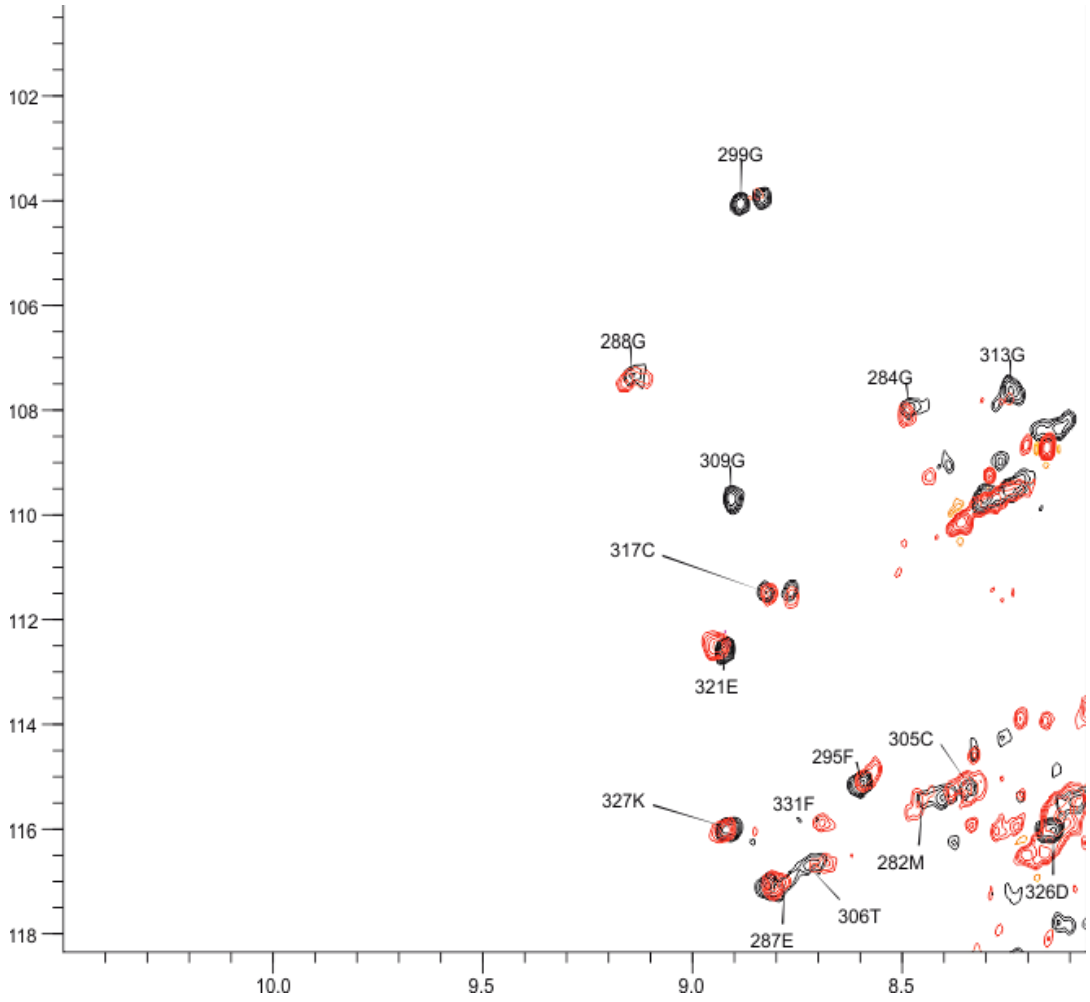


Figure 5-14: Part C of Figure 5-11. ^{1}H-^{15}N HSQC spectra of 222 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free 222 (black) and of 222 in the presence of type II gelatin (red) are superimposed. A 20:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in 222. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using the assigned spectra of module 2 and are labelled with their one-letter amino acid code and residue number.

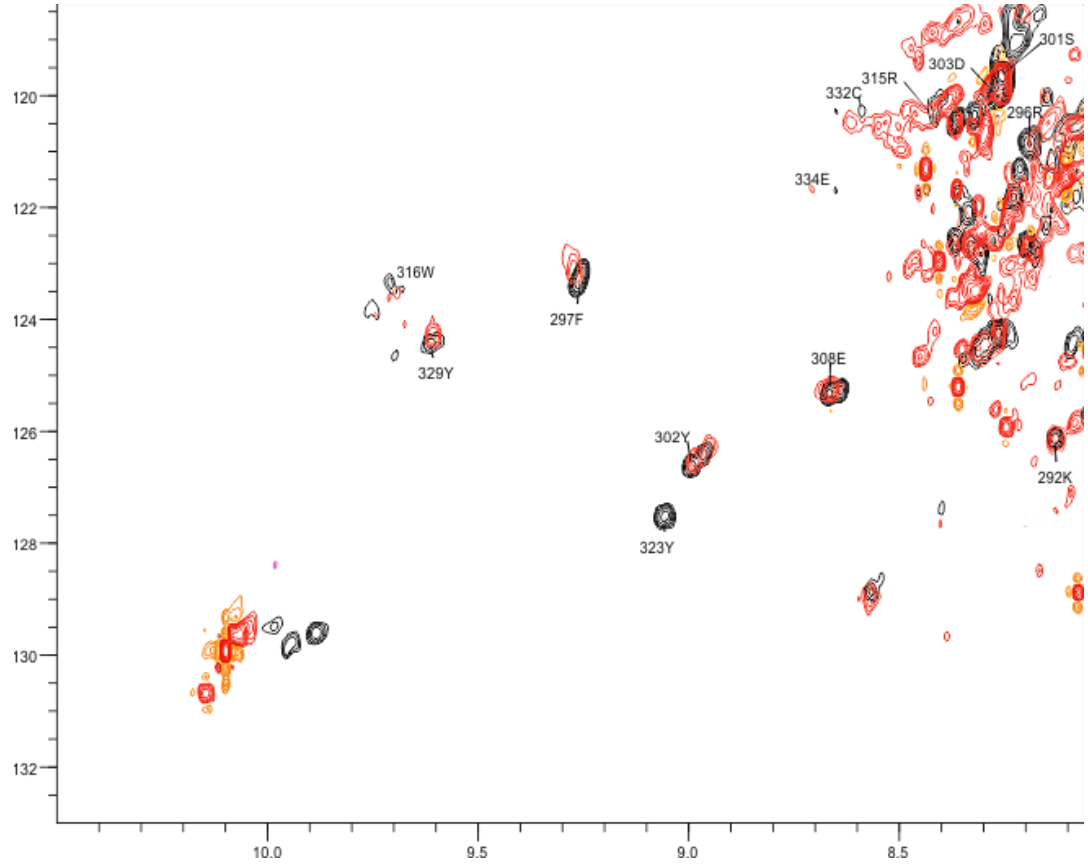


Figure 5-15: Part D of Figure 5-11. ^{1}H-^{15}N HSQC spectra of 222 before and after addition of type II gelatin at 298K in 25 mM sodium phosphate, pH 6.5, 10 % D_2O. The spectra of ligand-free 222 (black) and of 222 in the presence of type II gelatin (red) are superimposed. A 20:1 protein:gelatin molar ratio was used. The spectra were acquired on a 800Hz Bruker spectrometer instrument. Each peak on the spectrum corresponds to an amide proton in 222. Broadening of the peaks, leading to their disappearance, upon addition of the ligand indicates that the residues corresponding to these peaks are involved in the binding of the protein to the substrate. The peaks were assigned to their corresponding amino acid using the assigned spectra of module 2 and are labelled with their one-letter amino acid code and residue number.

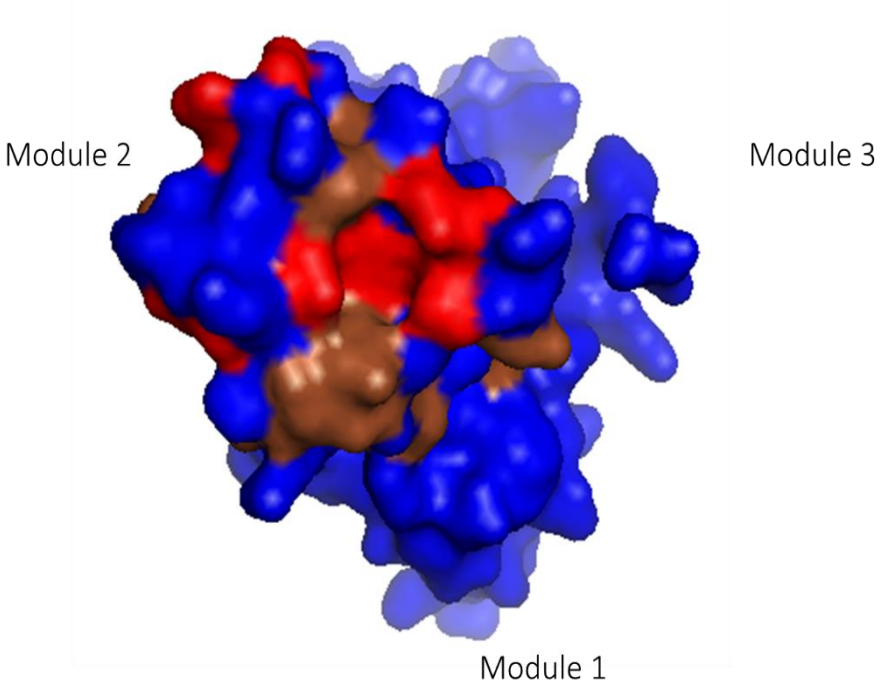


Figure 5-16: Three-dimensional surface representation of 222 showing the position of the residues that underwent peak broadening or shift changes in the presence of type II gelatin at a 20:1 protein:gelatin ratio. Residues undergoing peak broadening are indicated in red while residues undergoing shift changes are indicated in brown. Residues have arbitrarily been coloured on the second module of 222. The structure shown is derived from the reported crystal structure of MMP-2 (PDB Accession Number 1CK7). Residues from this structure were individually mutated in Pymol to match the sequence of 222.

5.4. Discussion

5.4.1. General conclusion

This chapter has described the generation of three unique proteins, namely the chimeric proteins 111, 222 and 333. They were carefully designed by ensuring that the residues involved in intramolecular interactions within the CBD were conserved in each construct. 222 and 333 were successfully expressed as soluble proteins in the cytoplasm of Shuffle cells, and purified as previously described. 111, however, was insoluble and had to be refolded, using a previously published method [141].

Although the DTNB assay indicated that all disulfide bridges were formed in all constructs, 111 was shown to be not properly folded by 1D NMR. The folding of this protein was investigated after the plate binding assay showed a very weak binding on both type I and type II gelatin.

222 and 333, on the other hand, showed a tight affinity in the plate binding assay as well as on the gelatin sepharose resin. As expected from the previous chapter, 333 was shown to bind better to type I than to type II gelatin. 222 was shown to bind with a slightly stronger affinity to type I gelatin than to type II, but, more importantly, displayed a dissociation constant which was nearly 14 times lower than that of the native CBD. The specificity of the binding of this protein to type II gelatin was further confirmed by NMR ligand binding studies, which showed that the residues involved in the binding were the same as those identified when studying module 2 alone or within the CBD.

Unfortunately, the impossibility of expressing 111 as a fully folded protein meant that we were not able to fully confirm the specificity of module 1 for type II gelatin, although the data on 333 do suggest that module 3 is more specific to type I.

5.4.2. Future work

Have there been other descriptions of gain of function mutations for the CBD?

To the best of our knowledge, this is the first time that mutational work has enabled an improvement in the affinity of the CBD of MMP-2 to one of its substrates. Most of the previously published mutational work performed on this protein were indeed alanine substitutions, which were done in an attempt to identify key residues involved in the binding [140,168].

However, one paper has reported mutations on the first module of the CBD of MMP-9 which improved its affinity [161]. In this paper, each module of the CBD9 as well as the full-length protein were fused to beta-galactosidase and their binding affinities to type I gelatin were measured. Module 2 was shown to provide most of the binding properties of the CBD, and to have a higher affinity for gelatin on its own than when combined to modules 1 and 3. The non-conserved residues within the boundaries of the conserved regions of modules 1 and 2 were then mutated. Residues A228 and A229 in module 1 were mutated for the cognate residues found in module 2 at these positions, *i.e.* K286 and P287. Similarly, a second construct was generated by mutating L253 and P254 in module 1 by the cognate residues Y311 and R312 found in the second module. Results showed that both of the mutants improved the binding of module 1 to gelatin, by nearly 8-fold and 6-fold, for the first and the second construct, respectively. However, even the first construct displayed an affinity which was still more than 3 times weaker than that of the second module of the CBD9, suggesting that other residues might be involved. Interestingly, these mutated residues are also non-conserved in each of the module of the CBD of MMP-2, as they correspond to residues 12 and 13 and 37 and 38 on Figure 5-17. In addition, residue 38 in module 3 corresponds to M373, an amino acid which we showed to be involved in the binding with type I gelatin only. This may indicate that these two regions are key for conferring specificity and affinity, and

performing mutations similar to those described in the CBD9 might provide further insights into the differences in affinity and specificity of each module of the CBD of MMP-2.

Are there any other mutations that could be done on the trimodular proteins to improve binding, specificity or solubility?

The insolubility of 111 was an issue that could not be solved, and the refolding procedure did not yield a properly folded protein. Module 1 alone was not more difficult to express than the other two, meaning that the insolubility is not due to the intrinsic sequence of the module, but rather due to difficulties in protein folding during protein expression.

Mutations could be designed in order to minimise insolubility, and to get a sequence which resembles more that of the original CBD. As the residues which are critical for the binding of module 1 to gelatin were identified in the previous chapter (showed in red and pink in Figure 5-17), one could decide to mutate only their cognate residues in modules 2 and 3, giving three modules with all the amino acids that are required for module 1 to bind.

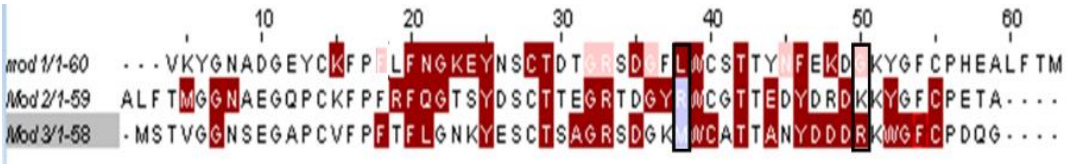


Figure 5-17: Alignment of modules 1, 2 and 3 of the CBD highlighting the residues involved in the binding to type I and type II gelatin. Residues common to both substrates are highlighted in red, while residues showing higher shifts for type II gelatin are in pink, and those showing higher shifts for type I gelatin in grey. G269 is at position 50 on this figure, and M373 at position 38, as shown by the black rectangles.

Another potential option would exploit the fact that we identified G269 in module 1 (corresponding to G50 in Figure 5-17) as a potential reason for the specificity of this module for type II gelatin. It would therefore be interesting to mutate the cognate residue of G269 in the other modules, in order to see if this would confer specificity for type II gelatin over type I. This would involve mutating K327 in module 2 and R385 in module 3 for a glycine. This mutation could also be performed on the 3 residues equivalent to K327 in 222, *i.e* K269, K327 and K385 as outlined in the Figure below (Figure 5-18).

E G Q V V F T M Y G N A E G Q P C K F P F R F Q G T S Y D S C T T E G R T
D G Y R W C G T T E D Y D R D K K Y G F C P H E A L F T M G G N A E G Q
P C K F P F R F Q G T S Y D S C T T E G R T D G Y R W C G T T E D Y D R D
K K Y G F C P E T A L F T M G G N A E G Q P C K F P F R F Q G T S Y D S C T
T E G R T D G Y R W C G T T E D Y D R D K K Y G F C P D Q G Y S L

Figure 5-18: Sequence of 222 highlighting, in red, the position of the lysines 269, 327 and 385 which could be mutated for a glycine.

Overall, 222 was the protein with the highest affinity for type II gelatin, and this was the protein progressed forward and used to coat MSCs in order to achieve the objective of increasing their binding to damaged cartilage in osteoarthritis.

6. Development of a new method for coating MSCs with 222

The previous chapter described the generation of a new protein, namely 222, with a high affinity for type II gelatin. The next phase of the project was to test whether coating MSCs with this protein would enhance their adhesion onto this substrate. Selecting the best method for coating these cells with 222 was therefore necessary.

6.1. Introduction: existing methods for cell surface engineering of MSCs

Cell therapy is a promising approach which relies on the delivery of cells to the target tissue. This is often not achieved, especially for MSCs [127] . When injected systemically, these cells are often trapped in the lungs [189], a phenomenon which is thought to be due to the strong adhesive properties of MSCs as well as to the small size of the capillaries [120]. Local delivery directly into the tissue of interest by-passes the lung barrier, but most of the time involves invasive surgery [189]. In addition, MSCs directly injected at the site of damage do not show high retention, with only 1 % of injected cells found in the heart following myocardial infarction [190].

Methods are therefore being developed to try to improve the efficiency of the delivery of MSCs to a damaged tissue. Achieving this goal could improve their therapeutic effects, but also reduce the required cell dose as well as the secondary effects due to their off-target accumulation [191].

Cell surface modification is one way to do so. It relies on coating the surface of MSCs with various proteins whose ligands are found in the tissue of interest. Three main ways can be used to achieve this coating: covalent coupling to proteins found at the cell surface, charge modification, or hydrophobic interaction of the protein of interest with the cell membrane [190]. The latter seems to be the most popular, as it reduces the risks of cytotoxicity and avoids permanent modification [190,192].

6.1.1. Palmitated Protein G (PPG)

One common method exploiting hydrophobic interactions relies on the use of a Palmitated Protein G (PPG), which intercalates into the cell membrane. The protein of interest, either an antibody or a protein fused to a Fc region, can then be bound to the PPG anchor, leaving the antigen binding site exposed outwards of the cell surface [127]. This method has been used to treat different conditions, such as osteoarthritis [127], inflammatory bowel disease [193], or cardiac diseases [194].

In all cases, the cells were successfully coated, and the modification was stable for a minimum of 3.4 hours [193] and up to more than 24 hours [194].

The coating method did not affect the cell viability nor the proliferation properties of cells [127,194]. The differentiation potential of the progenitors was also not affected in the work reported by Dennis et al. [127]. Fluorescent-labelled chonroprogenitors showed higher adhesion to a rabbit articular cartilage explant with a partial thickness defect when coated with antibodies directed against matrix molecules. A similar observation was made in the paper by Whitaker et al. [194], where 28 % of the infused functionalized Cardiosphere-Derived Cells (CDCs) were retained at the site of injection after 3h, due to their binding to P-selectins.

Even though the use of PPG has been proven to be efficient, the use of antibodies or fusion proteins can be cost-prohibitive. Other methods have therefore been developed to exploit the hydrophobic interaction of the protein of interest with the cell membrane.

6.1.2. Palmitated peptides

In two very similar papers [191,195], peptides of interest have been synthetically coupled to palmitic acids to provide a membrane anchor. Peptides were screened for their ability to bind to substrates found in either the ischemic heart or brain. MSCs were successfully coated with the palmitated peptides, and no cytotoxicity nor modification of their proliferation properties were observed. The coating was stable for periods from 3.5 up to 24 hours [191,195]

which was enough to enhance the engraftment of injected cells in the targeted tissues, although a clear quantification is lacking in both cases. In the paper by Huang et al.[191], however, the increased number of cells proved to have clinical benefits, as a reduction in brain deficits and a differentiation in neural cells were observed. However, it has to be noted that these MSCs were first transfected with the microRNA miR-133b, which could explain these observations. Although this method has shown some success, the coupling to palmitic acid seems to be limiting this methodology to the use of peptides as coating agents.

6.1.3. Polyethylene glycol (PEG)

The use of PEG as a membrane anchor has been shown to be successful for coating MSCs with both peptides [196] and proteins [190]. In the first case, MSCs were coated with PEG coupled to three different heparin binding peptides in order to prevent thromboinflammation by binding to heparin sulfate. Cells were successfully coated for up to 25 hours, and shown to bind to heparin sulfate. Their differentiation potential was not affected, and the method seemed to not be toxic, except for one of the peptides used. Coated cells reduced markers of complement activation and suppressed coagulation and platelet aggregation, showing the expected gain-of function from their cell surface modification.

In the second paper [190], MSCs were coated with the protein CXCR4 coupled to a lipid-PEG, in order to enhance their migration to the ischemic myocardium by exploiting the affinity of the exposed protein for SDF-1, which is upregulated in myocardial infarction for only 48 hours. The coating of cells with CXCR4 did not alter their proliferation and adhesion properties, and was not cytotoxic. The membrane-bound CXCR4, which was stably exposed at the surface for up to 180 minutes, was shown to be able to bind to SDF-1 by confocal microscopy, an interaction which favoured the migration of the coated cells towards a SDF-1 gradient.

6.1.4. Wheat Germ Agglutinin (WGA)

A similar approach was adopted by Kim et al. [189], except that the anchor to the protein membrane was Wheat Germ Agglutinin (WGA). It was covalently coupled to hyaluronate (HA) to favour the migration of coated MSCs to the liver, which expressed HA-binding partners. Again, the coating was shown to be non-cytotoxic, and was stable for one hour at 37°C, with no alterations of the cell properties. The exposure of HA at the cell surface was enough to direct them to the liver, while cells coated with WGA only remained trapped in the lungs [189].

6.1.5. Covalent modification

All of the previous cell surface modifications relied on hydrophobic interactions between the protein of interest and the cell membrane. Although most of them have shown, if not clinical benefits, at least an efficient targeting of cells to the tissue of interest, longer modification might be needed to improve the efficacy of MSCs treatment even further [192].

This has been explored by Takayama et al. [192], using a covalent modification of the cell surface by exploiting the avidin-biotin complex method. The method described in this paper used two fluorescent model proteins, namely Green Fluorescent Protein (GFP) and Nluciferase. The procedure relied on first biotinylating the amines of the proteins naturally present at the cell surface, prior to adding avidin. The biotinylated GFP or Nluc proteins were then added and could react with the avidinated cells. The covalent coupling did not affect the cell viability nor the differentiation potential, and was stable for 14 days, achieving a much longer modification than with non-covalent methods. Much more GFP-positive cells were obtained with the coating method compared to transfection (95 % against 7.3 %, respectively), which highlights the benefit of adopting a cell surface modification over a genetic one. Finally, coated cells injected into mice were still detected after 7 days, showing the persistence of the labelling *in vivo*.

6.1.6. Surfactant corona method

An ideal balance would be found if a non-covalent modification method was providing long-lasting effects. Both goals have been achieved using a new method [197], in which the protein is first cationised through the conversion of its aspartate and glutamate residues into synthetic amino acids containing amine side-chains. The cationised protein can then be electrostatically coupled to a negatively charged, PEG-based surfactant, which surrounds the protein like a corona. The conjugated protein can then be delivered into the cell membrane, with the surfactant acting as the membrane anchor. The method was first described with GFP and myoglobin. Coating MSCs with myoglobin led to the reduction of necrosis within cartilage constructs, due to the recruitment of oxygen at their centre by the exposed myoglobin [197].

This method not only is simpler than those described earlier, but also enables the exposure of the protein of interest at the cell surface for longer periods of time. It was therefore decided to use this surfactant-corona method for coating MSCs with 222 in an attempt to improve their binding to type II gelatin and, later, to damaged cartilage.

6.2. Materials and Methods

6.2.1. Protein conjugation

A FITC tag was added, when appropriate, to the 222 protein according to the manufacturer's instructions, giving FITC-222.

Protein samples were buffer exchanged against either 20 mM HEPES, pH 6.5 or 20 mM sodium phosphate (NaP), pH 6.5 using a PD 10 Sephadex G-25 M column. The protein was then concentrated up to 0.7 mg/mL before proceeding to the cationisation.

N,N`-dimethyl-1,3-propanediamine (DMPA) and N-(3-Dimethylaminopropyl)-N`-ethylcarbodiimide hydrochloride (EDC) were added to the protein samples and the pH

adjusted to 6.5. Samples were incubated for 20 to 24 hours at 4°C with stirring, before being purified from the excess reagents by spinning 5 times through Amicon Ultra 15 centricons with a 10 kDa Molecular Weight Cut-Off (MWCO). The negatively charged surfactant glycolic acid ethoxylate 4-nonylphenyl ether was then added to the cationised protein and incubated overnight at 4°C with stirring, giving the proteins Surfactant-222 or Surfactant-FITC-222.

6.2.2. Dynamic Light Scattering (DLS)

Particle size and zeta potential analysis were performed using a ZetaSizer Nano ZS (Malvern Instruments, UK) and ZetaSizer software (Malvern Instruments, UK). Samples were filtered through a 0.22 µm filter prior to the analysis. Zeta potential measurements were made after a 120-second equilibration with 10-100 runs per sample. Size measurements were made after a 120-second equilibration, with the backscatter measured at 173°. The values reported in this report correspond to the radius expressed in number of the total sample. Each zeta potential and DLS measurements were performed in triplicates.

6.2.3. Coating of adherent MSCs with Surfactant-FITC-222

The media from cells cultivated in T25 flasks was removed, and cells were treated with 18 µM of Surfactant-FITC-222 for 30 min at 37°C. The protein solution was then removed and cells were washed with 0.04 mg/mL heparin ammonium salt resuspended in PBS. Phenol-free DMEM medium was then added prior to imaging with a confocal microscope.

6.2.4. Coating MSCs in suspension with Surfactant-FITC-222

Cells cultivated in DMEM medium were trypsinised, counted, and resuspended at 1 million cells/mL.

The cells were then spun down before being resuspended in 300 µL of 5 µM Surfactant-222 or Surfactant-FITC-222, while control samples were incubated with 300 µL of native 222. Cells were incubated for 30 minutes at 37°C with shaking and manually agitated every 10 minutes to prevent cell aggregation.

Cells were then spun down and washed with 300 µL of 0.04 mg/mL heparin ammonium salt resuspended in PBS before being spun down again and resuspended in 1 mL of imaging medium.

6.2.5. Cell adhesion assay

MSCs were coated with 5 µM of either Surfactant-222 or Surfactant-FITC-222, and resuspended in DMEM media containing P/S and Glutamax only. Cells were seeded on either non-treated tissue culture plates, or gelatin-coated tissue culture plates (0.5 µg type II gelatin/well) and serially diluted over the plate. 4 replicates were used for each condition, and Wild Type (WT) cells were used as a control. Cells were cultivated for 24 hours and washed once with phenol-free DMEM before being imaged with an epifluorescent microscope following their staining with the membrane dye FM-4-64.

6.3. Results

6.3.1. Protein conjugation

The aim of the conjugation, which was first described by Armstrong et al. [197] is to surround the protein with a surfactant corona, enabling the protein to be incorporated into the membrane of stem cells via hydrophobic interactions, as illustrated in Figure 6-1.

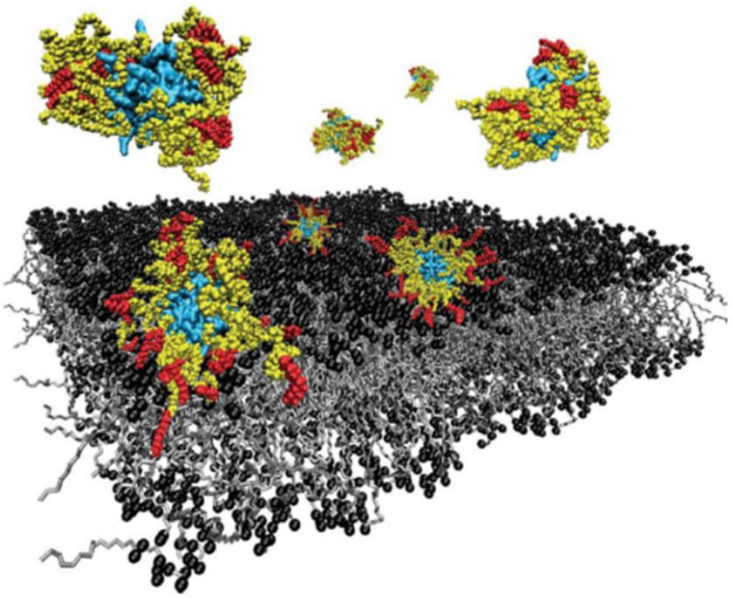


Figure 6-1: Anchoring of a surfactant-protein complex onto the cell membrane. The surfactant, represented in yellow and red, surrounds the protein, here represented in blue, creating a corona. In solution, the hydrophobic non-phenyl chains of the surfactant, in red, are buried by the PEG chains. At the contact with the hydrophobic membrane, a structural reorganisation occurs, enabling the alkyl chains to be exposed and to anchor the protein at the cell surface. Figure from [197].

The protein modification is a two-step process, with an initial cationisation following the surfactant addition. The cationisation involves the covalent coupling of N,N'-dimethyl-1,3-propanediamine (DMPA) to the carboxylic residues of the protein (Figure 6-2, **(3 and 4)**). Those carboxylic moieties first need to be activated by a carbodiimide (Figure 6-2, **(1 and 2)**), namely, N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC).

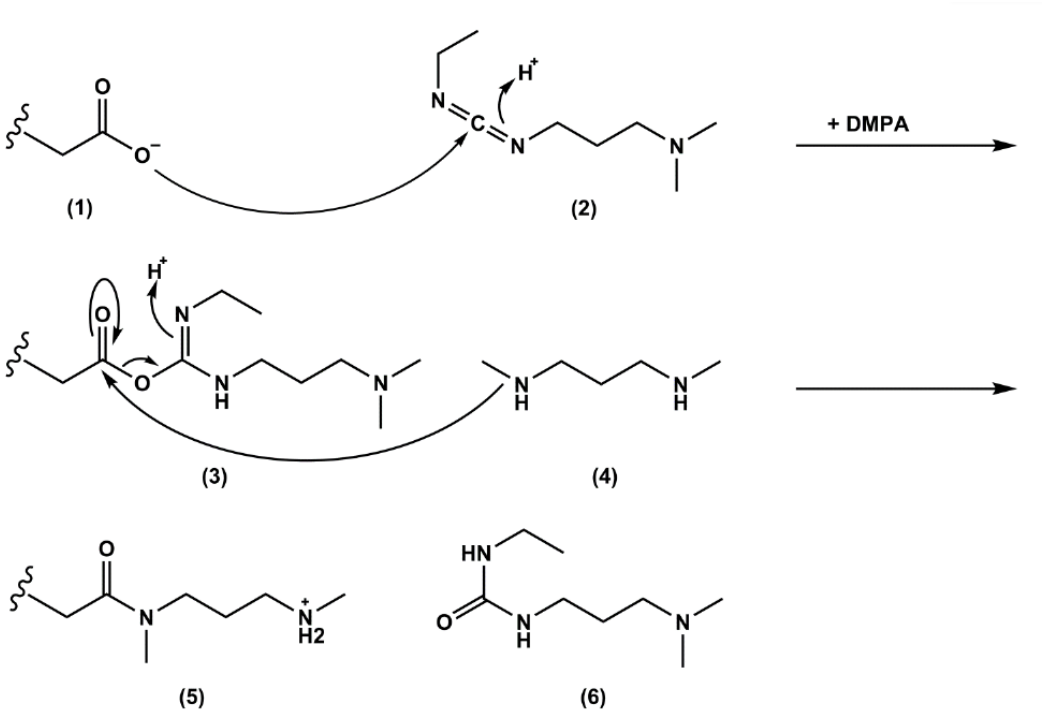


Figure 6-2: Chemical reaction describing the covalent coupling of DMPA to the carboxylic groups of amino acids [153]. The addition of DMPA enables the conversion of negatively charged residues, *i.e.* aspartate and glutamate **(1)**, into synthetic amino acids containing amine side-chains **(5)**. An isourea byproduct is also formed **(6).**

Following covalent coupling with DMPA, the cationised protein can then interact electrostatically with the negatively charged surfactant glycolic acid ethoxylate 4-nonylphenyl ether, which is obtained from the oxidation of Igepal CO-890, as shown in Figure 6-3.

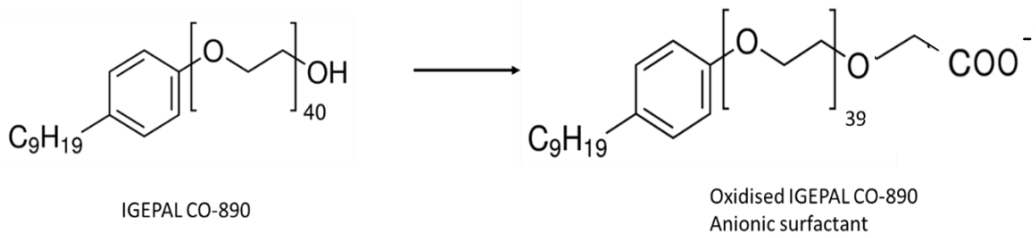


Figure 6-3: Chemical reaction describing the oxidation of IGEPAL CO-890 into the anionic surfactant used for the conjugation. This synthesis was performed by an external collaborator.

This process was applied to both 222 and FITC-222 in either 20 mM HEPES buffer, pH 6.5 or 20 mM sodium phosphate (NaP), pH 6.5, which were each used for different proteins in the paper published by Armstrong et al [197]. The two-step reaction was followed using Dynamic Light Scattering (DLS), as it enables the measurement of both the hydrodynamic radius and the zeta potential of a protein. As presented in Tables 6-1 and 6-2, the measurements were very similar for both 222 and FITC-222, suggesting that the presence of the fluorescent tag did not interfere with the process. In both buffers, the cationisation induced an increase in the charge of the protein and only a small change in the hydrodynamic radius. The fact that this radius could be slightly reduced after cationisation could come from a compaction of the molecule following the change in charge. The hydrodynamic radius was then increased following the addition of the surfactant, confirming the creation of a corona surrounding the protein. The surfactant addition also caused a neutralization of the charge which had been increased after the cationisation. Interestingly, the fold change was higher for both the charge and the size in HEPES compared to sodium phosphate. If it is known that the zeta potential is dependent on the dielectric constant of the buffer used for the measurement, it is unclear to us as to why the buffer would also have an influence on the hydrodynamic radius of the protein.

Table 6-1: DLS measurements of the hydrodynamic radius and the zeta potential of 222 in both sodium phosphate (NaP) and HEPES buffers. 222 was cationised and conjugated to the surfactant as previously described, in either 20 mM HEPES, pH 6.5 or 20 mM NaP, pH 6.5. A ZetaSizer instrument was used to measure both the hydrodynamic radius (size) and the zeta potential (charge) on each version of the protein. Each measurement was performed in triplicates, and the values reported below are their average.

	20 mM HEPES, pH 6.5		20 mM NaP, pH 6.5	
222	Size	Charge	Size	Charge
Native	3.26	-12.13	3.92	-2.26
Cationised	2.19	14.53	4.74	6.53
Surfactant	12.63	-2.31	6.89	-1.45

Table 6-2: DLS measurements of the hydrodynamic radius and the zeta potential of FITC-222 in both sodium phosphate (NaP) and HEPES buffers. 222 was cationised and conjugated to the surfactant as previously described, in either 20 mM HEPES, pH 6.5 or 20 mM NaP, pH 6.5. A ZetaSizer instrument was used to measure both the hydrodynamic radius (size) and the zeta potential (charge) on each version of the protein. Each measurement was performed in triplicates, and the values reported below are their average.

	20 mM HEPES, pH 6.5		20 mM NaP, pH 6.5	
FITC-222	Size	Charge	Size	Charge
Native	3.133	-15.133	3.652	-1.308
Cationised	3.179	9.110	4.721	3.890
Surfactant	13.42	-4.817	7.352	-0.588

6.3.2. Coating of adherent cells

The Surfactant-FITC-222 protein was then used at 18 µM in 20 mM HEPES, pH 6.5 to coat MSCs seeded onto plastic, as described in the original method [197]. Cells were imaged by confocal microscopy, which confirmed the presence of the protein at the cell surface (Figure 6-4).

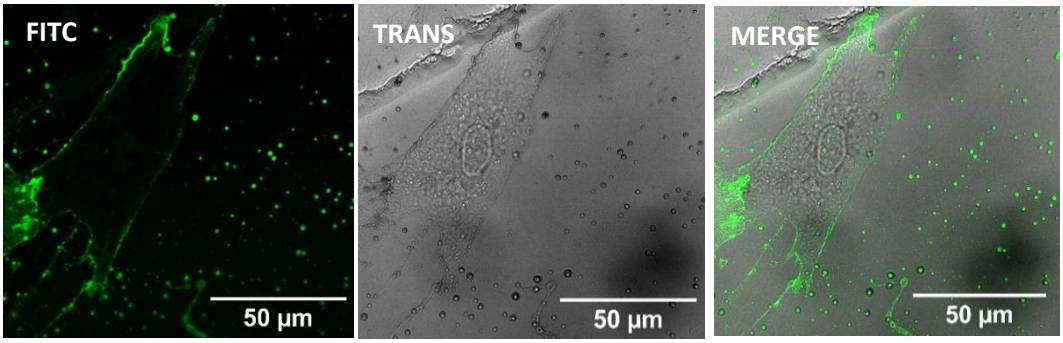


Figure 6-4: Confocal microscopy image showing the successful coating of MSCs attached to plastic with Surfactant-FITC-222. Cells were treated with 18 µM of Surfactant-FITC-222 for 30 min at 37°C before being washed with heparin ammonium salt. Phenol-free DMEM was added to the cells for imaging. Both the FITC and transmitted channels were used to acquire images.

Overall, this experiment validated the method developed by Armstrong et al. when applied to our new protein 222. However, several issues were encountered with this methodology: the amount of protein required to coat attached cells was very high, and was not normalised to the number of cells but to a surface area. Finally, it would be preferable for our project to be able to inject cells immediately after their coating, with no trypsinisation step, as such a step could, in theory, alter the integrity of the membrane and potentially remove the newly coated 222 from the cell surface. It was therefore decided that adapting the coating method for cells in suspension would be better for the purpose of this project.

6.3.3. Coating of cells in suspension

The new method, illustrated in Figure 6-5, involved trypsinising cells prior to incubating them with 300 µL of protein/million cell. Cells were incubated with the protein for 30 min with shaking, before being washed with the same volume of heparin ammonium salt and resuspended in phenol-free media.

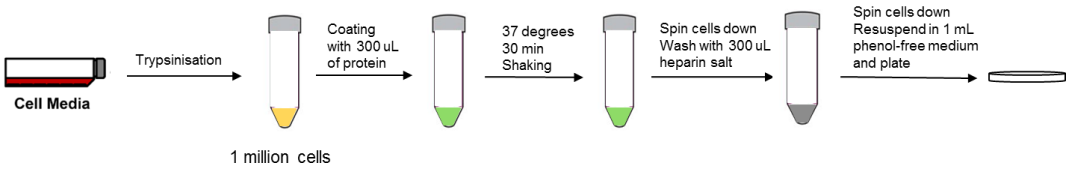
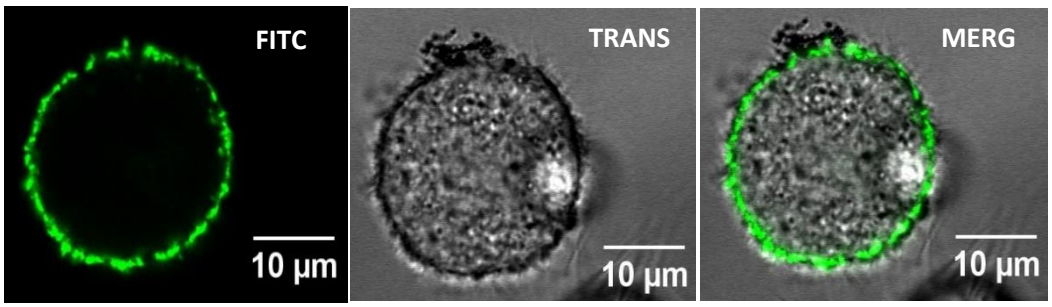


Figure 6-5: Schematic representation of the method for coating cells in suspension. Cells were trypsinised and 1 million cells were spun down before being resuspended in 300 µL of either Surfactant-222 or Surfactant-FITC-222. Various concentrations of protein were tested in the initial experiments. Cells were incubated with the protein for 30 minutes at 37°C with shaking. Cells were then spun down, washed in heparin ammoniun salt and resuspended in phenol-free medium for imaging.

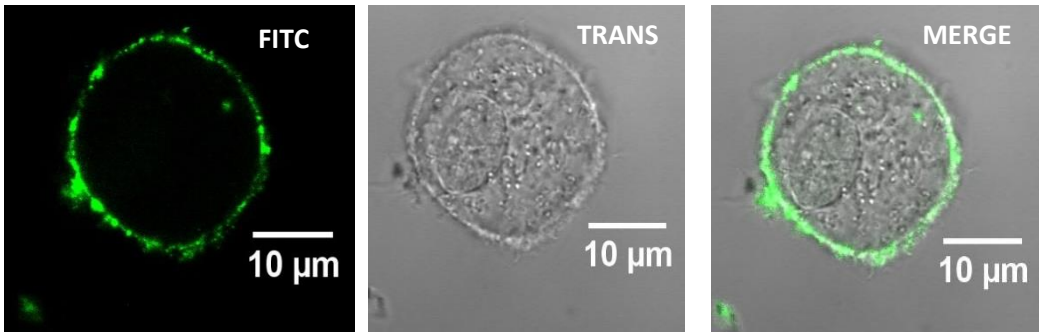
The methodology for coating cells in plate described the use of 18 µM of protein [197], but we aimed to reduce the concentration even further in order to improve the feasibility of this new method. It was therefore decided to test a range of concentrations, from 10 to 1.25 µM, all in HEPES buffer. Results showed successful coating at all concentrations tested, except for 1.25 µM, which gave a very weak signal (not shown). Overall, using a protein concentration of 5 µM seemed to give the best balance between intensity of signal and required amount of protein, and it was therefore decided to further optimise the coating with this condition.

Cells coated with 5 µM of Surfactant-FITC-222 in both NaP and HEPES were therefore imaged by confocal microscopy. The images confirmed the successful coating in both HEPES and NaP buffers (Figure 6-6, A and B, respectively), with a very similar staining pattern and intensity observed in both buffers when looking at a single cell. No signal was observed when incubating cells with the native FITC-222 protein (Figure 6-6, C), confirming that the observed fluorescence was due to the coating with Surfactant-FITC-222.

A



B



C

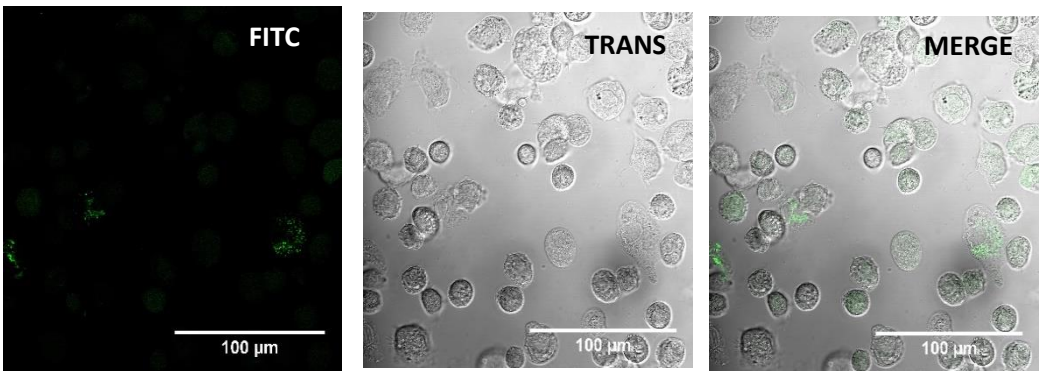


Figure 6-6: Confocal microscopy images showing the successful coating of MSCs in suspension with Surfactant-FITC-222. Cells were treated with 5 µM of Surfactant-FITC-222 for 30 min at 37°C in either 20 mM HEPES (A) or 20 mM NaP (B). Cells were also treated with Native-FITC-222 as a control (C, in NaP). Cells were then washed with heparin ammonium salt, and phenol-free DMEM was added to the cells for imaging. Both the FITC and transmitted channels were used to acquire images.

The protein was also shown to be co-localised with the membrane dye FM 4-64 (Figure 6-7), confirming its presence at the cell surface. However, some cells showed an internalisation of the dye, indicative of cell death (Figure 6-7).

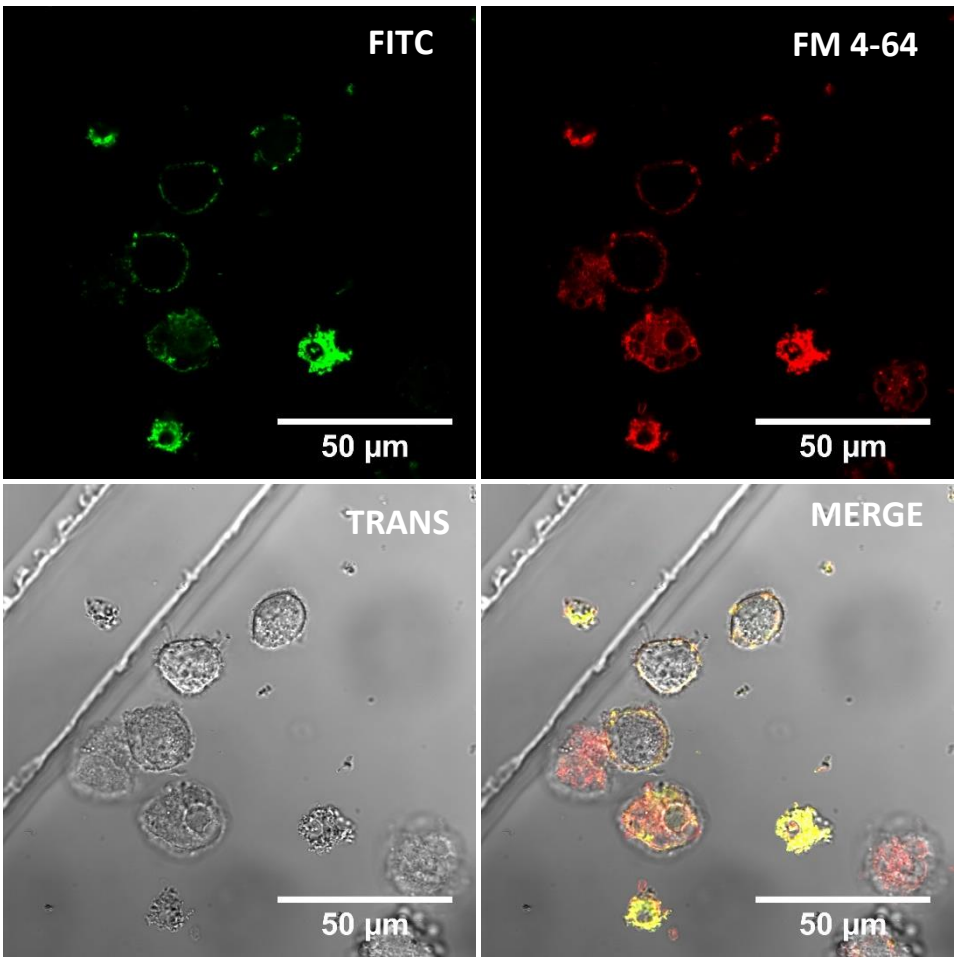
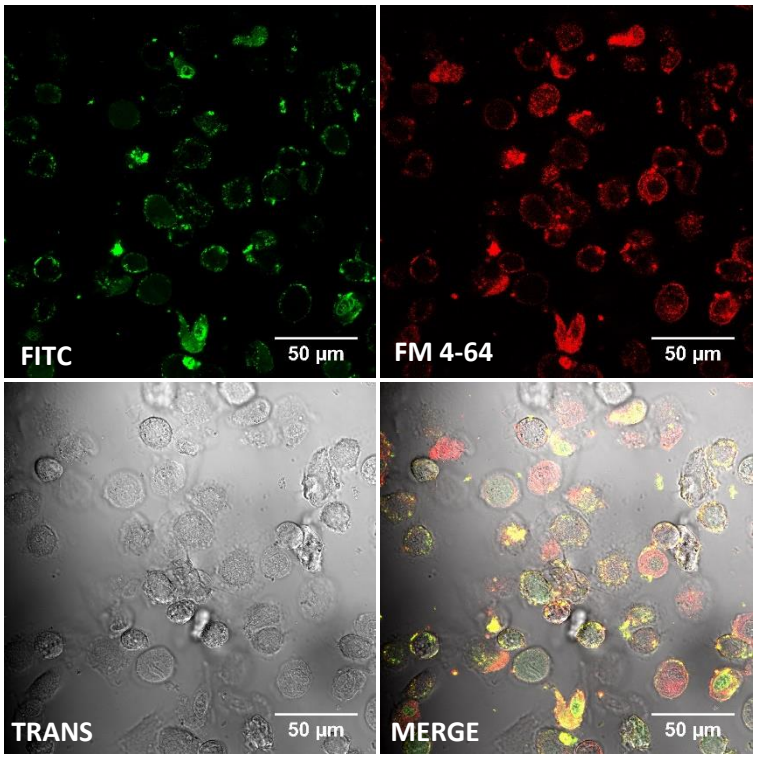


Figure 6-7: Confocal microscopy image showing the co-localisation of Surfactant-FITC-222 and the membrane dye FM 4-64. Cells were coated as previously described with 10 µM of Surfactant-FITC-222 and imaged with a confocal microscope. FM 4-64 was added at a 1: 1000 dilution to label the cell membrane. The merged image shows a yellow signal, indicative of the superposition of the signals coming from the FITC tag and the membrane dye. Some cells show internalisation of the dye, indicative of cell death.

When acquiring images with a lower magnification, we observed a much higher proportion of cells showing internalisation of the dye in HEPES than in NaP, suggesting that the latter might be better tolerated by MSCs, and that its use might result in a higher cell viability (Figure 6-8).

A



B

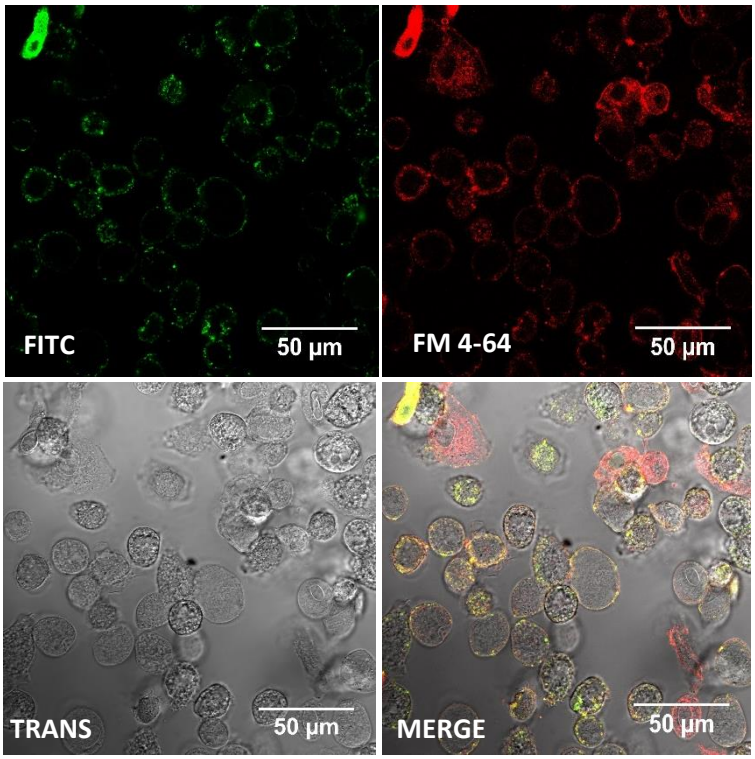


Figure 6-8: Confocal microscopy images showing the difference in coating for cells treated in 20 mM HEPES buffer (A) or 20 mM NaP buffer (B). Cells were coated as previously described with 5 µM of Surfactant-FITC-222 and imaged with a confocal microscope. FM 4-64 was added at a 1: 1000 dilution to label the cell membrane. The merged images show a yellow signal, indicative of the superposition of the signals coming from the FITC and the membrane dye. Some cells show internalisation of the dye, indicative of cell death.

It was then decided to investigate this observation further, by monitoring the viability of cells after coating with a trypan blue assay (Figure 6-9). Results showed that, even though each buffer on their own was not toxic for MSCs, coating MSCs with 5 µM of 222 in HEPES had deleterious effects on the cell viability, giving an average of 47.4 % of viable cells, compared to nearly 83 % in sodium phosphate. The difference in viability between the Surfactant-222 treated cells in HEPES and those left in DMEM was significant, while all the other treatments did not show any significant difference when compared to the DMEM sample.

The reason behind the difference in viability between cells treated with Surfactant-222 in HEPES and sodium phosphate is unclear. One hypothesis could be that coating MSCs weakens their membrane, leading to a leakage of the surrounding buffer into the cells. Sodium phosphate, which is more biologically relevant than HEPES, might be better tolerated in this process.

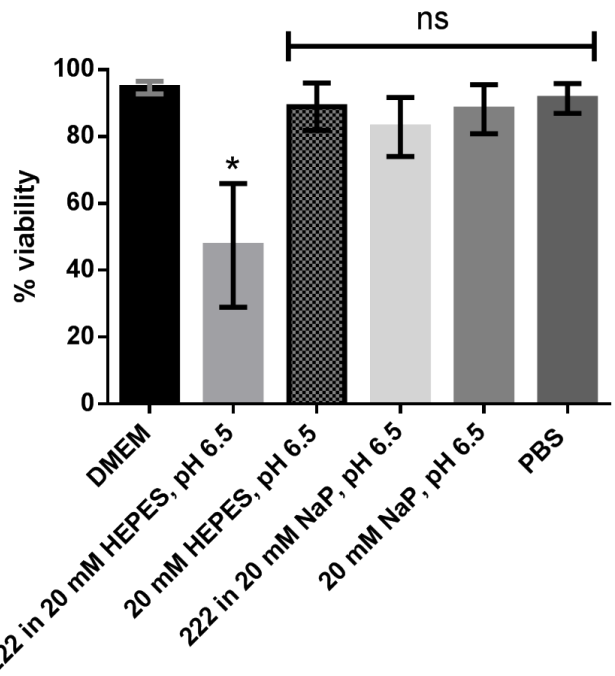


Figure 6-9: Viability of coated MSCs. The viability of cells was measured after coating with a trypan blue assay. MSCs were also incubated with either of the buffers, PBS and DMEM. Viability measurements were performed on at least 4 biological replicates for each sample. Error bars indicate standard deviation. ns: non significant, *: $p < 0.05$ compared with DMEM only, by Kruskall Wallis ANOVA with a Dunn post-hoc correction for multiple comparisons.

Overall, these results showed that using 5 µM of Surfactant-222 in 20 mM NaP, pH 6.5 were the optimal conditions for coating MSCs in suspension. The next step was to test whether this coating would improve the binding of MSCs to type II gelatin.

6.3.4. Cell adhesion assay

MSCs were then coated with either Surfactant-FITC-222 or Surfactant-222 and their binding to type II gelatin was monitored, and compared to that of Wild-Type (WT) cells. Cells were stained with the membrane dye FM 4-64 and imaged with an epifluorescent microscope.

In a first preliminary experiment, the coating was confirmed to be successful on the cells treated with Surfactant-FITC-222, as shown in Figure 6-10, where both the red channel for the membrane dye and the green channel for the FITC-tag were used to acquire images.

For each condition, 4 replicates were performed, and consistent results were obtained at both high (*i.e.* 200 000 cells/well) and low (*i.e.* 25 000 cells/well) density, as shown in Figure 6-11.

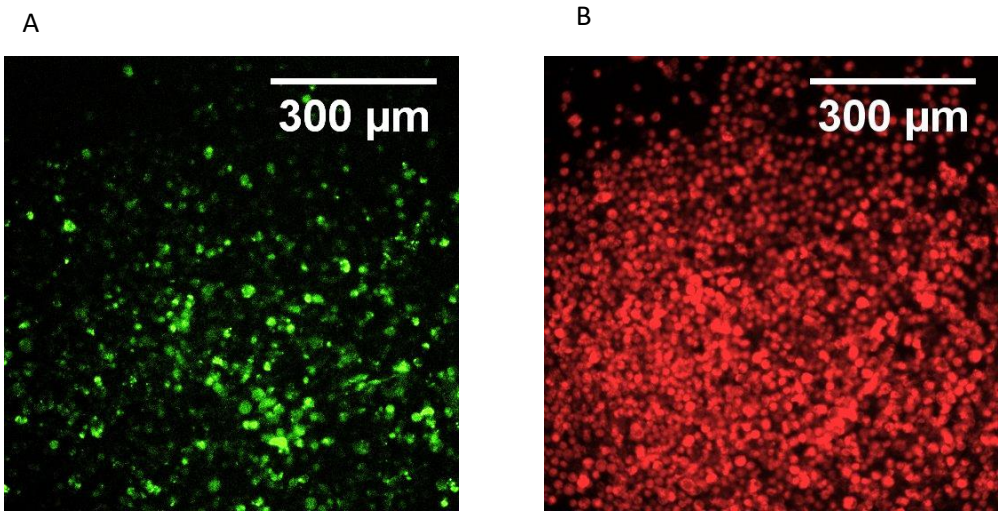


Figure 6-10: Epifluorescent images showing the successful coating of MSCs with Surfactant-FITC-222. MSCs were coated with 5 µM of Surfactant-FITC-222 as previously described before being seeded onto type II gelatin coated tissue culture plates and incubated for 24 hours. Cells were then washed once and phenol-free media was added before imaging. FM 4-64 was added at a 1:1000 dilution to enable the detection of the cells. This image corresponds to a cell seeding density of 200 000 cells/well. Both FITC (A) and FM 4-64 (B) channels were used.

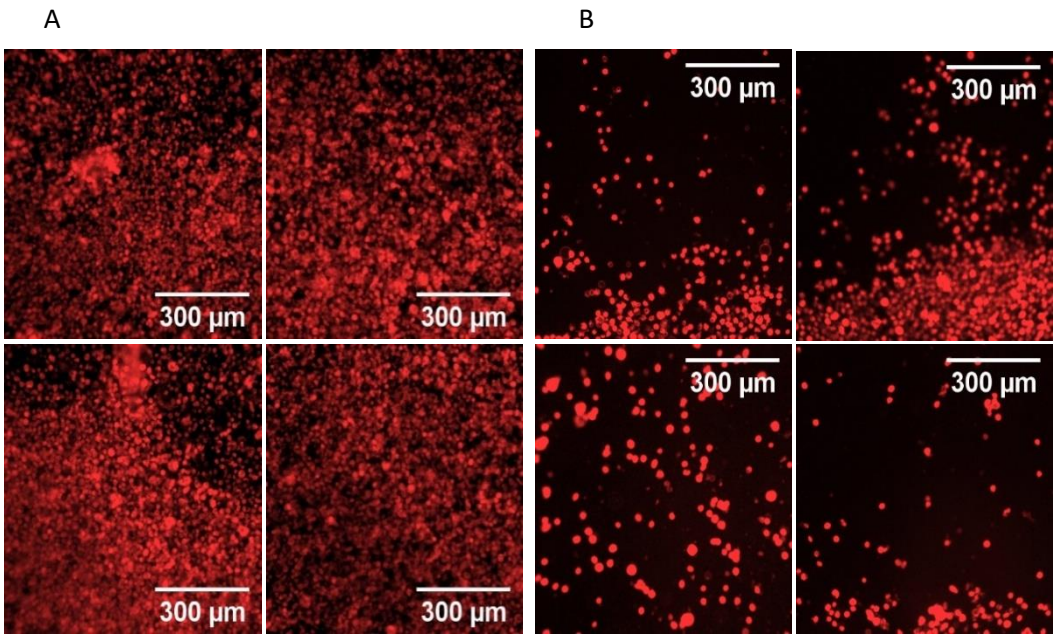


Figure 6-11: Epifluorescent images showing the consistency of 222-MSCs binding onto type II gelatin across 4 replicates, at both high (A) and low (B) cell density. MSCs were coated with 5 µM of Surfactant-222 as previously described before being seeded onto type II gelatin coated tissue culture plates and incubated for 24 hours. Cells were then washed once and phenol free media was added before imaging. FM 4-64 was added at a 1:1000 dilution to enable the detection of the cells.

The next step was therefore to compare the binding of MSCs in each condition, both on plastic and on gelatin. Overall, images showed that coated MSCs were able to bind better to type II gelatin than WT cells at each of the cell density tested (Figure 6-12, A). There did not seem to be any difference in the binding properties of cells coated with Surfactant-222 and Surfactant-FITC-222 (Figure 6-12, A, 222 and FITC-222 rows), meaning that the FITC tag does not interfere with the coating nor with the activity of 222. Finally, WT cells formed aggregates on gelatin (Figure 6-12, A, WT cells row). Interestingly, this phenomenon was not observed on plastic, except at the highest cell density (Figure 6-12, B, WT cells row). This difference could indicate that the aggregates observed on gelatin arise from some cell-cell interactions occurring to counterbalance the inability of these cells to bind directly to gelatin.

Finally, when looking at the binding properties of cells onto plastic, the opposite trend was observed compared to gelatin, as WT cells seemed to be able to bind better than coated MSCs (Figure 6-12, B). The difference was clearer at lower cell densities while, again, no clear difference was observed between cells coated with Surfactant-222 or Surfactant-FITC-222. The reduced attachment of coated cells on plastic might indicate that the presence of the protein at the cell surface creates a steric hindrance which prevents them from attaching.

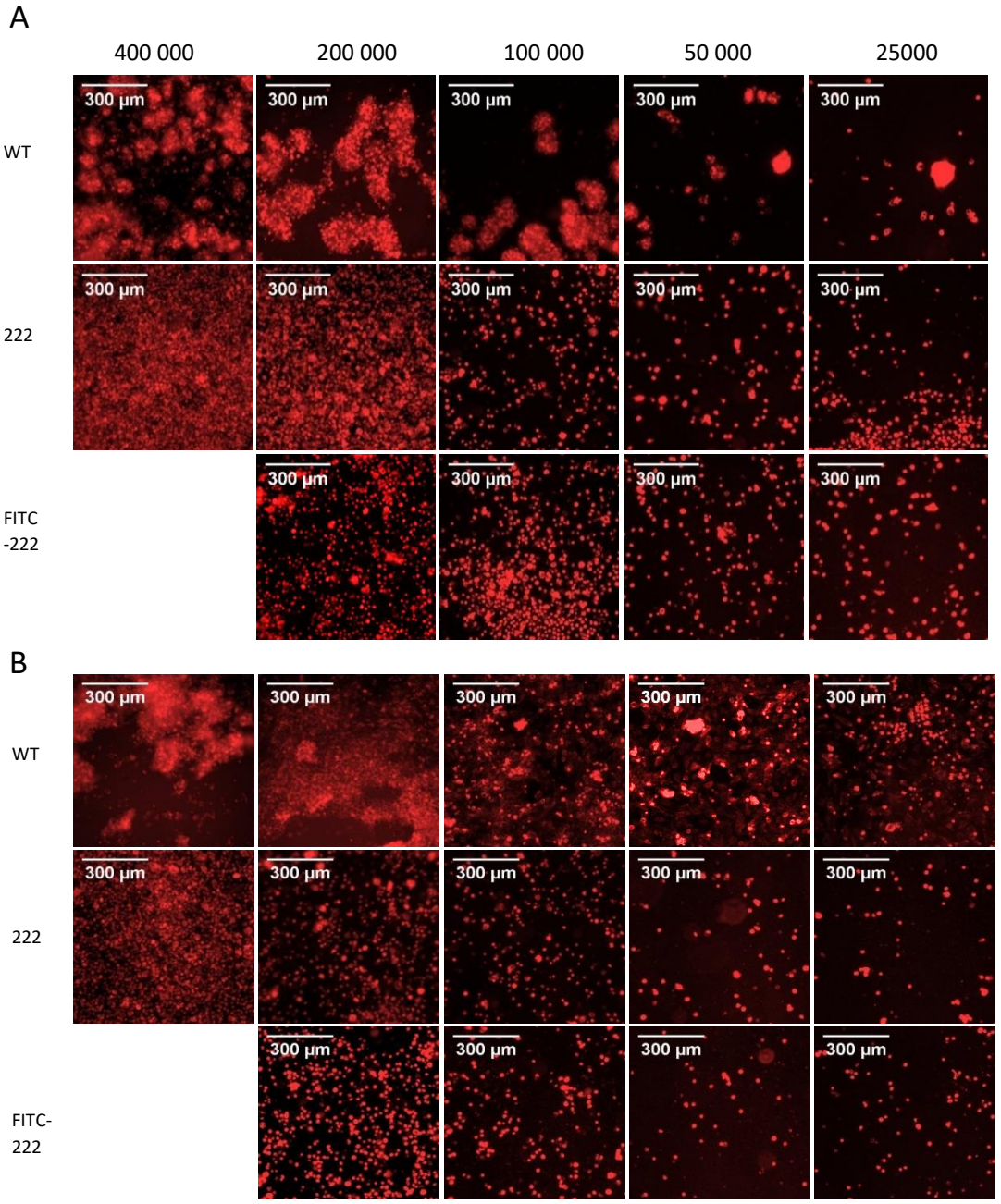


Figure 6-12: Epifluorescent images showing the difference in adhesion between WT, 222-MSCs, and FITC-222 MSCs on type II gelatin (A) and plastic (B). MSCs were coated with 5 µM of either Surfactant-222 or Surfactant-FITC-222 as previously described before being seeded onto type II gelatin coated tissue culture plates and incubated for 24 hours. Cells were then washed once and phenol-free media was added before imaging. FM 4-64 was added at a 1:1000 dilution to enable the detection of the cells. WT cells were used as a control. Cell densities are indicated at the top of each column.

The difference in attachment between WT and coated cells seeded at 25 000 cells/well was then quantified by counting the number of cells present in each well. In order to ensure that a representative sample was imaged, a tiling was performed with 16 squares (800 µmeters x 800 µmeters) being acquired in each well. This method allowed comparison of cell counts in the equivalent location of each culture well, although the method is at best semi-quantitative and hindered by the clumping of cells in some of the wells. The percentage of attached cells compared to plastic was calculated for each square, and the median and interquartile range determined (Figure 6-13). Results showed that the median attachment for coated cells on gelatin was 224.4% (range 31.6% - 5160 %) against 23.67 % for the WT (range 4.478 % – 78.39 %). This difference was shown to be significant when comparing both medians.

The trend observed here strongly suggests that coating of MSCs with 222 can enhance their gelatin-binding properties. This should in future be confirmed in repeat studies using cells from a range of patients and at a range of different passages.

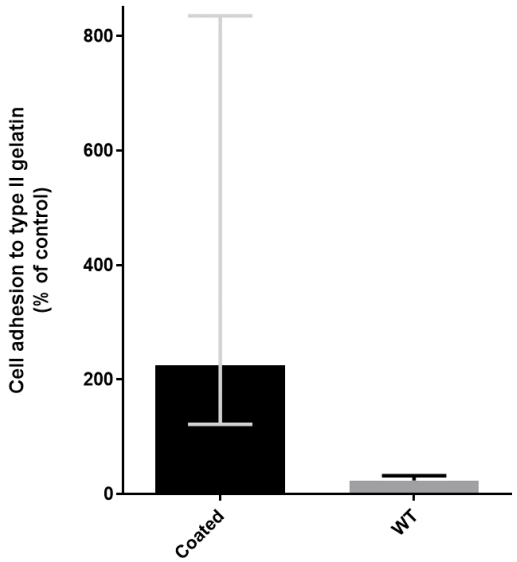


Figure 6-13: Difference in percentage of cell attachment of coated and WT MSCs on type II gelatin. 16 squares of 800 microns length were acquired in each of the wells seeded with 25 000 cells/well. Cells were automatically counted using the Image J software, and the percentage of attached cells on gelatin compared to plastic was calculated for each corresponding square. Results represent median with interquartile range. The difference was shown to be significant (P< 0.0001, Wilcoxon matched-pairs test).

6.4. Disucssion

6.4.1. Conclusion

This chapter has described the successful coating of MSCs with our most potent protein for binding to type II gelatin, namely 222. The chemistry on the protein was performed as previously described, and the presence of the surfactant around the protein was confirmed by the increase in its hydrodynamic radius. The charge was also neutralised by the addition of the surfactant following the cationisation of the protein.

The method for cell coating was however optimised compared to previously published results, as it was decided to coat MSCs in suspension, which helped to reduce the amount of protein needed as well as to inject cells directly after the coating. This method was shown to be toxic in HEPES buffer, but viability could be maintained when coating cells with 5 µM of protein in 20 mM NaP, pH 6.5 to levels as high as 80 %.

The presence of the protein at the surface of MSCs was then confirmed by confocal microscopy, which showed that Surfactant-FITC-222 was equally distributed around the cells and co-localised with the membrane dye FM-4-64.

Following up on this validation, we then tested the ability of MSCs to bind to type II gelatin, and compared it to that of WT cells. The results showed that the presence of the protein at the cell surface caused an increase in cell adhesion, confirming the initial hypothesis that coating MSCs with a gelatin binding protein would improve their binding onto type II gelatin, and could potentially improve their binding to damaged cartilage. The latter is currently being investigated with *in vivo* studies in osteoarthritic mice.

6.4.2. Future work

How could we confirm the improved binding properties of coated MSCs compared to WT cells?

Although the cell adhesion assay showed an increased binding for coated cells compared to WT ones, other experiments could be performed to further confirm this trend. First, repeats of this assay on at least 3 more patients will be needed. Comparing 222-coated MSCs with CBD-coated MSCs could also be interesting in order to determine if the difference in affinity observed at the protein level is maintained once the protein is exposed at the cell surface. Another important point to consider in the design of this experiment is that we used tissue-culture plastic plates which are pre-treated to favour cell adhesion. Even though the adhesion of both WT and coated cells onto gelatin was normalised to this plastic control, we could in the future use low-adherence plastic plates to get a better control. Using these plates would indeed rule out any effect of the pretreatment on the adhesion of WT and coated MSCs.

In order to further extend this work, it would also be interesting to study the difference in binding of coated and WT cells *ex-vivo*, onto cartilage explants from both normal and osteoarthritic patients. This would indicate whether improving the binding of MSCs to type II gelatin is enough to also improve their engraftment onto damaged cartilage.

Finally, the *in vivo* experiments which are currently underway should indicate whether or not coated MSCs are able to bind better to damaged cartilage, and, at a later stage, whether they lead to cartilage repair.

How could we characterise the coated protein further?

Although 222 was characterised in the previous chapter, more information could be obtained on the protein once it is cationised and conjugated to the surfactant.

Even though the presence of the protein at the cell surface seemed to confer more affinity to MSCs for gelatin, suggesting that the protein has remained structurally and functionally active following its insertion into the membrane, several tests should be done to confirm it. The residues modified during the cationisation could be identified by NMR, and the overall fold of the protein checked. It would also be interesting to study whether or not the residues that are modified are the same from one cationisation experiment to the other.

Carrying out this set of experiments would also indicate whether the chemistry affects the mechanism of binding of 222 onto gelatin. Indeed, we know that the cationisation modifies the residues aspartate and glutamate. When looking back at our NMR data on 222, we identified that residues E321, D326 and E334 were involved in the binding. If these residues are modified during the cationisation, it is likely that the binding affinity of 222 would be affected. If that was the case, one could imagine replacing the cationisation step by some site-directed mutagenesis work, where the aspartates and glutamates which are further away from the binding site would be mutated for lysines and arginines. According to our NMR data, the glutamates 287 and 308 as well as the aspartates 312, 322 and 324, which did not show involvement in the binding to gelatin, could be targeted for this mutagenesis work. However, it is important to note that it would be harder to do a similar work for the surfactant addition step. As this chemical reacts with lysines and arginines, a strategy to protect the lysines 296, 310 and 315, which were shown to be involved in the binding, would be useful.

This work should be coupled to more investigation on the functionality of the protein after the cationisation and the surfactant addition. However, this has proven to be difficult when

first working on the CBD and GFP as model proteins, as the surfactant made both proteins "sticky", leading to a binding to gelatin which was not necessarily specific. This means that it would be difficult to estimate the binding affinity of 222 once it is conjugated to the surfactant, although this test could be done on the cationised version.

Which further characterisation is needed for the coated cells?

As outlined in the papers mentioned in the introduction, it is important to check whether the cells coated with the protein of interest still retain the properties of their native counterparts. It would therefore be interesting to determine whether 222-coated MSCs can still proliferate and differentiate. The previous work from Armstrong et al. seems to suggest that they would, as it was shown that MSCs coated with GFP with this surfactant could still proliferate and differentiate into cartilage, bone and adipose cells [197]. However, confirming it with 222 will be needed.

Another point to investigate would be to determine the duration of 222 exposure at the cell surface. Again, previous findings from Armstrong et al. showed that GFP was maintained at the cell surface for up to 10 days, before being endocytosed via clathrin-mediated vesicles [197]. Early stage studies on MSCs coated with the CBD have shown that cells were coated for at least 5 days (not shown). It would however be interesting to find out whether 222 exposure at the cell surface is stable for the same amount of time. The mechanism by which the protein is endocytosed could also be investigated in order to determine whether the mode of clearance is surfactant or protein-specific.

How does this method compare to the previously published ones?

Overall, this chapter has described how we have developed a new, simple, one-step method for coating MSCs with a protein which binds to type II gelatin with a very high affinity. Our coating method is simpler than those relying on PPG or avidin described in the introduction of this chapter, which are both two-step methods. Our method is also more versatile, as, contrary to those which rely on the use of palmitic acid or even PPG, various proteins can be used to be conjugated to the surfactant. However, the cationisation step could alter the function of proteins with a binding site rich in aspartates and glutamates. The method is finally cheaper, as it does not require the use of protein G nor antibodies. Finally, it should enable the persistence of the protein at the cell surface for up to 10 days, which is longer than all the other non-covalent coating methods which have previously been described. Whether this is long enough to provide clinical benefits, through the secretion of trophic factors by the injected MSCs or through their direct differentiation, remains to be seen.

In conclusion, when comparing our method to that described in the introduction for the treatment of osteoarthritis, it appears that the method discovered in this thesis has the additional advantage of targeting a substrate which is specific to damaged cartilage, namely, type II gelatin. Indeed, the work performed by Dennis et al.[127] described the use of antibodies directed against molecules from the ECM. Even though the authors reported an enhanced adhesion of coated cells to damaged compared to healthy cartilage, our method still appears to be more specific. Indeed, 222 binds to type II gelatin, and the original CBD showed no binding to type II collagen [145]. Even though it would be interesting to check, it is likely that 222 would also show a low affinity for the native type II collagen found in healthy cartilage, as module 2 does not seem to have any affinity for this substrate. Mutations on 222 could also be envisioned to further increase the specificity of the protein for our substrate of

interest, as mentioned in the previous chapter. This would further reinforce the novelty of our method and would provide better specificity to damaged cartilage.

7. General conclusion

Previous clinical studies on the injection of MSCs in OA joints have described promising results in terms of pain management and improved motor functions. However, the structural benefits of this therapy were unclear and inconsistent results were obtained. In parallel, lack of engraftment was also observed in both clinical and preclinical studies. This issue was addressed in preclinical models by enhancing the attachment of MSCs to the site of damage, leading to MSCs differentiation and cartilage repair. However, no such strategies were translated in clinical trials.

The aim of this project was to develop a strategy to improve the delivery of MSCs when injected into OA joints, by enhancing their retention at the site of injection, where their properties are the most needed. The strategy relied on the coating of MSCs with a protein that binds to type II gelatin, a substrate which is specifically found in OA cartilage.

This thesis has described the successful expression of a fully folded and functional CBD of MMP-2 in the *E.coli* Shuffle cells. To the best of our knowledge, this is the first time that this protein was expressed as a soluble protein in the cytoplasm of a bacterial host. The protein was shown to bind to type II gelatin with a Kd of 20.4 nM in a plate binding assay, which was in line with previous findings. This result also validated the correct folding of the protein in the bacterial host, which was further confirmed by Mass Spectrometry, NMR and a DTNB assay assessing the number of disulfide bridges formed.

Further detailed characterisation of the interaction of the CBD with type II gelatin by NMR enabled the identification of module 2 as the most potent module. The expression, purification and characterisation of these individual modules further validated this observation, and a specificity of binding was observed for modules 1 and 3. The former displayed some specificity towards type II gelatin, while the latter seemed more specific towards type I gelatin. The binding site of each module was also investigated and a difference

in the residues involved in the binding of module 1 to each substrate was observed, which could explain the difference in specificity observed with this protein.

The generation of the chimeric proteins made of identical modules 111, 222 and 333 partially validated the different specificities of each module: 333 was shown to have a much tighter affinity for type I gelatin than for type II. Unfortunately, 111 was highly insoluble, and refolding strategies failed to obtain a correctly folded protein. However, and more importantly for this project, 222 was shown to bind with the highest affinity for type II gelatin. Its Kd was measured at 1.46 nM, which corresponds to a 14-fold improvement compared to that of the original CBD.

This protein was therefore used to coat the membrane of MSCs. The coating method relies on the use of a surfactant which surrounds the cationised protein and enables its delivery within the cytoplasmic membrane. This method was initially developed on adherent cells, and work presented in this thesis has enabled its translation onto cells in suspension. This optimisation renders the injection of coated cells into OA joints easier and safer, as no trypsinisation step is required, hence avoiding the risk of damaging the membrane and removing the newly incorporated 222 from the cell surface. Extensive optimisation of the conditions of coating also led to a reduction in the amount of 222 needed for the coating. The success of the coating was validated by confocal microscopy, which demonstrated the co-localisation of 222 with the membrane dye FM-4-64. Cell viability was not significantly decreased after the coating, and MSCs coated with 222 were shown to bind significantly better to type II gelatin than WT cells in an *in vitro* binding assay, suggesting that 222 retained its ability to bind to its substrate once incorporated in the cell membrane.

As mentioned in the last chapter, more detailed characterisation could be carried out on the cationised and conjugated forms of 222 in order to determine their binding affinity. Another important point to consider would be the retention time of the protein at the cell surface,

which was not investigated fully in this thesis. Previous work has however shown that proteins coated with this method remained at the cell surface for up to 10 days [197]. Whether the protein is exposed at the cell surface for long enough to lead to an increased engraftment of injected MSCs *in vivo* is currently being investigated in a mouse model of OA.

Another validation which is lacking in this thesis is whether or not MSCs coated with 222 maintain their trilineage differentiation potential, as well as their proliferative properties. As mentioned in the previous chapter, these properties were shown to be maintained when using the same coating method to introduce GFP and myoglobin in the membrane of MSCs [197,153]. However, it would be important to validate this observation with 222.

The importance of the ability of coated MSCs to differentiate into chondrocytes will however depend on their mode of action once injected. As mentioned in the introduction of this thesis, MSCs are thought to act mostly through paracrine effects, by releasing factors that induce the differentiation of resident cells and decrease inflammation. This hypothesis was advanced not only in cartilage, but also in other tissues, by several authors who have demonstrated that, in spite of the lack of engraftment of MSCs at the injected site, therapeutic effects were observed for sustained period of times.

One study showed that adipose derived stem cells, which only engrafted for 4 weeks and had previously shown limited potential for osteogenic differentiation, contributed to cranial defects repair [198].

Another study showed that, in spite of improving the engraftment of MSCs to the infarcted myocardium using MCP-3 as a homing factor, MSCs did not differentiate into cardiomyocytes, but rather induced a favourable remodelling of the infarct zone, leading to improved heart function [199]. This observation led the authors to suggest trophic factors as the main mode of action of MSCs in the infarcted myocardium. This hypothesis was further confirmed by other groups, who proceeded to either intra-venous [200] or intramuscular

injection of MSCs for the treatment of infarcted myocardium [201]. Only 0.02 % of the injected cells reached the target tissue in the first study [200], while cells were shown to not have migrated to the heart in the second [201]. In both cases, however, cardiac function was improved, an observation which was accompanied by a reduction of the size of the infarct and of inflammatory responses in the first study [200], while the second also reported tissue remodelling, reduced apoptosis and fibrosis, recruitment of bone marrow progenitor cells and increased mobilisation of myocardial progenitors and myocytes [201]. Notably, both studies observed increased levels of proteins such as the anti-inflammatory protein TSG-6 [200], and IGF, VEGF and Leukemia Inhibitory Factor (LIF) [201], which strongly suggest that MSCs acted through paracrine effects. This hypothesis was further supported by the fact that MSCs transduced with an anti TSG-6 siRNA did not exert any therapeutic effects in the first study [200], and that MSCs conditioned medium had the same effect as MSCs on the improvement of cardiac function in the second [201].

Altogether, these observations have led several groups to think that MSCs act via a “hit and run” effect [202], and some have even suggested to modify their name to “medicinal signalling cells” in order to reflect this mode of action [203].

However, other authors have suggested that, if MSCs were acting via paracrine effects on distant organs, this mode of action would involve the release of bioactive molecules in a burst and at high concentrations, a phenomenon which could be toxic. Another explanation which was advanced was that it is the small portion of cells engrafted into the target tissue that secrete these cytokines, but their short-term engraftment does not seem to correlate with the duration of the therapeutic benefits [204].

One could also argue that it is the lack of engraftment of MSCs at the site of damage which prevents their differentiation into the correct cell type. As mentioned in the introduction of this thesis, several pre-clinical studies have indeed shown that, when engrafted, MSCs could

differentiate into chondrocytes. However, these studies were limited to small animals, and were not recapitulated in larger animals with joints that more resemble those of humans.

These observations on the engraftment and differentiation of MSCs, however, are again not limited to the cartilage tissue. In patient suffering from Osteogenesis Imperfecta, a genetic disease which causes osteoblasts to produce a faulty type I collagen, MSCs were used in an attempt to promote bone differentiation. Several studies have reported a very low level of engraftment of the injected cells (between 1 and 2 %, [205,206]). However, these cells were shown to differentiate into osteoblasts, promoting bone growth and providing a better bone structure [206]. Interestingly, a potential correlation was observed between the cell dose and the therapeutic effect, meaning that improving engraftment of MSCs in this condition would be crucial [206]. Finally, a study by Cowan et al. on the regeneration of cranial defects following injection of adipose derived MSCs showed exceptionally high levels of engraftment. After 12 weeks, 95 % of the cells in the newly regenerated bone were donor-derived [198].

Overall, it seems that improved engraftment of MSCs onto a target tissue could lead to their subsequent differentiation into the desired cell type and enhance therapeutic benefit. If 222-coated MSCs show increased engraftment compared to WT cells *in vivo*, it would therefore be interesting to determine the fate of the coated cells once injected.

Interestingly, targeting MSCs to the damaged tissue is not only pursued in OA, but also in Osteogenesis Imperfecta and in the infarcted myocardium. Notably, this thesis has described the generation of a protein which is very potent to bind to type I denatured collagen, namely 333. As degradation of type I collagen was shown to be increased in infarcted myocardium [207, 208], one could envision the use of 333-coated MSCs for targeting cells to this tissue.

Finally, even if the mode of action of MSCs rely on paracrine effects only, the concentration of these cells at the site of injury will likely improve their therapeutic effects. Better cartilage

repair could be observed, as well as longer-lasting benefits. Another advantage on increased engraftment would be a reduction in the cell dose required for beneficial effects. These potential outcomes will be investigated in this project in the next phase of the *in vivo* trials.

8. References

1. Juneja P, H. J. Anatomy, Joints. *StatPearls* (2018). Available from: https://www.ncbi.nlm.nih.gov/books/NBK507893/

2. Mow, V. C., Ratcliffe, A. & Robin Poole, A. Cartilage and diarthrodial joints as paradigms for hierarchical materials and structures. *Biomaterials* **13,** 67–97 (1992).

3. Tamer, T. M. Hyaluronan and synovial joint: Function, distribution and healing. *Interdiscip. Toxicol.* **6,** 111–125 (2013).

4. Camarero-espinosa, S. & Weder, C. Articular cartilage: from formation to tissue engineering. *Biomater. Sci.* **4,** 734–767 (2016).

5. Bhosale, A. M. & Richardson, J. B. Articular cartilage: Structure, injuries and review of management. *Br. Med. Bull.* **87,** 77–95 (2008).

6. Mansour, J. M. & Ph, D. Biomechanics of Cartilage. *Kinesiol. Mech. Pathomechanics Hum. Mov.* 66–79 (2009).

7. Sophia Fox, A. J., Bedi, A. & Rodeo, S. A. The Basic Science of Articular Cartilage: Structure, Composition, and Function. *Sport. Heal. A Multidiscip. Approach* **1,** 461–468 (2009).

8. Buckwalter, J. A., Mankin, H. & Grodzinsky, A. Articular Cartilage and osteoarthritis. *AAOS Instr. Course Lect.* **54,** 465–480 (2005).

9. Ricard-blum, S. The Collagen Family. *Cold Spring Harb Perspect Biol* **3,** 1–19 (2011).

10. Cremer, M. A., Rosloniec, E. F. & Kang, A. H. The cartilage collagens : a review of their structure , organization , and role in the pathogenesis of experimental arthritis in animals and in human rheumatic disease. *J Mol Med* **76,** 275–288 (1998).

11. Gelse, K., Po, E. & Aigner, T. Collagens — structure , function , and biosynthesis. *Adv. Drug Deliv. Rev.* **55,** 1531–1546 (2003).

12. Eyre, D. R. Collagens and Cartilage Matrix Homeostasis. *Clin. Orthop. Relat. Res.* **427,** 118–122 (2004).

13. Knudson, C. B. & Knudson, W. Cartilage proteoglycans. *CELL Dev. Biol.* **12,** 69–78 (2001).

14. Roughley, P. J. The structure and function of cartilage proteoglycans. *Eur. Cells Mater.* **12,** 92–101 (2006).

15. Ni, G., Li, Z. & Zhou, Y. The role of small leucine-rich proteoglycans in osteoarthritis pathogenesis. *Osteoarthr. Cartil.* **22,** 896–903 (2014).

16. Rédini, F. Structure et régulation de l ' expression des protéoglycanes du cartilage articulaire. *Pathol Biol* **49,** 364–375 (2001).

17. Becerra, J. *et al.* Articular Cartilage : Structure and Regeneration. *Tissue Eng.* **16,** 617–627 (2010).

18. Ham, O. *et al.* Therapeutic potential of differentiated mesenchymal stem cells for treatment of osteoarthritis. *Int. J. Mol. Sci.* **16,** 14961–14978 (2015).

19. Krishnan, Y. & Grodzinsky, A. J. Cartilage diseases. *Matrix Biol.* 1–19 (2017).

doi:10.1016/j.matbio.2018.05.005

20. Mazor, M. *et al.* Mesenchymal stem-cell potential in cartilage repair: An update. *J. Cell. Mol. Med.* **18,** 2340–2350 (2014).

21. Loeser, R. F., Goldring, S. R., Scanzello, C. R. & Goldring, M. B. Osteoarthritis A Disease of the Joint as an Organ. *Arthritis Rheum.* **64,** 1697–1707 (2012).

22. Goldring, M. B. & Goldring, S. R. Osteoarthritis. *J. Cell. Physiol.* **213,** 626–634 (2007).

23. Mobasheri, A., Kalamegam, G., Musumeci, G. & Batt, M. E. Chondrocyte and mesenchymal stem cell-based therapies for cartilage repair in osteoarthritis and related orthopaedic conditions. *Maturitas* **78,** 188–198 (2014).

24. Silawal, S., Triebel, J., Bertsch, T. & Schulze-Tanzil, G. Osteoarthritis and the complement cascade. *Clin. Med. Insights Arthritis Musculoskelet. Disord.* **11,** 1–12 (2018).

25. Jenei-Lanzl, Z., Meurer, A. & Zaucke, F. Interleukin-1β signaling in osteoarthritis – chondrocytes in focus. *Cell. Signal.* **53,** 212–223 (2019).

26. Vinatier, C., Domínguez, E., Guicheux, J. & Caramés, B. Role of the inflammation-autophagy-senescence integrative network in Osteoarthritis. *Front. Physiol.* **9,** 1–25 (2018).

27. Mort, J. S. & Billington, C. J. Articular cartilage and changes in arthritis matrix degradation. *Arthritis Res.* **3,** 337–341 (2001).

28. Sherwood, J. C., Bertrand, J., Eldridge, S. E. & Accio, F. D. Cellular and molecular mechanisms of cartilage damage and repair. *Drug Discov. Today* **19,** 1172–1177 (2014).

29. Singh, P., Marcu, K. B., Goldring, M. B. & Otero, M. Phenotypic instability of chondrocytes in osteoarthritis: on a path to hypertrophy. *Ann. N. Y. Acad. Sci.* **1442,** 17–34 (2019).

30. Sandell, L. J. *et al.* Articular cartilage and changes in Arthritis: Cell biology of osteoarthritis. *Arthritis Res.* **3,** 107 (2001).

31. Edith, C. *et al.* Chondrocyte dedifferentiation and osteoarthritis (OA). *Biochem. Pharmacol.* 1–17 (2019). doi:10.1016/j.bcp.2019.02.036

32. Hashimoto, S., Takahashi, K., Amiel, D., Coutts, R. D. & Lotz, M. Chondrocyte apoptosis and nitric oxide production during experimentally induced osteoarthritis. *Arthritis Rheum.* **41,** 1266–1274 (1998).

33. Thomas, C. M., Fuller, C. J., Whittles, C. E. & Sharif, M. Chondrocyte death by apoptosis is associated with the initiation and severity of articular cartilage degradation. *Int. J. Rheum. Dis.* **14,** 191–198 (2011).

34. Hashimoto, S., Ochs, R. L., Komiya, S. & Lotz, M. Linkage of chondrocyte apoptosis and cartilage degradation in human osteoarthritis. *Arthritis Rheum.* **41,** 1632–1638 (1998).

35. Mobasheri, A. Role of chondrocyte death and hypocellularity in ageing human articular cartilage and the pathogenesis of osteoarthritis. *Med. Hypotheses* **58,** 193–197 (2002).

36. Manley, G. Autophagy is a Protective Mechanism in Normal Cartilage and its Aging-related Loss is Linked with Cell Death and Osteoarthritis. *Arthritis Rheum* **62,** 791–801 (2013).

37. Musumeci, G. *et al.* Biomarkers of Chondrocyte Apoptosis and Autophagy in Osteoarthritis. *Int. J. Mol. Sci. ·* **16,** 20560–20575 (2015).

38. Lotz, M., Hashimoto, S. & Kühn, K. Mechanisms of chondrocyte apoptosis. *Osteoarthr. Cartil.* **7,** 389–391 (1999).

39. Hashimoto, S. *et al.* Chondrocyte-derived apoptotic bodies and calcification of articular cartilage. *Proc. Natl. Acad. Sci.* **95,** 3094–3099 (2002).

40. Cross, M. *et al.* The global burden of hip and knee osteoarthritis: Estimates from the Global Burden of Disease 2010 study. *Ann. Rheum. Dis.* **73,** 1323–1330 (2014).

41. Shah, K., Zhao, A. G. & Sumer, H. New Approaches to Treat Osteoarthritis with Mesenchymal Stem Cells. *Stem Cells Int.* **2018,** 1–9 (2018).

42. Medvedeva, E. V. *et al.* Repair of damaged articular cartilage: Current approaches and future directions. *Int. J. Mol. Sci.* **19,** (2018).

43. Makris, E. A., Gomoll, A. H., Malizos, K. N., Hu, J. C. & Athanasiou, K. A. Repair and tissue engineering techniques for articular cartilage. *Nat. Rev. Rheumatol.* **11,** 21–34 (2015).

44. Zylińska, B., Silmanowicz, P., Sobczyńska-Rak, A., Jarosz, Ł. & Szponder, T. Treatment of articular cartilage defects: Focus on tissue engineering. *In Vivo (Brooklyn).* **32,** 1289–1300 (2018).

45. Ornetti, P. *et al.* - Does platelet-rich plasma have a role in the treatment of osteoarthritis? *Jt. Bone Spine* **7,** 104–109 (2015).

46. Kennedy, M. I., Whitney, K., Evans, T. & LaPrade, R. F. Platelet-Rich Plasma and Cartilage Repair. *Curr. Rev. Musculoskelet. Med.* **11,** 573–582 (2018).

47. Zhu, Y. *et al.* Basic science and clinical application of platelet-rich plasma for cartilage defects and osteoarthritis: A review. *Osteoarthr. Cartil.* **21,** 1627–1637 (2013).

48. Chang, K. V. *et al.* Comparative effectiveness of platelet-rich plasma injections for treating knee joint cartilage degenerative pathology: A systematic review and meta-analysis. *Arch. Phys. Med. Rehabil.* **95,** 562–575 (2014).

49. Khoshbin, A. *et al.* The Efficacy of Platelet-Rich Plasma in the Treatment of Symptomatic Knee Osteoarthritis: A Systematic Review With Quantitative Synthesis. *Arthrosc. J. Arthrosc. Relat. Surg.* **29,** 2037–2048 (2013).

50. Campbell, K. A. *et al.* Does Intra-articular Platelet-Rich Plasma Injection Provide Clinically Superior Outcomes Compared With Other Therapies in the Treatment of Knee Osteoarthritis? A Systematic Review of Overlapping Meta-analyses. *Arthrosc. - J. Arthrosc. Relat. Surg.* 1–9 (2015). doi:10.1016/j.arthro.2015.03.041

51. McConnell, S., Kolopack, P. & Davis, A. M. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum.* **45,** 453–461 (2002).

52. Collins, N. J., Misra, D., Felson, D. T., Crossley, K. M. & Roos, E. M. Measures of knee

function. *Arthritis Care Res.* **63,** 208–228 (2011).

53. Peterfy, C. G. *et al.* Whole-organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis. *Osteoarthr. Cartil.* **12,** 177–190 (2004).

54. Hawker, G. A., Mian, S., Kendzerska, T. & French, M. Measures of adult pain. *Arthritis Care Res.* **63,** 240–252 (2011).

55. Mistry, H. *et al.* Autologous chondrocyte implantation in the knee: Systematic review and economic evaluation. *Health Technol. Assess. (Rockv).* **21,** V–160 (2017).

56. Brittberg, M. *et al.* Treatment of Deep Cartilage Defects in the Knee with Autologous Chondrocyte Transplantation. *N. Engl. J. Med.* **331,** 889–895 (1994).

57. Knutsen, G. *et al.* A Randomized Trial Comparing Autologous Chondrocyte Implantation with Microfracture. *J. Bone Jt. Surgery-American Vol.* **89,** 2105–2112 (2016).

58. Davies, R. & Kuiper, N. Regenerative Medicine: A Review of the Evolution of Autologous Chondrocyte Implantation (ACI) Therapy. *Bioengineering* **6,** 1–16 (2019).

59. Bartlett, W. *et al.* Autologous chondrocyte implantation versus matrix-induced autologous chondrocyte implantation for osteochondral defects of the knee . *J. Bone Joint Surg. Br.* **87-B,** 640–645 (2005).

60. Ebert, J. R., Fallon, M., Wood, D. J. & Janes, G. C. A Prospective Clinical and Radiological Evaluation at 5 Years after Arthroscopic Matrix-Induced Autologous Chondrocyte Implantation. *Am. J. Sports Med.* **45,** 59–69 (2017).

61. Gille, J., Behrens, P., Schulz, A. P., Oheim, R. & Kienast, B. Matrix-Associated Autologous Chondrocyte Implantation. *Cartilage* **7,** 309–315 (2016).

62. Brittberg, M., Recker, D., Ilgenfritz, J. & Saris, D. B. F. Matrix-Applied Characterized Autologous Cultured Chondrocytes Versus Microfracture: Five-Year Follow-up of a Prospective Randomized Trial. *Am. J. Sports Med.* **46,** 1343–1351 (2018).

63. Mastbergen, S. C., Saris, D. B. & Lafeber, F. P. Functional articular cartilage repair: here, near, or is the best approach not yet clear? *Nat Rev Rheumatol* **9,** 277–290 (2013).

64. Grol, M. W. & Lee, B. H. Gene therapy for repair and regeneration of bone and cartilage. *Curr. Opin. Pharmacol.* **40,** 59–66 (2018).

65. Angelica, M. D. & Fong, Y. Getting arthritis gene therapy into the clinic. *Nat Rev Rheumatol* **7,** 244–249 (2011).

66. Ha, C.-W., Noh, M. J., Choi, K. B. & Lee, K. H. Initial phase I safety of retrovirally transduced human chondrocytes expressing transforming growth factor-beta-1 in degenerative arthritis patients. *Cytotherapy* **14,** 247–56 (2012).

67. Ha, C. W. *et al.* A Multicenter, Single-Blind, Phase IIa Clinical Trial to Evaluate the Efficacy and Safety of a Cell-Mediated Gene Therapy in Degenerative Knee Arthritis Patients. *Hum. Gene Ther. Clin. Dev.* **26,** 125–130 (2015).

68. Lee, B. INVOSSA, a first-in-class of cell and gene therapy for osteoarthritis treatment: the phase III trial. *Osteoarthr. Cartil.* **26,** S43–S44 (2018).

69. Karsdal, M. A. *et al.* Disease-modifying treatments for osteoarthritis (DMOADs) of

the knee and hip: lessons learned from failures and opportunities for the future. *Osteoarthr. Cartil.* **24,** 2013–2021 (2016).

70. Daley, G. Q. Stem cells and the evolving notion of cellular identity. *Philos. Trans. R. Soc. B Biol. Sci.* **370,** (2015).

71. Sharma, R. iPS Cells—The Triumphs and Tribulations. *Dent. J.* **4,** 1–23 (2016).

72. Martin, U. Therapeutic Application of Pluripotent Stem Cells: Challenges and Risks. *Front. Med.* **4,** 1–8 (2017).

73. Nowak, W., Dulak, J., Józkowicz, A., Szade, K. & Szade, A. Adult stem cells: hopes and hypes of regenerative medicine. *Acta Biochim. Pol.* **62,** 329–337 (2015).

74. Barry, F. & Murphy, M. Mesenchymal stem cells in joint disease and repair. *Nat. Rev. Rheumatol.* **9,** 584–94 (2013).

75. Filardo, G. *et al.* Mesenchymal stem cells for the treatment of cartilage lesions: From preclinical findings to clinical application in orthopaedics. *Knee Surgery, Sport. Traumatol. Arthrosc.* **21,** 1717–1729 (2013).

76. Baugé, C. & Boumédiene, K. Use of Adult Stem Cells for Cartilage Tissue Engineering: Current Status and Future Developments. *Stem Cells Int.* **2015,** 438026 (2015).

77. Hollander, A. P., Dickinson, S. C. & Kafienah, W. Stem cells and cartilage development: Complexities of a simple tissue. *Stem Cells* **28,** 1992–1996 (2010).

78. Freitag, J. *et al.* Mesenchymal stem cell therapy in the treatment of osteoarthritis: Reparative pathways, safety and efficacy - A review. *BMC Musculoskelet. Disord.* **17,** 1–13 (2016).

79. Murphy, M. B., Moncivais, K. & Caplan, A. I. Mesenchymal stem cells: Environmentally responsive therapeutics for regenerative medicine. *Exp. Mol. Med.* **45,** 1–16 (2013).

80. Tyndall, A. *et al.* Immunomodulatory properties of mesenchymal stem cells: A review based on an interdisciplinary meeting held at the Kennedy Institute of Rheumatology Division, London, UK, 31 October 2005. *Arthritis Res. Ther.* **9,** 1–10 (2007).

81. da Silva Meirelles, L., Fontes, A. M., Covas, D. T. & Caplan, A. I. Mechanisms involved in the therapeutic properties of mesenchymal stem cells. *Cytokine Growth Factor Rev.* **20,** 419–427 (2009).

82. Jevotovsky, D. S., Alfonso, A. R., Einhorn, T. A. & Chiu, E. S. Osteoarthritis and stem cell therapy in humans: a systematic review. *Osteoarthr. Cartil.* **26,** 711–729 (2018).

83. Jones, I. A., Togashi, R., Wilson, M. L., Heckmann, N. & Vangsness, C. T. Intra-articular treatment options for knee osteoarthritis. *Nat. Rev. Rheumatol.* **15,** 77–90 (2019).

84. Kafienah, W. *et al.* Three-dimensional cartilage tissue engineering using adult stem cells from osteoarthritis patients. *Arthritis Rheum.* **56,** 177–187 (2007).

85. Fatemehsadat, T. *et al.* Scaffold-free adipose-derived stem cells (ASCs) improve experimentally induced osteoarthritis in rabbits. *Arch. Iran. Med.* **15,** 495–499 (2012).

86. Al Faqeh, H., Nor Hamdan, B. M. Y., Chen, H. C., Aminuddin, B. S. & Ruszymah, B. H. I. The potential of intra-articular injection of chondrogenic-induced bone marrow stem cells to retard the progression of osteoarthritis in a sheep model. *Exp. Gerontol.* **47,** 458–464 (2012).

87. Frisbie, D. D., Kisiday, J. D., Kawcak, C. E., Werpy, N. M. & McIlwraith, C. W. Evaluation of adipose-derived stromal vascular fraction or bone marrow-derived mesenchymal stem cells for treatment of osteoarthritis. *J. Orthop. Res.* **27,** 1675–1680 (2009).

88. Li, M. *et al.* In vivo human adipose-derived mesenchymal stem cell tracking after intra-articular delivery in a rat osteoarthritis model. *Stem Cell Res. Ther.* **7,** 1–13 (2016).

89. Wang, W. *et al.* Human adipose-derived mesenchymal progenitor cells engraft into rabbit articular cartilage. *Int. J. Mol. Sci.* **16,** 12076–12091 (2015).

90. Sato, M. *et al.* Direct transplantation of mesenchymal stem cells into the knee joints of Hartley strain guinea pigs with spontaneous osteoarthritis. *Arthritis Res. Ther.* **14,** 1–9 (2012).

91. Feng, C. *et al.* Efficacy and Persistence of Allogeneic Adipose-Derived Mesenchymal Stem Cells Combined with Hyaluronic Acid in Osteoarthritis After Intra-articular Injection in a Sheep Model. *Tissue Eng. Part A* **24,** 219–233 (2017).

92. Ter Huurne, M. *et al.* Antiinflammatory and chondroprotective effects of intraarticular injection of adipose-derived stem cells in experimental osteoarthritis. *Arthritis Rheum.* **64,** 3604–3613 (2012).

93. Kim, Y. S. *et al.* Comparative matched-pair analysis of the injection versus implantation of mesenchymal stem cells for knee osteoarthritis. *Am. J. Sports Med.* **43,** 2738–2746 (2015).

94. Davatchi, F., Abdollahi, B. S., Mohyeddin, M., Shahram, F. & Nikbin, B. Mesenchymal stem cell therapy for knee osteoarthritis. Preliminary report of four patients. *Int. J. Rheum. Dis.* **14,** 211–215 (2011).

95. Emadedin, M., Aghdami, N., Taghiyar, L., Fazeli, R., Moghadasali, R., Jahangir, S., Farjad, R., Eslaminejad, M. . Intra-articular injection of autologous mesenchymal stem cells in six patients with knee Osteoarthritis. *Arch. Iran. Med.* **15,** 422–428 (2012).

96. Al-Najar, M. *et al.* Intra-articular injection of expanded autologous bone marrow mesenchymal cells in moderate and severe knee osteoarthritis is safe: A phase I/II study. *J. Orthop. Surg. Res.* **12,** 1–6 (2017).

97. Emadedin, M. *et al.* Long-Term Follow-up of Intra-articular Injection of Autologous Mesenchymal Stem Cells in Patients with Knee, Ankle, or Hip Osteoarthritis. *Arch. Iran. Med.* **18,** 336–344 (2015).

98. Lopa, S., Colombini, A., Moretti, M. & de Girolamo, L. Injective mesenchymal stem cell-based treatments for knee osteoarthritis: from mechanisms of action to current clinical evidences. *Knee Surgery, Sport. Traumatol. Arthrosc.* (2018). doi:10.1007/s00167-018-5118-9

99. Jo, C. H. *et al.* Intra-Articular Injection of Mesenchymal Stem Cells for the Treatment of Osteoarthritis of the Knee: A Proof-of-Concept Clinical Trial. *Stem Cells* **32,** 1254–

1266 (2014).

100. Gupta, P. K. *et al.* Efficacy and safety of adult human bone marrow-derived, cultured, pooled, allogeneic mesenchymal stromal cells (Stempeucel®): Preclinical and clinical trial in osteoarthritis of the knee joint. *Arthritis Res. Ther.* **18,** 1–18 (2016).

101. Vega, A. *et al.* Treatment of Knee Osteoarthritis With Allogeneic Bone Marrow Mesenchymal Stem Cells: A Randomized Controlled Trial. *Transplantation* **00,** 1 (2015).

102. Orozco, L. *et al.* Treatment of knee osteoarthritis with autologous mesenchymal stem cells: A pilot study. *Transplantation* **95,** 1535–1541 (2013).

103. Soler, R. *et al.* Final results of a phase I-II trial using ex vivo expanded autologous Mesenchymal Stromal Cells for the treatment of osteoarthritis of the knee confirming safety and suggesting cartilage regeneration. *Knee* **23,** 647–654 (2016).

104. Kim, Y. S. *et al.* Assessment of clinical and MRI outcomes after mesenchymal stem cell implantation in patients with knee osteoarthritis: A prospective study. *Osteoarthr. Cartil.* **24,** 237–245 (2016).

105. Fodor, P. B. & Paulseth, S. G. Adipose Derived Stromal Cell (ADSC) Injections for Pain Management of Osteoarthritis in the Human Knee Joint. *Aesthetic Surg. J.* **36,** 229–236 (2016).

106. Lluis Orozco, Anna Munar, Robert Soler, Mercedes Alberca, Francesc Soler, Marina Huguet, Joan Sentıs, A. S. & Javier, G.-S. Treatment of Knee Osteoarthritis With Autologous Mesenchymal Stem Cells: Two-Year Follow-up Results. *Transplantation* **97,** e65–e66 (2014).

107. Davatchi, F., Abdollahi, B. S., Mohyeddin, M., Shahram, F. & Nikbin, B. Mesenchymal stem cell therapy for knee osteoarthritis: 5 years follow-up of three patients. *Int. J. Rheum. Dis.* **14,** 211–215 (2011).

108. Jo, C. H. *et al.* Intra-articular Injection of Mesenchymal Stem Cells for the Treatment of Osteoarthritis of the Knee: A 2-Year Follow-up Study. *Am. J. Sports Med.* **45,** 2774–2783 (2017).

109. Pers, Y.-M. M., Rackwitz, L. & Ferreira, R. Adipose Mesenchymal Stromal Cell-Based Therapy for Severe Osteoarthritis of the Knee : *Stem Cells Transl. Med.* **5,** 847–856 (2016).

110. Lamo-Espinosa, J. M. *et al.* Intra-articular injection of two different doses of autologous bone marrow mesenchymal stem cells versus hyaluronic acid in the treatment of knee osteoarthritis: Long-term follow up of a multicenter randomized controlled clinical trial (phase I/II). *J. Transl. Med.* **16,** 1–9 (2018).

111. Gibbs, N., Diamond, R., Sekyere, E. O. & Thomas, W. D. Management of knee osteoarthritis by combined stromal vascular fraction cell therapy, platelet-rich plasma, and musculoskeletal exercises: A case series. *J. Pain Res.* **8,** 799–806 (2015).

112. Bansal, H. *et al.* Intra-articular injection in the knee of adipose derived stromal cells (stromal vascular fraction) and platelet rich plasma for osteoarthritis. *J. Transl. Med.* **15,** 1–11 (2017).

113. Van Pham, P. *et al.* Symptomatic knee osteoarthritis treatment using autologous

adipose derived stem cells and platelet-rich plasma: a clinical study. *Biomed. Res. Ther.* **1,** 2019 (2014).

114. Koh, Y. G., Choi, Y. J., Kwon, S. K., Kim, Y. S. & Yeo, J. E. Clinical results and second-look arthroscopic findings after treatment with adipose-derived stem cells for knee osteoarthritis. *Knee Surgery, Sport. Traumatol. Arthrosc.* **23,** 1308–1316 (2015).

115. Pintat, J. *et al.* Intra-articular Injection of Mesenchymal Stem Cells and Platelet-Rich Plasma to Treat Patellofemoral Osteoarthritis: Preliminary Results of a Long-Term Pilot Study. *J. Vasc. Interv. Radiol.* **28,** 1708–1713 (2017).

116. Bastos, R. *et al.* Intra-articular injections of expanded mesenchymal stem cells with and without addition of platelet-rich plasma are safe and effective for knee osteoarthritis. *Knee Surgery, Sport. Traumatol. Arthrosc.* **26,** 3342–3350 (2018).

117. Pak, J., Lee, J. H., Park, K. S., Jeong, B. C. & Lee, S. H. Regeneration of Cartilage in Human Knee Osteoarthritis with Autologous Adipose Tissue-Derived Stem Cells and Autologous Extracellular Matrix. *Biores. Open Access* **5,** 192–200 (2016).

118. Pak, J. Regeneration of human bones in hip osteonecrosis and human cartilage in knee osteoarthritis with autologous adipose-tissue-derived stem cells: A case series. *J. Med. Case Rep.* **5,** 1–8 (2011).

119. Liu, W. *et al.* Microcryogels as injectable 3-D cellular microniches for site-directed and augmented cell delivery. *Acta Biomater.* **10,** 1864–1875 (2014).

120. Kurtz, A. Mesenchymal Stem Cell Delivery Routes and Fate. *Int. J. Stem Cells* **1,** 1–7 (2008).

121. Broeckx, S. Y. *et al.* Equine allogeneic chondrogenic induced mesenchymal stem cells: A GCP target animal safety and biodistribution study. *Res. Vet. Sci.* **117,** 246–254 (2018).

122. Mamidi, M. K., Das, A. K., Zakaria, Z. & Bhonde, R. Mesenchymal stromal cells for cartilage repair in osteoarthritis. *Osteoarthr. Cartil.* **24,** 1307–1316 (2016).

123. Iijima, H., Isho, T., Kuroki, H., Takahashi, M. & Aoyama, T. Effectiveness of mesenchymal stem cells for treating patients with knee osteoarthritis: a meta-analysis toward the establishment of effective regenerative rehabilitation. *npj Regen. Med.* **3,** 1–13 (2018).

124. Kamei, G. *et al.* Articular cartilage repair with magnetic mesenchymal stem cells. *Am. J. Sports Med.* **41,** 1255–1264 (2013).

125. Koga, H. *et al.* Local adherent technique for transplanting mesenchymal stem cells as a potential treatment of cartilage defect. *Arthritis Res. Ther.* **10,** 1–10 (2008).

126. Uvebrandt, K. *et al.* Integrin α10β1 selected Equine MSCs Have Improved Chondrogenic Differentiation Immunomodulatory and Cartilage Adhesion Capacity. *Ann. Stem Cell Res.* **2,** 1–5 (2019).

127. Dennis, J. E., Cohen, N., Goldberg, V. M. & Caplan, A. I. Targeted delivery of progenitor cells for cartilage repair. *J. Orthop. Res.* **22,** 735–741 (2004).

128. Leijs, M. J. C. *et al.* Pre-Treatment of Human Mesenchymal Stem Cells with Inflammatory Factors or Hypoxia Does Not Influence Migration to Osteoarthritic Cartilage and Synovium. *Am. J. Sports Med.* **45,** 1151–1161 (2017).

129. Hollander, A. P. *et al.* Increased damage to type II collagen in osteoarthritic articular cartilage detected by a new immunoassay. *J. Clin. Invest.* **93,** 1722–1732 (1994).

130. Dodge, G. R. & Poole, A. R. Immunohistochemical detection and immunochemical analysis of type II collagen degradation in human normal, rheumatoid, and osteoarthritic articular cartilages and in explants of bovine articular cartilage cultured with interleukin 1. *J. Clin. Invest.* **83,** 647–661 (1989).

131. Wu, W. *et al.* Sites of collagenase cleavage and denaturation of type II collagen in aging and osteoarthritic articular cartilage and their relationship to the distribution of matrix metalloproteinase 1 and matrix metalloproteinase 13. *Arthritis Rheum.* **46,** 2087–2094 (2002).

132. Hollander, A. P. *et al.* Damage to type II collagen in aging and osteoarthritis starts at the articular surface, originates around chondrocytes, and extends into the cartilage with progressive degeneration. *J. Clin. Invest.* **96,** 2859–2869 (1995).

133. Lopresti-Morrow, L. L. *et al.* Cloning, expression, and type II collagenolytic activity of matrix metalloproteinase-13 from human osteoarthritic cartilage. *J. Clin. Invest.* **97,** 761–768 (2008).

134. Miller, E. J. *et al.* Cleavage of Type II and III Collagens with Mammalian Collagenase: Site of Cleavage and Primary Structure at the NH2-Terminal Portion of the Smaller Fragment Released from Both Collagens. *Biochemistry* **15,** 787–792 (1976).

135. Billinghurst, R. C. *et al.* Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage. *J. Clin. Invest.* **99,** 1534–1545 (1997).

136. Hollander, A. P. *et al.* Increased damage to type II collagen in osteoarthritic articular cartilage detected by a new immunoassay. *J. Clin. Invest.* **93,** 1722–1732 (1994).

137. Xu, X., Wang, Y., Lauer-Fields, J. L., Fields, G. B. & Steffensen, B. Contributions of the MMP-2 collagen binding domain to gelatin cleavage: Substrate binding via the collagen binding domain is required for hydrolysis of gelatin but not short peptides. *Matrix Biol.* **23,** 171–181 (2004).

138. Klein, T. & Bischoff, R. Physiology and pathophysiology of matrix metalloproteases. *Amino Acids* **41,** 271–290 (2011).

139. Morgunova, E. *et al.* Structural properties of matrix metalloproteinases. *Matrix Met. Inhib. Cancer Ther.* **55,** 39–66 (1999).

140. Xu, X. *et al.* Nuclear magnetic resonance mapping and functional confirmation of the collagen binding sites of matrix metalloproteinase-2. *Biochemistry* **48,** 5822–5831 (2009).

141. Steffensen, B., Margaretha, W. & Overall, C. M. The Extracellular matrix binding properties of recombinant Fn-type II like modules of human 72 kDa Gelatinase. *J. Biol. Chem.* **270,** 11555–11566 (1995).

142. Morgunova, E. *et al.* Structure of Human Pro-Matrix Metalloproteinase-2: Activation Mechanism Revealed. *Science (80-.).* **284,** 1667–1670 (1999).

143. Bányai, L., Tordai, H. & Patthty, L. Structure and domain-domain interactions of the gelatin binding site of human 72-kilodalton type IV collagenase (gelatinase A, matrix metalloproteinase 2). *J. Biol. Chem.* **271,** 12003–12008 (1996).

144. Bányai, L., Tordai, H. & Patthy, L. The gelatin-binding site of human 72 kDa type IV

collagenase (gelatinase A). *Biochem. J.* **298 (Pt 2,** 403–7 (1994).

145. Steffensen, B., Xu, X., Martin, P. A. & Zardeneta, G. Human fibronectin and MMP-2 collagen binding domains compete for collagen binding sites and modify cellular activation of MMP-2. *Matrix Biol.* **21,** 399–414 (2002).

146. Murphy, G. *et al.* Assessment of the Role of the Fibronectin-Like Domain of Gelatinase-A by Analysis of a Deletion Mutant. *J. Biol. Chem.* **Vol 269,** 6632–6636 (1994).

147. Tam, E. M., Moore, T. B., Butler, G. S. & Overall, C. M. Characterization of the distinct collagen binding, helicase and cleavage mechanisms of matrix metalloproteinase 2 and 14 (gelatinase A and MT1-MMP): The differential roles of the MMP hemopexin C domains and the MMP-2 fibronectin type II modules in collage. *J. Biol. Chem.* **279,** 43336–43344 (2004).

148. Xu, X., Chen, Z., Wang, Y., Yamada, Y. & Steffensen, B. Functional basis for the overlap in ligand interactions and substrate specificities of matrix metalloproteinases-9 and -2. *Biochem J* **392,** 127–134 (2005).

149. Mikhailova, M. *et al.* Identification of collagen binding domain residues that govern catalytic activities of matrix metalloproteinase-2 (MMP-2). *Matrix Biol* **25,** 713–724 (2015).

150. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5,** 1–17 (2014).

151. Dyson, M. R. Fundamentals of Expression in Mammalian Cells. *Adv. Exp. Med. Biol.* **896,** 217–224 (2016).

152. Geisse, S. Reflections on more than 10 years of TGE approaches. *Protein Expr. Purif.* **64,** 99–107 (2009).

153. Armstrong, J. P. Artificial Membrane-Binding Proteins. (2014).

154. Domon, B. & Aebersold, R. Mass Spectrometry and Protein Analysis. *Science (80-.).* **312,** 212–218 (2006).

155. Stults, J. T. Matrix-assisted laser desorption / ionization mass spectrometry (MALDI-MS). *Curr. Opin. Struct. Biol.* **5,** 691–698 (1995).

156. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1,** 2876–2890 (2006).

157. Kelly, S. M. & Price, N. C. The use of circular dichroism in the investigation of protein structure and function. *Curr. Protein Pept. Sci.* **1,** 349–84 (2000).

158. Gehrmann, M., Briknarová, K., Bányai, L., Patthy, L. & Llinás, M. The col-1 module of human matrix metalloproteinase-2 (MMP-2): Structural/functional relatedness between gelatin-binding fibronectin type II modules and lysine-binding kringle domains. *Biol. Chem.* **383,** 137–148 (2002).

159. Briknarová, K. *et al.* Gelatin-binding region of human matrix metalloproteinase-2: Solution structure, dynamics, and function of the COL-23 two-domain construct. *J. Biol. Chem.* **276,** 27613–27621 (2001).

160. Gehrmann, M. L. *et al.* Modular autonomy, ligand specificity, and functional cooperativity of the three in-tandem fibronectin type II repeats from human matrix

metalloproteinase. *J. Biol. Chem.* **279,** 46921–46929 (2004).

161. Collier, I. E., Krasnov, P. A., Strongin, A. Y., Birkedal-Hansen, H. & Goldberg, G. I. Alanine scanning mutagenesis and functional analysis of the fibronectin-like collagen-binding domain from human 92-kDa type IV collagenase. *J. Biol. Chem.* **267,** 6776–6781 (1992).

162. Demain, A. L. & Vaishnav, P. Production of Recombinant Proteins by Microbes and Higher Organisms. *Compr. Biotechnol. Second Ed.* **3,** 333–345 (2011).

163. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307,** 249–264 (2008).

164. Ferrer-miralles, N., Saccardo, P., Corchero, J. L., Xu, Z. & García-fruitós, E. General Introduction : Recombinant Protein Production and Purification of Insoluble Proteins. **1258,** (2015).

165. Georgiou, G. & Valax, P. Expression of correctly folded proteins in Escherichia coli. *Curr Opin Biotechnol* **7,** 190–197 (1996).

166. Ventura, S. & Villaverde, A. Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **24,** 179–185 (2006).

167. Ke, N. & Berkmen, M. Production of Disulfide-Bonded Proteins in Escherichia coli. *Curr. Protoc. Mol. Biol.* **108,** 16.1B.1–16.1B.21 (2014).

168. Tordai, H. & Patthy, L. The gelatin-binding site of the second type-II domain of gelatinase A/MMP-2. *Eur. J. Biochem.* **259,** 513–518 (1999).

169. Burgess-Brown, N. A. *et al.* Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expr. Purif.* **59,** 94–102 (2008).

170. Malakhov, M. P. *et al.* SUMO Fusions and SUMO-pecific Protease for Efficient Expression and Purification of Proteins. *J. Struct. Funct. Genomics* **5,** 75–86 (2004).

171. Marblestone, J. G. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **15,** 182–189 (2006).

172. Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8,** 1668–1674 (1999).

173. Butt, T. R., Edavettal, S. C., Hall, J. P. & Mattern, M. R. SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* **43,** 1–9 (2005).

174. Makino, T., Skretas, G. & Georgiou, G. Strain engineering for improved expression of recombinant proteins in bacteria. *Microb. Cell Fact.* **10,** 32 (2011).

175. Peisley, A. A. & Gooley, P. R. High-level expression of a soluble and functional fibronectin type II domain from MMP-2 in the Escherichia coli cytoplasm for solution NMR studies. *Protein Expr. Purif.* **53,** 124–131 (2007).

176. García-Fruitós, E. in *Insoluble Proteins: Methods and Protocols* **1258,** 1–422 (2014).

177. Kaestner, L., Scholz, A. & Lipp, P. Conceptual and technical aspects of transfection and gene delivery. *Bioorganic Med. Chem. Lett.* **25,** 1171–1176 (2015).

178. Samuelson, J. C. Disulfide-Bonded Protein Production in *E. coli*. *Genet. Eng. Biotechnol. News* **32,** 35–35 (2012).

179. Pickford, A. R., Potts, J. R., Bright, J. R., Phan, I. & Campbell, I. D. Solution structure of a type 2 module from fibronectin: Implications for the structure and function of the gelatin-binding domain. *Structure* **5,** 359–370 (1997).

180. Briknarová, K. *et al.* The second type II module from human matrix metalloproteinase 2: Structure, function and dynamics. *Structure* **7,** 1235–1245 (1999).

181. Xu, X., Chen, Z., Wang, Y., Bonewald, L. & Steffensen, B. Inhibition of MMP-2 gelatinolysis by targeting exodomain-substrate interactions. *Biochem. J.* **406,** 147–155 (2007).

182. Berton, A. *et al.* Involvement of Fibronectin Type II Repeats in the Efficient Inhibition of Gelatinases A and B by Long-chain Unsaturated Fatty Acids. *J. Biol. Chem.* **276,** 20458–20465 (2001).

183. Shipley, J. M. *et al.* The Structural Basis for the Elastolytic Activity of the 92-kDa and 72-kDa Gelatinases. *J. Biol. Chem.* **271,** 4335–4341 (1996).

184. Overall, C. M. Matrix Metalloproteinase Substrate Binding Domains, Modules and Exosites. *Methods Mol. Biol.* **151,** 79–120

185. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349,** 293–299 (1991).

186. Krah, S. *et al.* Engineering bispecific antibodies with defined chain pairing. *N. Biotechnol.* **39,** 167–173 (2017).

187. Brinkmann, U. *et al.* The making of bispecific antibodies. *MAbs* **9,** 182–212 (2017).

188. Han, X. *et al.* Adnectin-Based Design of Chimeric Antigen Receptor for T Cell Engineering. *Mol. Ther.* **25,** 2466–2476 (2017).

189. Kim, Y. S., Kong, W. H., Kim, H. & Hahn, S. K. Targeted systemic mesenchymal stem cell delivery using hyaluronate – wheat germ agglutinin conjugate. *Biomaterials* **106,** 217–227 (2016).

190. Won, Y. W., Patel, A. N. & Bull, D. A. Cell surface engineering to enhance mesenchymal stem cell migration toward an SDF-1 gradient. *Biomaterials* **35,** 5627–5635 (2014).

191. Huang, B. *et al.* Peptide modified mesenchymal stem cells as targeting delivery system transfected with miR-133b for the treatment of cerebral ischemia. *Int. J. Pharm.* **531,** 90–100 (2017).

192. Takayama, Y. *et al.* Long-term drug modification to the surface of mesenchymal stem cells by the avidin-biotin complex method. *Sci. Rep.* **7,** 1–10 (2017).

193. Ko, I. K. *et al.* Targeting improves MSC treatment of inflammatory bowel disease. *Mol. Ther.* **18,** 1365–1372 (2010).

194. Whitaker, A. M., Farooq, M. A., Edwards, S. & Gilpin, N. W. Cell surface glycoengineering improves selectin-mediated adhesion of mesenchymal stem cells (MSCs) and cardiosphere- derived cells (CDCs): pilot validation in porcine ischemia-reperfusion models. *Biomaterials* **74,** 19–30 (2016).

195. Kean, T. J. *et al.* Development of a peptide-targeted, myocardial ischemia-homing, mesenchymal stem cell. *J. Drug Target.* **20,** 23–32 (2012).

196. Asif, S. *et al.* Heparinization of cell surfaces with short peptide-conjugated PEG-lipid regulates thromboinflammation in transplantation of human MSCs and hepatocytes. *Acta Biomater.* **35,** 194–205 (2016).

197. Armstrong, J. P. K. *et al.* Artificial membrane-binding proteins stimulate oxygenation of stem cells during engineering of large cartilage tissue. *Nat. Commun.* **6,** 7405 (2015).

198. Ruetze, M. & Richter, W. Adipose-derived stromal cells for osteoarticular repair: Trophic function versus stem cell activity. *Expert Rev. Mol. Med.* **16,** 1–19 (2014).

199. Schenk, S. *et al.* Monocyte Chemotactic Protein-3 Is a Myocardial Mesenchymal Stem Cell Homing Factor. *Stem Cells* **25,** 245–251 (2006).

200. Cabello, C. M. *et al.* Intravenous hMSCs Improve Myocardial Infarction in Mice because Cells Embolized in Lung Are Activated to Secrete the Anti-inflammatory Protein TSG-6 Ryang. *Cell Stem Cell* **5,** 54–63 (2009).

201. Shabbir, A., Zisa, D., Suzuki, G. & Lee, T. Heart failure therapy mediated by the trophic activities of bone marrow mesenchymal stem cells: a noninvasive therapeutic regimen. *Am. J. Physiol. Heart Circ. Physiol.* **296,** H1888–H1897 (2009).

202. Von Bahr, L. *et al.* Analysis of tissues following mesenchymal stromal cell therapy in humans indicates limited long-term engraftment and no ectopic tissue formation. *Stem Cells* **30,** 1575–1578 (2012).

203. Arnold I Caplan. Mesenchymal Stem Cells : Time to Change the Name ! *Stem Cells Transl. Med.* 1445–1451 (2017).

204. Prockop, D. J. Repair of tissues by adult stem/progenitor cells (MSCs): Controversies, myths, and changing paradigms. *Mol. Ther.* **17,** 939–946 (2009).

205. Horwitz, E. M. *et al.* Isolated allogeneic bone marrow-derived mesenchymal cells engraft and stimulate growth in children with osteogenesis imperfecta: Implications for cell therapy of bone. *Proc. Natl. Acad. Sci.* **99,** 8932–8937 (2002).

206. Horwitz, E. M. *et al.* Transplantability and therapeutic effects of bone marrow-derived mesenchymal cells in children with osteogenesis imperfecta. *Nat. Med.* **5,** 309–313 (1999).

207. Bertelsen, D. M. *et al.* Matrix Metalloproteinase Mediated Type I Collagen Degradation is an Independent Predictor of Increased Risk of Acute Myocardial Infarction in Postmenopausal Women. *Sci. Rep.* **8,** 1–7 (2018).

208. Cleutjens, J. P. M., Kandala, J. C., Guarda, E., Guntaka, R. V. & Weber, K. T. Regulation of collagen degradation in the rat myocardium after infarction. *J. Mol. Cell. Cardiol.* **27,** 1281–1292 (1995).

209. Ala-Aho, R. & Kähäri, V. M. Collagenases in cancer. *Biochimie* **87,** 273–286 (2005).