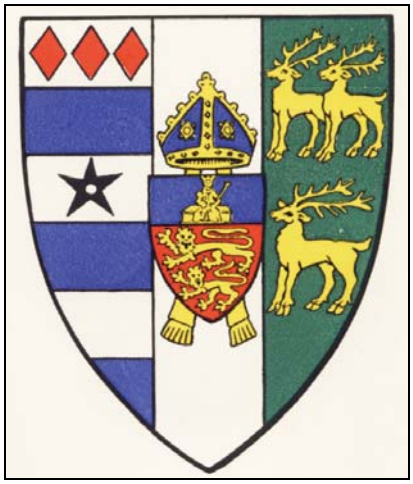
# Structural and functional characterisation of the collagen binding domain of fibronectin



Christopher John Millard

Lincoln College, Oxford

Submitted for the degree of Doctor of Philosophy

Michaelmas Term 2007

# Abstract

Fibronectin is an extracellular multidomain glycoprotein that directs and regulates a variety of cell processes such as proliferation, development, haemostasis, embryogenesis, and wound healing. As a major component of blood, fibronectin exists as a soluble disulphide linked dimer, but it can also be incorporated into an insoluble cross-linked fibrillar network to form a major component of the extracellular matrix. Fibronectin is composed of an extended chain of module repeats termed Fn1, Fn2, and Fn3 that bind to a wide range of transmembrane receptors and extracellular matrix components, including collagen.

The gelatin binding domain of fibronectin was first isolated as a 45kDa proteolytic fragment and has since been found to be composed of six modules: $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1 (in this notation $^{n}$FX represents the $n^{th}$ type X module in the native protein). This domain has been reported to bind to both collagen and denatured collagen (gelatin), but with 10-100 times higher affinity to the latter; it can be purified to homogeneity on a gelatin affinity column.

In the work presented here, fragments of the gelatin binding domain are expressed in *P. pastoris*, purified to homogeneity, and investigated at the molecular level. Through a dissection approach, surface plasmon resonance (SPR) is used to characterise the recombinantly produced protein, to accumulate more information about the function of the full domain. NMR is used to assess the folding of the protein fragments at atomic resolution. In particular, the secondary structure of $^{8}$Fn1-$^{9}$Fn1 is mapped using inter-strand NOEs, which suggests that the construct takes the fold of a pair of typical Fn1 modules.

Gelatin affinity chromatography is used to confirm that both Fn1 and Fn2 modules contribute to gelatin binding, possibly in two clusters ($^{1}$Fn2-$^{2}$Fn2 and $^{8}$Fn1-$^{9}$Fn1). The $^{7}$Fn1 module may perform a structural role in linking together these two interaction sites, in the same way as suggested for $^{6}$Fn1, which is thought to act in a structural manner to enhance the binding of $^{1}$Fn2-$^{2}$Fn2 to gelatin.

Three carbohydrate moieties are found on this domain, one on $^{2}$Fn2 and two on $^{8}$Fn1. Here, by means of expressing different protein length fragments, and by site directed mutagenesis, the role of each sugar chain is investigated independently. The sugar chain on $^{2}$Fn2 does not appear to promote binding to collagen, nor does the first sugar chain on $^{8}$Fn1 (N-linked to N497), implying another role for these sugars such as protection from proteolysis. However, the presence of at least a single GlcNAc sugar residue on the second sugar chain site on $^{8}$Fn1 (N- linked to N511) is essential for full affinity binding to collagen.

Direct binding of the $^{8}$Fn1-$^{9}$Fn1 module pair to collagen is assessed with a short collagen peptide and the binding is monitored by NMR. The peptide appears to bind, predominantly to the final strand of $^{8}$Fn1, the first β- strand of $^{9}$Fn1, and the linker between the two modules, with μM affinity. A model for bound peptide is proposed.

The highly conserved amino acid motif Ile-Gly-Asp (IGD) is found on four of the nine N-terminal Fn1 modules of fibronectin. Tetrapeptides containing the IGD were demonstrated to promote the migration of fibroblast cells into a native collagen matrix. Two of these "bioactive" IGD motifs are found within the gelatin binding domain, one on $^{7}$Fn1 and one on $^{9}$Fn1. In this study, the motif in the $^{8}$Fn1-$^{9}$Fn1 module pair is shown to be located in a tightly constrained loop within $^{9}$Fn1. By site directed mutagenesis, the IGD motifs of $^{7}$Fn1 and $^{9}$Fn1 are subjected to single amino acid substitutions, and their ability to stimulate cell migration assessed in our assay.

By NMR, the fold of the IGD mutant proteins is found to be unaffected by the mutation with respect to the wild type, with the exception of small perturbations around the substitution site. While the wild type module is able to stimulate fibroblast migration, the mutant proteins show reduced or negligible bioactivity. The larger fragments show far more potency in stimulating fibroblast migration, with $^{8}$Fn1-$^{9}$Fn1 (one IGD motif) $10^{4}$ times more potent than the IGD peptide, and the full gelatin binding domain (two IGD motifs) $10^{6}$ times more potent than the $^{8}$Fn1-$^{9}$Fn1. Potential mechanisms for this enormous enhancement of the IGD potency in different contexts are discussed.

# Acknowledgements

Over the past three years I have learnt a great deal about how to conduct scientific research and the process of building a biological story through the use of many different biochemical techniques. This as only been possible through the generous day-to-day help and support from my lab; Iain Campbell who has allowed me time to explore my own ideas but at the same time steered me in the right direction, Andy Pickford for his teaching and guidance during the first year of my project, Ioannis Vakonakis and Dave Staunton for sharing their expertise in NMR and molecular biology, and all the other members who studied alongside me in the IDC group and made our group day and Christmas party such enjoyable events (Pengju, Sonja, Gregor, Nick, Ivan, Camilla, Kate, Max, Alice, Charlie, Luke, Uli, and Amanda to name just a few).

My three years would have not been the same without so many clubs and societies; Friday lunchtime football led by Terry Butters in the University Parks, training and racing with OUCCC, events and socials with OUOC scattered throughout Europe (largely thanks to Ian, SGB, Cerys, Ali, Scott, Benjamin and Joe), participation in the London marathon, and the love and support offered by St Ebbe's Church.

I would like to acknowledge George Wadhams and Mark Roberts for their academic support, Lincoln College for having me as a student for over seven years, and the BBSRC for funding me throughout my studentship.

Lastly I would like to thank my parents and family for their support in encouraging me throughout my work. I would especially like to thank my beautiful wife Katherine, who allowed me to take her hand during the second year of my D.Phil, and without her I would have certainly struggled to maintain the motivation to complete this work.

# Publications

- Millard, C.J., Campbell, I.D. and Pickford, A.R. (2005) Gelatin binding to the $^{8}$F1-$^{9}$F1 module pair of human fibronectin requires site-specific N-glycosylation. *FEBS Lett*, **579**, 4529-4534.

- Millard, C.J., Ellis, I.R., Pickford, A.R., Schor, A.M., Schor, S.L. and Campbell, I.D. (2007) The role of the fibronectin IGD motif in stimulating fibroblast migration. *J Biol Chem*, **276**, 177-187.

- Staunton, D., Millard, C.J., Aricescu, A.R. and Campbell, I.D. Preparation of recombinant fibronectin fragments for functional and structural studies. *Methods Mol Biol, (in press).*

# Abbreviations

- cFN - cellular fibronectin
- ECM - extracellular matrix
- ESI-MS - electrospray ionization mass spectrometry
- FN - fibronectin
- Fn1 - type I fibronectin module
- Fn2 - type II fibronectin module
- Fn3 - type III fibronectin module
- $^{n}$FX - n$^{th}$ type X module
- GBD - gelatin binding domain (proteolytically derived; native sequence)
- GlcNAc - N-acetylglucosamine
- Hex - $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 (recombinant; R503K, N497Q)
- IGD - Ile-Gly-Asp (amino acid motif)
- MMP - matrix metallo-proteinases
- NMR - nuclear magnetic resonance
- NOE - nuclear overhauser effect
- Quad - $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1 (recombinant; native sequence)
- pFN - plasma fibronectin
- RP-HPLC - reverse-phase high performance liquid chromatography
- SDS-PAGE - sodium dodecyl sulfate polyacrylamide gel electrophoresis
- SPR - surface plasmon resonance
- Triple - $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2 (recombinant; native sequence)

# Contents

**1. The Extracellular Matrix ..... 1**
1.1. Matrix composition ..... 1
1.2. Matrix proteins ..... 2
1.3. Fibronectin ..... 3
1.3.1. Modular architecture ..... 4
1.3.2. Domain structure ..... 5
1.3.3. Post-translational modification ..... 8
1.3.4. Fibril formation and matrix assembly ..... 10
1.4. Collagen ..... 12
1.5. Binding partners of fibronectin ..... 15
1.5.1. Collagen binding to fibronectin ..... 16
1.5.2. Module contribution to binding ..... 17
1.5.3. Biological significance of collagen binding to fibronectin ..... 18
1.6. Integrins ..... 19
1.6.1. Integrin activation by fibronectin ..... 21
1.7. Aims of this thesis ..... 22
1.8. References ..... 24

**2. Expression and purification ..... 29**
2.1. Protein expression systems ..... 29
2.1.1. Expression of recombinant proteins in *Pichia pastoris* ..... 30
2.1.2. Transient expression in mammalian HEK293T cells ..... 32
2.2. Methods ..... 34
2.2.1. PCR amplification ..... 34
2.2.2. Primer design ..... 35
2.2.3. Vector construction and gene fragment insertion ..... 36
2.2.4. Transformation of *E. coli* ..... 36
2.2.5. Preparation of *P. pastoris* cells for electroporation ..... 39
2.2.6. Preparation of DNA for transformation into *P. pastoris* ..... 39
2.2.7. Transformation of *P. pastoris* ..... 39
2.2.8. Small scale protein expression screens ..... 40
2.2.9. *P. pastoris* fermentation ..... 41
2.2.10. Harvesting the protein supernatant ..... 41
2.2.11. Cation exchange chromatography ..... 42
2.2.12. Trimming of high mannose oligosaccharides ..... 42
2.2.13. Preparative gelatin affinity chromatography ..... 43
2.2.14. Reverse-phase high performance liquid chromatography ..... 43
2.2.15. Transient protein expression ..... 43
2.2.16. Purification of transiently expressed protein ..... 44
2.2.17. Sodium dodecyl sulphate polyacrylamide gel electrophoresis ..... 44
2.2.18. Western blotting ..... 45
2.2.19. The 2100 Bioanalyzer System ..... 46
2.2.20. Electrospray ionization mass spectrometry ..... 46
2.2.21. N-terminal sequencing ..... 46
2.3. Results ..... 47
2.3.1. Design of constructs ..... 47
2.3.2. Gene fragment amplification and insertion into pPICZλ ..... 49

2.3.3. Transformation of *P. pastoris* ..... 49
2.3.4. *P. pastoris* fermentation ..... 49
2.3.5. Protein expression and purification ..... 50
2.3.6. Cation exchange chromatography ..... 51
2.3.7. Reverse-phase high performance liquid chromatography ..... 52
2.3.8. Affinity chromatography ..... 54
2.3.9. Mass spectrometry ..... 56
2.3.10. N-terminal sequencing ..... 57
2.3.11. Western blotting ..... 57
2.3.12. Transient mammalian protein expression ..... 58
2.3.13. Protein purification of transiently expressed Hex ..... 59
2.4. Discussion ..... 60
2.4.1. Vector design and cloning ..... 60
2.4.2. Protein expression, purification, and analysis ..... 60
2.5. References ..... 62

**3. Structural characterisation of fibronectin fragments ..... 64**
3.1. Protein structure at atomic resolution ..... 64
3.1.1. Protein NMR ..... 65
3.1.2. The basic NMR experiment ..... 65
3.1.3. One-dimensional $^{1}$H NMR ..... 67
3.1.4. Two-dimensional NMR ..... 68
3.1.5. COSY, TOCSY, and NOESY ..... 68
3.1.6. [$^{1}$H-$^{15}$N]-HSQC ..... 68
3.1.7. Heteronuclear {$^{1}$H-}$^{15}$N NOE ..... 69
3.1.8. Three-dimensional NMR ..... 70
3.1.9. NMR assignment ..... 70
3.2. Methods ..... 73
3.2.1. NMR spectroscopy ..... 73
3.2.2. NMR processing and analysis ..... 73
3.2.3. Data conversion ..... 73
3.2.4. Data processing ..... 74
3.2.5. Data analysis ..... 74
3.3. Results ..... 75
3.3.1. NMR analysis of $^{8}$Fn1**-$^{9}$Fn1 ..... 75
3.3.2. Sequential assignment of $^{8}$Fn1*-$^{9}$Fn1 ..... 77
3.3.3. Secondary structure of $^{8}$Fn1*-$^{9}$Fn1 ..... 80
3.3.5. NMR analysis of $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 ..... 81
3.3.6. NMR characterisation of the Hex at 950MHz ..... 83
3.3.7. Crystallisation trials of the Hex ..... 84
3.4. Discussion ..... 85
3.4.1. The recombinant fibronectin fragments are folded ..... 85
3.4.2. The N497Q mutation subtly alters the behaviour of the protein ..... 85
3.4.3. The $^{8}$Fn1**-$^{9}$Fn1 module pair contains typical Fn1 folds ..... 85
3.4.4. Further study following NMR assignment ..... 86
3.5. References ..... 87

**4. The interaction between collagen and fibronectin ..... 88**
4.1. Experimental design ..... 88
4.1.1. Binding properties of collagen and gelatin to Fibronectin ..... 88
4.1.2. Nature of the binding interaction ..... 89
4.1.3. Gelatin affinity chromatography ..... 91
4.1.4. Surface plasmon resonance ..... 92
4.1.5. Analysis of surface plasmon resonance data ..... 93
4.1.6. NMR titration ..... 94
4.1.7. Binding analysis through synthetic peptide NMR titration ..... 95
4.2. Methods ..... 96
4.2.1. Surface plasmon resonance ..... 96
4.2.2. Analytical gelatin affinity chromatography ..... 97
4.2.3. NMR collagen peptide titration ..... 97
4.2.4. NMR chemical shift mapping ..... 99
4.3. Results ..... 100
4.3.1. Characterisation of Hex by surface plasmon resonance ..... 100
4.3.2. Studying the property of glycosylation on fibronectin fragments ..... 102
4.3.3. Further fibronectin fragment binding properties ..... 107
4.3.4. NMR collagen peptide titration ..... 108
4.3.5. Collagen peptide interaction site on $^{8}$Fn1*-$^{9}$Fn1 ..... 111
4.4. Discussion ..... 113
4.4.1. Recombinantly expressed protein interacts with collagen ..... 113
4.4.2. Role of glycosylation on the gelatin binding domain ..... 114
4.4.3. Fibronectin fragments bind to a synthetic collagen peptide ..... 116
4.4.4. The role of the each module in the gelatin binding domain ..... 118
4.5. References ..... 121

**5. Stimulating fibroblast migration with fibronectin ..... 123**
5.1. Fibroblast migration ..... 123
5.1.1. The IGD motif ..... 124
5.2. Methods ..... 127
5.2.1. Protein production ..... 127
5.2.2. Cells ..... 127
5.2.3. Migration assays ..... 128
5.2.4. NMR spectroscopy ..... 128
5.3. Results ..... 129
5.3.1. The location of the IGD motif ..... 129
5.3.2. Expression, purification, and characterisation of IGD mutants ..... 130
5.3.3. An intact IGD motif is required to stimulate fibroblast migration ..... 132
5.3.4. The relative roles of IGD in $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 ..... 133
5.3.5. The IGD motifs in $^{7}$F1 and $^{9}$F1 can act independently ..... 134
5.4. Discussion ..... 136
5.4.1. Fibronectin fragments containing the IGD motif are bioactive ..... 136
5.4.2. Increased potency in the context of a folded protein ..... 137
5.4.3. A potential IGD signalling pathway ..... 138
5.4.4. The IGD motif appears multiple times in fibronectin ..... 140
5.4.5. The bioactive IGD motif may be cryptically hidden ..... 141
5.5. References ..... 143

**6. The gelatin binding domain ..... 145**
6.1. A multifaceted domain ..... 145
6.1.1. Successful expression of gelatin binding domain fragments ..... 145
6.2. Collagen binding in vivo ..... 146
6.2.1. Is collagen the ligand for fibronectin? ..... 147
6.2.2. Orientation of binding ..... 148
6.3. The role of the fibronectin IGD motif ..... 150
6.4. Cooperativity within the gelatin binding domain ..... 151
6.4.1. Future work ..... 152
6.5. References ..... 154

**7. Appendices ..... 155**
7.1. pPICZλ vector sequence ..... 155
7.2. Gelatin binding domain fibronectin sequence ..... 157
7.3. Backbone NH and Hα assignments for $^{8}$Fn1*-$^{9}$Fn1 ..... 159
7.4. Publications ..... 163
7.4.1. Gelatin binding to the $^{8}$Fn1-$^{9}$Fn1 module pair of human fibronectin requires site-specific N-glycosylation ..... 163
7.4.2. The role of the fibronectin IGD motif in stimulating fibroblast migration ..... 169
7.4.3. Preparation of recombinant fibronectin fragments for functional and structural studies ..... 175

# 1. The Extracellular Matrix

## 1.1. Matrix composition

The extracellular matrix (ECM) is the structurally stable network of proteins and polysaccharides that underlies epithelia and surrounds connective tissue cells. It provides a supporting surface for cells to adhere, migrate, and proliferate, and its properties are modulated by the composition and positioning of the component proteins (Figure 1:1). The ECM gives skin the unique properties of elasticity, flexibility, and strength; it gives support to structures such as bone and teeth, and can organise to form tendons with enormous tensile strength (Alberts et al., 2002), (Bosman and Stamenkovic, 2003).

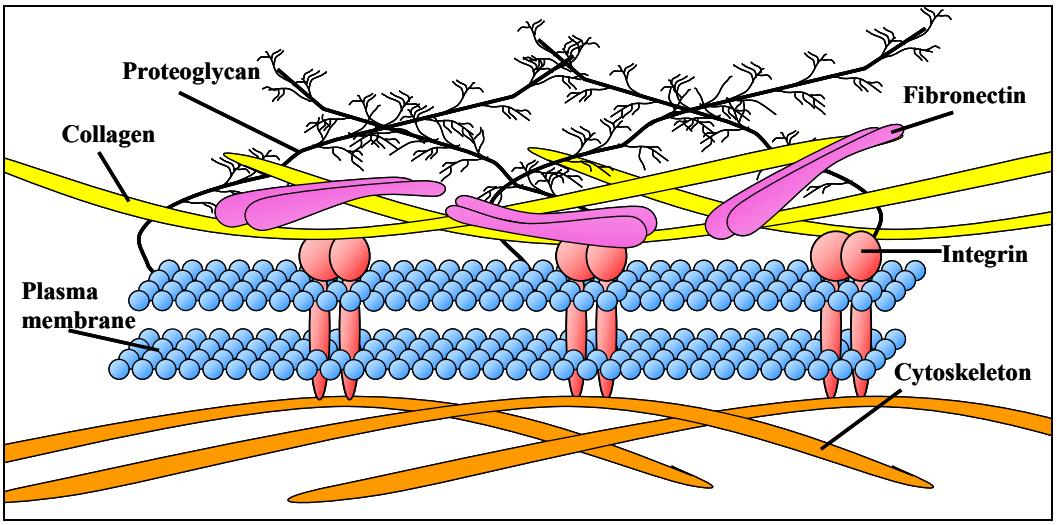



**Figure 1:1** A schematic diagram of the key extracellular matrix components. The external scaffold is linked to the intracellular cytoskeleton by integral membrane proteins (such as integrins) resulting in a cross-talk through the plasma membrane.

The ECM is far more than a structurally inert scaffold. It is able to influence the behaviour of cells by regulating migration, development, proliferation, shape, tissue

homeostasis and wound healing (Berrier and Yamada, 2007). The cross-talk and interaction between components within this intricate network is incredibly complex (Larsen et al., 2006). Although several of the more abundant components of the matrix (such as collagen and fibronectin) have been studied extensively, the subtlety of interaction between components is only just being realised, and much work is needed to understand the signalling mechanisms between cell and matrix.

The components of the ECM can be broadly divided into two classes: i) proteoglycans (e.g. aggrecan, versican, decorin, and biglycan), which are heavily glycosylated proteins, decorated with long polysaccharides known as glycosaminoglycans (e.g. chondroitin sulphate, dermatan sulfate, and heparan sulphate), and ii) fibrous proteins (e.g. collagen, elastin, laminin, and fibronectin), which lie outside the cell membrane in the extracellular space (Tanzer, 2006). Some members of the proteoglycan family are integral components of the plasma membrane, with the protein domain spanning the lipid bilayer, and the large polysaccharide chain filling the extracellular space. While some of the membrane spanning regions can act as co-receptors in cell signalling, the highly extended hydrophilic polysaccharides form a hydrated gel matrix, allowing the passage of fluids and nutrients around the cell. The fibrous proteins have both structural and adhesive properties which provide a framework for cells to attach and order themselves within a tissue (Alberts et al., 2002).

## 1.2. Matrix proteins

The ECM is composed of a highly diverse range of proteins which provides anchorage and support to cells but also allows cell migration within the matrix. They

form a wide range of structures in order to fulfil all the roles needed for these processes to occur. A common theme among extracellular proteins is their modular composition, each made up of a limited number of protein modules or repeating units, which fold up to create functional domains. These large proteins have been studied by dissection into smaller units of manageable size (Campbell and Downing, 1998). Fortunately many of the modules can fold independently (Bork et al., 1996) and the protein fragments often retain activity after proteolysis. Study of the full length protein would be possible with techniques such as atomic force microscopy and cryo-electron microscopy, but at the expense of lower resolution. Although modules may be repeated many times within one particular protein, it is impossible to predict their function from their structural homology (Hohenester and Engel, 2002). Their biological characteristics must be examined experimentally, within individual domains, to give insight into the characteristics of full length extracellular proteins (Potts and Campbell, 1994).

## 1.3. Fibronectin

Fibronectin (FN) is a large extracellular matrix glycoprotein which is involved in a range of important physiological processes such as embryogenesis, wound healing, haemostasis, and thrombosis (reviewed in (Hynes, 1990), (Yamada, 1983), (Pankov and Yamada, 2002)). This key fibrous protein exists in at least two forms: i) as a soluble disulphide-linked dimer in plasma, involved in blood coagulation through its affinity for fibrin and platelets, and ii) as an insoluble multimer in the extracellular matrix, which interacts with cell receptors to promote cellular adhesion and cytoskeletal organisation. FN was originally given a variety of different names such

as cold insoluble globulin, anti-gelatin factor, collagen-cell attachment protein (Hynes, 1990) due to its coprecipitation and interaction with many different molecules. FN participates in many complex interactions with other matrix proteins (e.g. collagen, heparin, fibrin, and fibulin) and cell receptors (integrins and syndecans). The role of FN as a key protein in development was confirmed when the FN mouse knockout was shown to cause early embryonic lethality (George et al., 1993).

### 1.3.1. Modular architecture

FN is a mosaic protein composed almost entirely of three types of module, known as Fn1, Fn2, and Fn3 (Potts and Campbell, 1996) (Figure 1:2). The modules are arranged into domains which act in combination to enable the protein to perform a variety of different functions. The domains of FN were initially characterised by studying fragments of partially digested native FN, but more recently studied in recombinantly expressed fragments. The function of each domain cannot be predicted solely from structural similarity, but a "divide and conquer" approach has been used to learn more about the domain properties of FN, and its interaction with other macromolecules.

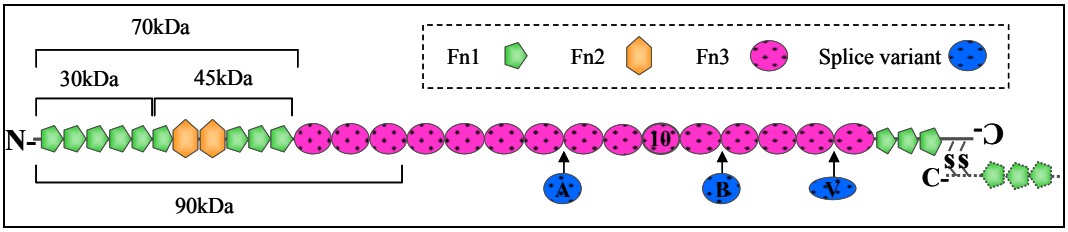



**Figure 1:2** Modular structure of FN. The positions of the Fn1, Fn2, and Fn3 modules are shown. The FN molecule is the product of a single gene but multiple forms can arise due to alternate splicing in the A, B, and V regions. The N-terminal binding domains resistant to proteolysis are labelled with their classical molecular weights.

Many of the FN modules have had their structure solved by crystallography either individually (Dickinson et al., 1994), in series (Rudino-Pinera et al., 2007), and/or their solution structure solved by NMR (Baron et al., 1990), (Pickford et al., 2001). The Fn1, Fn2, and Fn3 modules are composed of 45, 60, and 90 residues respectively. Whereas the Fn1 and Fn2 modules are held together by disulphide bonds in a 1-3, 2-4 configuration, the Fn3 modules do not require any intramolecular disulphide bonds. All three FN modules contain at least two short antiparallel β-sheets which fold over each other to form a central core. The FN module folds are found in other related proteins; Fn1 modules have been identified in tissue-type plasminogen activator (t-PA) and the blood coagulation factor XII; Fn2 modules have been found in the type IV collagenases and PDC-109 (phosphorylcholine-binding protein); the Fn3 module is found in numerous extracellular and intracellular proteins (Campbell and Spitzfaden, 1994). Although the tertiary structure of the Fn1, Fn2, and Fn3 module folds are highly conserved, the primary protein sequence homology is considerably lower than expected for this level of similarity, making prediction of function simply based on fold similarity particularly difficult. Discovering how similar modules can lend themselves to completely different functions is of great interest and a better understanding would aid future structure-function predictions.

### 1.3.2. Domain structure

The independently folded modules of FN were originally thought to exist as "beads on a string", with each module joined to the next by a short peptide linker, but they are now known to have a far more complex assembly. The modules appear to associate to form larger domains which give FN the ability to bind to itself, other extracellular matrix proteins, and cell-surface receptors.

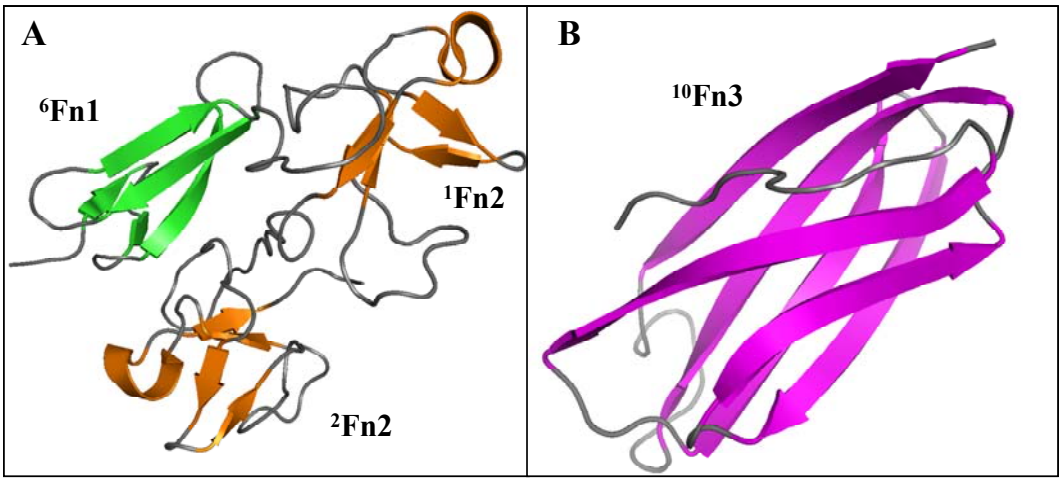



**Figure 1:3** Ribbon diagrams of modules from FN. A, the first three modules from the gelatin binding domain of FN (Fn1 = green, Fn2 = orange) (pdb code 1e8b, (Pickford et al., 2001)). B, a single Fn3 module taken from the cell binding domain (Fn3 = purple) (pdb code 1FNA, (Dickinson et al., 1994)). Diagrams drawn in Pymol.

The 30kDa N-terminal fragment is composed of five Fn1 modules, and was first isolated from full length FN by means of a double digest with cathepsin D and chymotrypsin, before purification on a heparin column (Hynes, 1990). As well as binding to heparin and fibrin, these five Fn1 modules can self associate with other regions of FN, and are thought to be critical in forming FN fibrils (Schwarzbauer, 1991). This domain is targeted by pathogens, such as *Staphylococcus aureus* and *Streptococcus pyogenes*, as a way of gaining entry to the human cell. Proteins, which show a high affinity for Fn1 modules, are displayed on the bacterial cell surface and provide a mechanism to aid internalisation and infection (Schwarz-Linek et al., 2003).

Adjacent to the 30kDa N-terminal fragment is a proteolytically stable domain of 45kDa which has the module composition $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1 (in this notation $^{n}$FX represents the n$^{th}$ type X module in the native protein). This domain has been termed the collagen or gelatin binding domain (GBD) (Balian et al., 1979) as

this fragment will bind both isolated collagen α chains (Ingham et al., 1988), (Guidry et al., 1990) and gelatin (heat-denatured collagen) (Hahn and Yamada, 1979), (Johansson and Hook, 1980). The GBD contains the only two Fn2 modules that occur within FN (Figure 1:3A). These modules display a solvent-exposed hydrophobic surface which is a possible binding site for collagen (Pickford et al., 1997). The Fn1 modules promote binding to collagen, both stabilising the domain by forming a hairpin core (Pickford et al., 2001), and contributing to the affinity of the collagen binding directly (Katagiri et al., 2003).

The Fn3 modules (Figure 1:3B) are all found C-terminal to the GBD. These modules provide further binding domains which promote FN binding to heparin (Ingham et al., 1993) and fibulin (Balbona et al., 1992), as well as directly to cells (Pierschbacher et al., 1981). Active units within these binding domains were first isolated through sequential proteolytic digestion of FN in order to identify the smallest active fragment in a process known as fragment mapping. In most cases these have been restricted to small proteolytic fragments, but in the most prominent example, the active fragment in the cell-binding domain was mapped to the short amino acid sequence Arg-Gly-Asp-Ser (RGDS) of the $^{10}$Fn3 (Pierschbacher et al., 1985). This motif in FN has been shown to promote cell attachment and some effects can be mimicked by the tetra-peptide RGDS (Pierschbacher and Ruoslahti, 1984); the RGDS sequence is highly conserved in FN from different species.

The Fn3 modules number 15-17 (depending on the alternative splicing pattern), and are followed by three further Fn1 modules, which bind to fibrin (Hayashi and

Yamada, 1983). Within the last 20 amino acids of FN there are two interchain disulphide bonds which are responsible for covalently linking the two FN polypeptides to form a dimer (Garcia-Pardo et al., 1984).

### 1.3.3. Post-translational modification

The structure of FN can be modified by alternative splicing of the molecule at three different points (Schwarzbauer et al., 1987). Two extra Fn3 modules and an extra linker "V" (variable in length) (see Figure 1:2) can be inserted to give rise to variant forms of FN. The two extra Fn3 repeats are most commonly found in cellular FN (cFN) as the splicing pattern of plasma FN (pFN) tends to be simpler. The addition of extra repeats in FN may act as a mechanism for controlling FN properties, either by introducing new interaction sites, or by modulating existing interactions with cell surface receptors such as integrins. For instance, the V region containing the cell binding LDV sequence is reintroduced into FN by alternative splicing to aid tissue repair (ffrench-Constant, 1995). Through alternative splicing of the pre-mRNA, 12 different forms of rat FN and up to 20 splice variants of human FN (Pankov and Yamada, 2002) have been isolated to date.

All known forms of FN are post-translationally modified to allow the protein to function correctly and to protect the protein from proteolysis. Certain modules are targeted to be disulphide bonded, sulphated, phosphorylated, and glycosylated (Paul and Hynes, 1984). There are between three and seven N- and O- linked sugars chains attached to FN which vary in carbohydrate composition and length depending on the species (Fukuda et al., 1982) and type of cell expressing the protein (Ingham et al.,

1995). The predominant forms of N-linked carbohydrates on human FN are complex-type biantennary glycans (Figure 1:4) which can constitute as much as 5% of the protein mass (Takasaki et al., 1980).

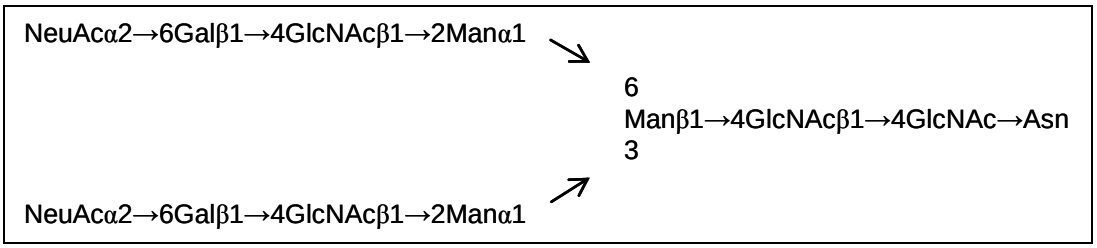



**Figure 1:4** The structure of the major N-glycans of human FN. GlcNAc = N-acetylglucosamine; Man = mannose; Gal = galactose; NeuAc = N-acetylneuraminic acid (adapted from (Takasaki et al., 1980)).

The carbohydrate moieties at the N-terminus of FN may fulfil a number of different roles with each glycosylation site potentially fulfilling a specialised role. One study investigated the overall role of carbohydrate on FN by inhibiting all glycosylation with tunicamycin and observed that FN became twice as sensitive to proteolytic hydrolysis as glycosylated protein (Olden et al., 1979). In a different study, the effect of the carbohydrate on the thermal stability was examined; this involved the complete removal of the attached sugar residues with Endo F in order to compare the native and deglycosylated fragments (Ingham et al., 1995).

The thermal stability of $^{8}$Fn1-$^{9}$Fn1 was shown to be greatly decreased after deglycosylation, lowering the melting temperature by around 20°C at pH 8.2; an even greater effect was observed as the pH was adjusted down to pH 7.0. In a more sophisticated study, the selective deglycosylation of certain carbohydrate moieties was assessed with respect to their role in maintaining the structure of a pair of Fn1 modules. One of the two carbohydrate moieties on the $^{8}$Fn1 module was shown to be

more critical in stabilising the protein fold than the other, although both sugars affected the ability of the protein to bind gelatin in solution (Millard et al., 2005).

The restricted appearance of carbohydrate on FN may also imply that the sugar chains have functional roles. The occupancy of the three potential N-linked glycosylation sites in the gelatin binding domain differs between tissue types (Zhu and Laine, 1985). While three N-linked chains are found on cFN, there is only an average of two N-linked chains on pFN, suggesting the presence of sugar residues may be deliberately altered between tissues to allow FN to perform different functional roles.

### 1.3.4. Fibril formation and matrix assembly

The structure of FN can be further modified by the assembly of soluble compact dimers into an insoluble extended fibrillar network in a cell dependent process. FN dimers bind to transmembrane integrin receptors, and perhaps by undergoing conformational changes, intermolecular binding sites are revealed which promote fibril formation (Mao and Schwarzbauer, 2005) (see Figure 1:5). The N-terminal $^{1}$Fn1-$^{5}$Fn1 "matrix assembly domain" is essential for fibril formation (Sottile and Mosher, 1997) and recombinant material lacking this domain is not active in fibril assembly (Schwarzbauer, 1991). The collagen binding domain may also be involved to some extent, as the N-terminal 70kDa proteolytic fragment of FN (containing both domains) was shown to block FN matrix assembly (McKeown-Longo and Mosher, 1985), and polyclonal antibodies to the collagen binding domain also inhibited fibril formation (McDonald et al., 1987). FN must display the RGD sequence on the $^{10}$Fn3 to allow interaction with integrins, as monoclonal antibodies to the cell binding region

blocked matrix assembly by cultured fibroblasts (McDonald et al., 1987). In addition, FN must be present as a disulphide bonded dimer for incorporation into fibrils (Schwarzbauer, 1991).

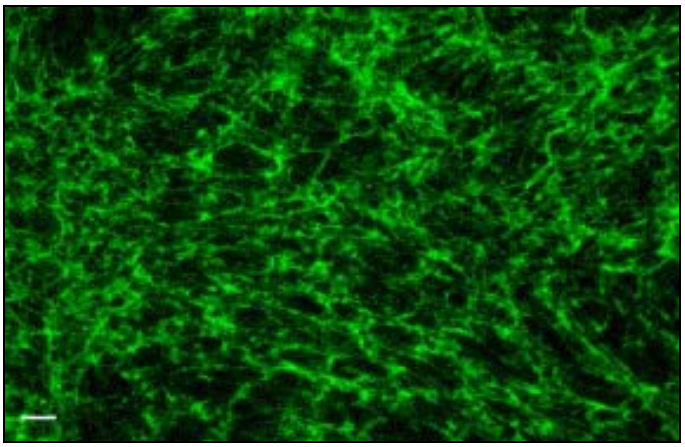


**Figure 1:5** The assembly of a FN matrix by Chinese Hamster Ovary (CHO) cells over a 16h period visualised in real time by immunofluorescence. FN fibrils were detected with an anti-rat FN monoclonal antibody and visualised by fluorescein-conjugated secondary antibody. The scale bar = 20 μm (adapted from (Mao and Schwarzbauer, 2005)).

Fibrillogenesis is highly regulated by both intracellular signalling components and by the extracellular environment. FN does not form fibrils spontaneously from purified components, as the dimer is folded into a conformation that does not undergo fibril formation, but instead requires components from both sides of the cell membrane to come together to coordinate the cell mediated process (Wierzbicka-Patynowski and Schwarzbauer, 2003). The initiation step of fibrillogenesis is still unclear but it is thought to require the revealing of "cryptic sites" on FN which are hidden until activation. Two models exist; either FN binds to integrins on the cell surface, via the RGD loop on $^{10}$Fn3, leading to a tension-induced conformational change that allows FN-FN interactions via the N-terminal region, or, the N-terminal region undergoes a conformational change that allows FN to bind to integrins on the cell surface via its RGD loop. There are four known sites on FN which can associate during

fibrillogenesis; the N-terminal $^{1}$Fn1-$^{5}$Fn1 has been shown to interact with both $^{1}$Fn3-$^{2}$Fn3 (Aguirre et al., 1994) and $^{12}$Fn3-$^{13}$Fn3-$^{14}$Fn3 (Bultmann et al., 1998), the $^{1}$Fn3 module can interact with $^{7}$Fn3 (Ingham et al., 1997), and the $^{2}$Fn3-$^{3}$Fn3 module has been shown to interact with $^{12}$Fn3-$^{13}$Fn3-$^{14}$Fn3 (Johnson et al., 1999). These multiple FN-FN interaction sites could give rise to variable alignments of the fibrils, which would bring the other binding domains (collagen, heparin, cells, etc.) into different environments, and would add to the complexity of fibril formation (Wierzbicka-Patynowski and Schwarzbauer, 2003).

## 1.4. Collagen

The collagens are a family of highly characteristic extracellular matrix proteins present in all multicellular animals, sometimes constituting more than 25% of the total protein in adult organisms (Alberts et al., 2002). The collagen scaffold forms the major structural element of all connective tissues, and acts directly or indirectly as an agent to promote cell attachment and differentiation.

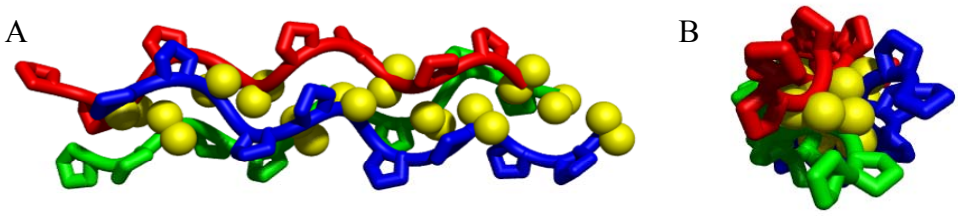



**Figure 1:6** The collagen triple helix. A, side on view. B, end on view. The three helices (red, green and blue) intertwine to produce a right-handed superhelix with Pro and Hyp (hydroxyproline) side chains shown in their respective backbone chain colours. αHydrogen atoms attached to Gly are shown in yellow (pdb code 1Q7D (Emsley et al., 2004)). Diagram produced in MolMol.

More than 20 genetically-distinct types of human collagen have been identified (Bosman and Stamenkovic, 2003) with 28 named entries to date in the Swiss Prot

database (http://expasy.org). These vary greatly in their size, function, and tissue distribution. Almost 90 percent of the family is represented by the fibril-forming collagens (types I, II, III, V and XI) which are able to form rope-like fibres of high tensile strength (Hulmes, 2002). The other collagen proteins have very diverse structures and functions. For example, type IV collagens are able to form loosely woven fibres that are flexible enough to allow contraction and expansion. Other collagens are able to associate with the membrane and form anchoring fibrils and membrane proteins (Gelse et al., 2003).

The collagens are not modular in the same sense as FN but are characterized by molecular rod-like triple helices interspersed by globular domains (van der Rest and Garrone, 1990). Each chain is composed of a repeating tripeptide sequence of Gly-x-y, where x is often Pro and y is often Hyp (4-hydroxyproline), which forms a left-handed helix. Three left-handed polypeptides (two α1(I) chains and one α2(I) chain) are supercoiled together in a right-handed manner to form a long, stiff, triple helical structure with 3.3 amino acid residues per turn (Traub and Piez, 1971), (Fraser et al., 1979) (Figure 1:6). The centre of the triple helix is held together by an extensive

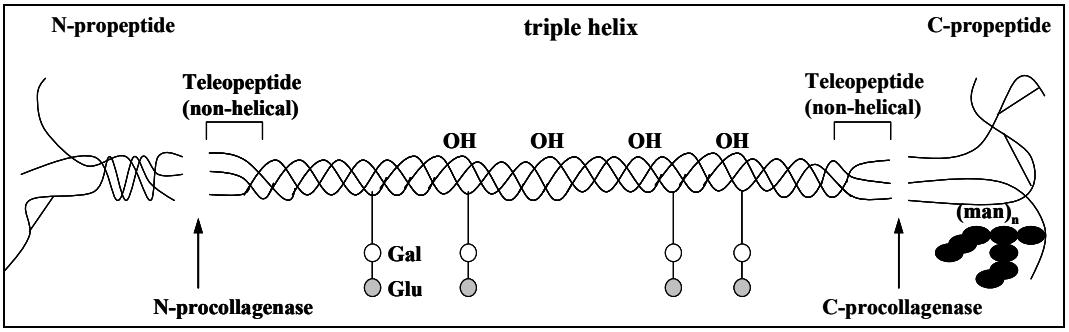



**Figure 1:7** The molecular structure of fibrillar collagen. The pro-peptide is cleaved by procollagen proteases to leave exposed non-helical ends. The terminal extension peptides are able to form intra- and intermolecular cross-links which stabilise the molecular arrangement within fibrils (adapted from (Gelse et al., 2003)).

network of hydrogen bonds, and the relatively inflexible Pro and Hyp residues confer rigidity on the collagen structure (Bella et al., 1994). Some collagens are able to form fibrils. At the extremities of these highly-ordered structures are non-helical regions at the $NH_2$ and COOH-terminal ends that are far more susceptible to proteolytic digestion and play an important functional role in fibril formation (Traub and Piez, 1971) (Figure 1:7).

The triple helices spontaneously aggregate to form fibrils during cleavage by procollagen proteinases as they are secreted into the extracellular matrix (Hulmes, 2002). The fibrils are thin structures of up to ~300nm in diameter and many hundreds of micrometers in length. They provide a scaffold for the attachment of other matrix molecules and cells, thereby conferring connectivity, long range structure, and tensile strength in tissues (Alberts et al., 2002).

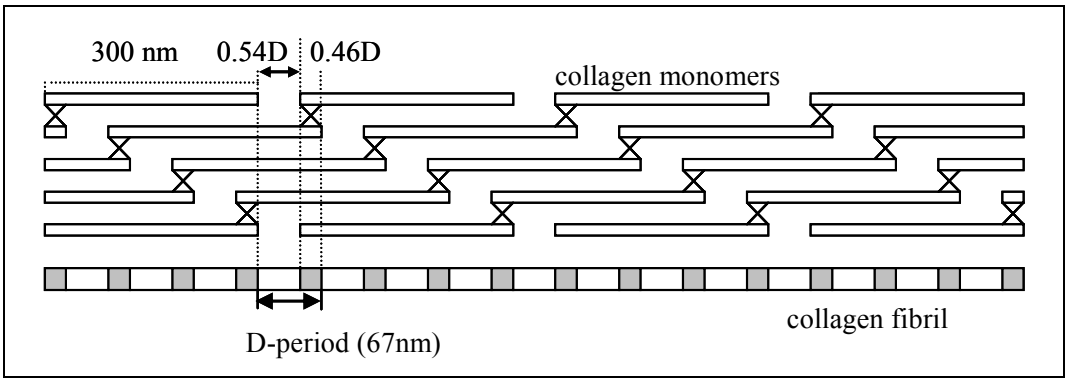



**Figure 1:8** The axially-staggered array present in fibrillar collagens. Triple helical fibrils are formed spontaneously upon cleavage of the propeptides from the N- and C- termini of procollagen monomers. The fibrils have been shown, using negative staining microscopy, to have a basic repeating unit (D = 67nm) of gap and overlap regions.

The classic fibril-forming collagens can assemble into highly oriented axially-staggered fibril arrays by forming intramolecular and intermolecular cross-links at

their non-triple-helical ends (Figure 1:8). Individual molecules are approximately 4.4 times the length of the repeat period with gaps of 0.54D between the molecular ends (Orgel et al., 2001). This arrangement results in maximal electrostatic and hydrophobic interactions (Brodsky and Ramshaw, 1997) and allows specific Lys and Hyl (5-hydroxylysine) residues in the helical and non-helical regions within the 0.46D overlap region to form stable cross-links (Orgel et al., 2001).

Many models have been proposed to explain how the collagen fibrils form into a lateral supermolecular structure, but even with much X-ray diffraction, electron microscopy, and biochemical data, no consensus structure has emerged (Orgel et al., 2001). The specific packing arrangement of collagen fibrils *in situ* has been difficult to define due to the high degree of disorder in the gap regions of electron density maps. In the most detailed study of fibrils to date, X-ray fibre diffraction data was combined with computer simulation, to establish a quasihexagonally packing model of the collagen molecules (Orgel et al., 2006). The new structural information has offered insight into collagen binding by other ECM proteins such as FN and collagenases.

## 1.5. Binding partners of fibronectin

FN has many binding partners in the extracellular matrix (e.g. heparin, fibrin, integrins etc), but in particular, it can bind specifically to members of the collagen family. The 45kDa proteolytic fragment can be isolated through proteolytic digestion of whole FN and then be purified by single-step affinity chromatography on gelatin-sepharose or gelatin-agarose (Ruoslahti et al., 1982). Detailed protocols for purifying

both full length FN and the GBD by gelatin affinity chromatography date back to 1977 (Engvall and Ruoslahti, 1977).

### 1.5.1. Collagen binding to fibronectin

FN will bind to both native collagen and denatured collagen (gelatin), but binds with 10-100 times higher affinity to the latter (Hynes 1990). The binding of collagen/gelatin to FN involves a mixture of electrostatic and hydrophobic forces (Vuento et al., 1982, 1983), but not enough is known about the nature of these interactions to explain the preferential affinity for gelatin. Experimental evidence suggests that the strong-binding sites are located on the more highly ordered collagenous domains rather than in the globular regions (Ingham et al., 1985). Some studies have suggested that the triple-helical domains must unwind slightly to bind FN (Engvall et al., 1981). It is possible that the binding domains exist in a much more open form in gelatin and are therefore more accessible for binding.

Local unwinding of the triple helix is known to be part of the matrix metallo-proteinases-1 (MMP-1) cleavage mechanism, where the collagen is only cleaved at one specific site, once the triple helix has relaxed (Chung et al., 2004). This region is less rich in Pro and Hyp; this lowers the stability of the triple helix and probably increases the likelihood of unwinding at this point. Exposing hidden binding sites by the local unwinding of the collagen helix may also be part of the FN binding mechanism since FN and collagenase MMP-1 have very similar collagen binding properties and both bind through Fn2 modules. This would suggest that unwound regions of collagen preferentially bind FN and other binding partners.

These cryptic sites may be exposed during tissue damage and could serve as a mechanism to initiate wound healing. Plasma FN levels increase soon after wounding (Singer and Clark, 1999), and apart from promoting adhesion and migration of cells to repair the damaged area, FN could play a role in binding to and removing collagenous debris (La Celle et al., 1990). This unwinding mechanism may be used in other processes such as tissue remodelling where FN is recruited during collagen degradation only when cryptic sites are revealed to trigger the process. There is a cryptic site in the triple helical region of type VI collagen which is required for angiogenesis *in vivo* (Xu et al., 2001). The binding of a monoclonal antibody (HUIV26) blocks exposure of this site and potently inhibits angiogenesis and tumour growth. FN cryptic sites are also thought to contribute to the mechanism of fibril formation and are revealed by conformational change during matrix assembly (Hocking et al., 1994).

### 1.5.2. Module contribution to binding

The collagen binding properties of FN were first attributed to the two Fn2 modules which are unique to the gelatin binding domain of FN. These modules share homology with other collagen binding proteins and have been shown to bind independently to gelatin (Banyai et al., 1990). The Fn2 modules from several collagenases (e.g. MMP-2 and MMP-9) also bind to gelatin and the binding sites have been mapped through alanine scanning mutagenesis (Collier et al., 1992). A solvent exposed hydrophobic pocket has been identified as a possible interaction site which includes two charged residues (Arg34 and Asp36). These residues are absolutely conserved in FN and MMP Fn2 modules, which could form a surface salt bridge, and may contribute directly to gelatin binding (Pickford et al., 1997).

The Fn2 modules can be shown to enhance binding to gelatin, but are not critical, as isolated Fn1 modules from this domain will also bind to gelatin. Proteolytic cleavage of the 45kDa collagen binding domain produces three non-overlapping module pairs ($^{6}$Fn1-$^{1}$Fn2, $^{2}$Fn2-$^{7}$Fn1, and $^{8}$Fn1-$^{9}$Fn1), each of which can bind to gelatin to some degree. The affinity is 10-20 fold weaker than that of whole FN (Ingham et al., 1989) which suggests that all six modules must act in concert to bind gelatin with maximum affinity. Gelatin affinity studies have shown that increasing the number of modules in recombinant fragments of this domain enhances the binding affinity for gelatin (Katagiri et al., 2003). The $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2 modules have been previously shown to form a globular hairpin-like structure (Pickford et al., 2001) and this may provide the nucleus around which the other modules fold to enhance gelatin binding.

### 1.5.3. Biological significance of collagen binding to fibronectin

Despite the growing evidence to support a direct interaction between collagen and FN in the extracellular matrix, the biological significance of this interaction is still routinely questioned. A recent study into the lethal mutations in osteogenesis imperfecta has highlighted possible crucial interactions between FN and collagen (Marini et al., 2007). Although this interaction has been comprehensively demonstrated by light microscopy (Vitellaro-Zuccarello et al., 1992), (Akhtar et al., 1999), this could simply be a fortuitous interaction, and further evidence for an interaction *in vivo* is needed. But as FN is the only major component from plasma to bind to gelatin-sepharose, there is a high probability that two such abundant proteins in plasma and the extracellular matrix would interact. Therefore, it remains a priority to investigate the binding properties of collagen and FN further. Here, in agreement

with many others, we primarily use gelatin as the main binding partner and working model, but the potential for binding to collagen itself is always borne in mind.

## 1.6. Integrins

Integrins are large modular cell surface receptors that are able to signal bidirectionally across the lipid bilayer. They modulate interactions between the cytoskeleton and the extracellular matrix, and play a critical role in adhesion, migration, differentiation, and survival (Xiong et al., 2007). Integrins have been shown to bind to a wide range of extracellular ligands such as ECM proteins (e.g. collagen, tenascin, fibrillin) and counter receptors (e.g. ICAM (intercellular adhesion molecule), E-cadherin, VCAM (vascular cell adhesion molecule)), as well as components of the cytoskeleton (e.g. talin, filamin, parvin, vinculin) (Humphries et al., 2006). Integrins are heterodimers composed of two non-covalently linked transmembrane glycoprotein subunits (Figure 1:9). In mammals, one of eight β-subunits and one of 18 α-subunits associate to form 24 distinct receptors (Alberts et al., 2002).

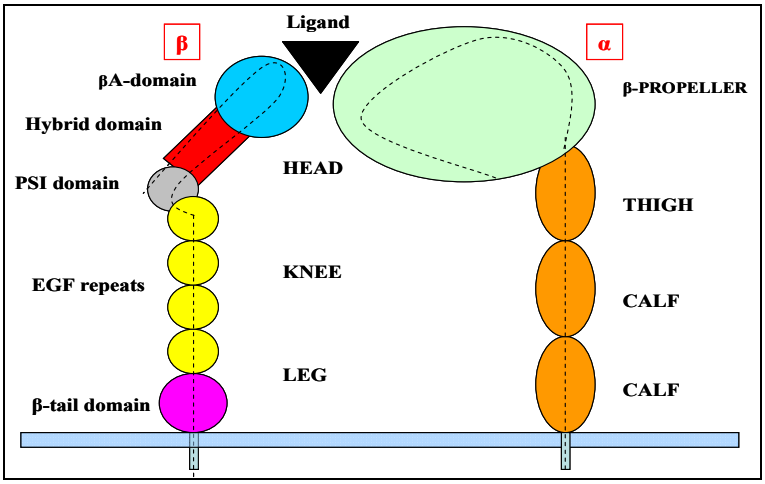



**Figure 1:9** Schematic representation of the α-subunit and β-subunits of integrins. Ligands can bind between the β-propeller and βA-domain in the integrin headpiece (adapted from (Mould and Humphries, 2004)).

Integrins transmit signals by integrating chemical and mechanical cues, and form a direct and dynamic link between proteins on either side of the membrane (Arnaout, 2002), (Hynes, 2002). The ability to recognise a broad range of ligands allows the integrin to orchestrate multiple signals and bring about a variety of responses. This property is balanced with safeguards, such as an "OFF" conformation in which they do not bind ligands and do not signal, thereby blocking unnecessary activation (Hynes, 2002).

Initially integrins were thought to exist in three conformations; inactive, primed, and ligand-activated, but there appear to be many intermediate states which allow a more subtle form of signalling (Mould and Humphries, 2004). Integrins can be triggered by inside-out signals which release intracellular constraints and are translated into large tertiary and quaternary structural changes through the protein to the ECM. The reverse can also occur as ligand binding to the extracellular part can cause outside-in signalling by making further conformational changes to the integrin structure.

Most of the integrins are linked through their β-tails to the actin cytoskeleton either directly or through intermediary proteins such as talin, α-actinin, and filamin. In response to integrin activation by extracellular ligand occupancy and clustering, these intracellular proteins act with G-coupled or kinase receptors to initiate signal transduction cascades. The transduction cascades ultimately influence cytoskeletal organisation, cell motility, and cell proliferation.

### 1.6.1. Integrin activation by fibronectin

Integrins are able to bind to a diverse range of ligands in the extracellular matrix (various collagens, fibronectins, vitronectin, laminins, von Willebrand factor and thrombospondins). These ligands often have an exposed aspartic acid or glutamic acid residue, found within an extended flexible loop, which is critical for integrin recognition. The Arg-Gly-Asp (RGD) sequence of the $^{10}$Fn3 module of FN was the first recognised integrin recognition signal (Ruoslahti and Pierschbacher, 1987) and has since been crystallised as a cyclic peptide in complex with the integrin αvβ3 (Xiong et al., 2001). Through a combination of cell biology and genetic analysis, other aspartic acid based sequences such as LDV and REDV have been identified in FN, which bind to and activate integrins (Ruoslahti, 1996).

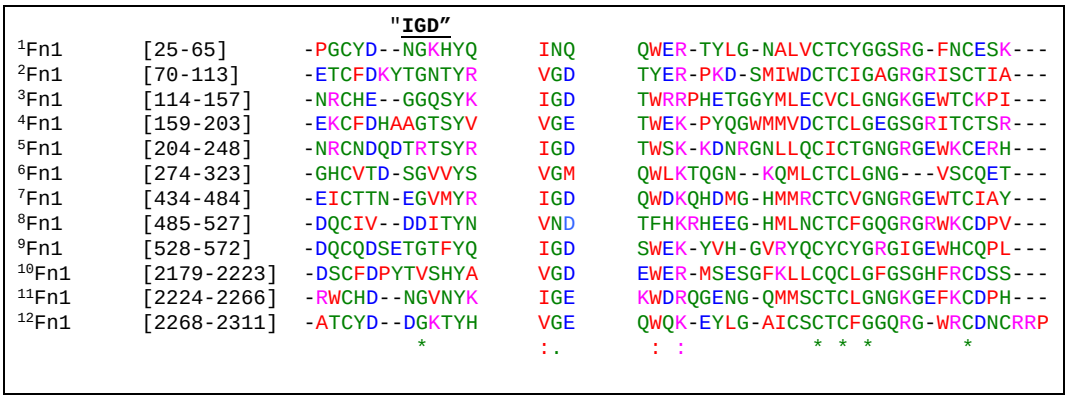



**Figure 1:10** Multiple sequence alignment of Fn1 modules from human FN ($^{1}$Fn1-$^{12}$Fn1) with the residue numbers for each module shown (mature FN). The relative position of the IGD motif is indicated. Red = small + hydrophobic + Y; blue = acidic; magenta = basic; green = hydroxyl + amine + basic + Q; '*' = identical in all; ':' = conserved in all; '.' = semi-conserved in all. This alignment was generated using the Swiss-Prot protein database (http://expasy.org) and ClustalW multiple sequence alignment (blosum matrix) (http://www.ebi.ac.uk).

A possible further integrin recognition sequence is the amino acid motif Ile-Gly-Asp (IGD) motif, which is a highly conserved feature in FN, being found in 4/9 of the amino terminal Fn1 modules (Figure 1:10). The position of this motif within the modules has been demonstrated to be at the apex of the main type I loop (Baron et al.,

1990). IGD peptides have been shown to stimulate a migratory response in fibroblasts. This response is thought to be elicited by a signal transduction cascade which, once activated, initiates a persistent alteration in cell phenotype (Schor et al., 1999). These studies showed that stimulation of fibroblast migration by IGD peptides is dependent on both αvβ3 functionality and tyrosine phosphorylation of focal adhesion kinase. Whether cell activation by IGD peptides occurs by a direct integrin binding, or via an intermediary protein, remains to be determined.

## 1.7. Aims of this thesis

FN is a key protein in the extracellular matrix and is able to interact with many different binding partners. Knowledge of this molecule's structure is a crucial step in understanding these interactions and will inevitably aid interpretation of biological data. The original "beads on a string" model for FN has been superseded by data which indicates that the structure of FN is much more complex. The modules of FN assemble into carefully defined molecular arrangements, which promote interactions internally within the protein, and externally with various extracellular ligands.

In this study, the gelatin binding domain of FN is explored at the molecular level in order to understand how the structural assembly of the domain affects its function. Solution state NMR is used to characterise the protein fold and detect the molecular environment of each amino acid under conditions close to physiological. Structural changes caused by single amino acid substitutions are monitored and used to show that the protein fold is not significantly perturbed. Those residues directly or indirectly involved in ligand binding are detected and data are collected to allow

calculation of binding constants. Complementary techniques, such as surface plasmon resonance and differential scanning calorimetry support the NMR results.

Having characterised the structure of the protein, the biological functions are explored in more detail, with the knowledge that the protein is structurally intact. Gelatin affinity chromatography is used to assess the behaviour of the protein in solution and characterise its interactions with a natural ligand. The design and interpretation of cell assays is made more straightforward with the available structural information. The information gained through working at both the molecular level and cellular level can only add to our understanding of the gelatin binding domain of FN.

Through a dissection approach, this study explores the structure and function of different length fragments of the gelatin binding domain of FN, with the aim of building up a working model for the whole domain. Fragments are expressed in *Pichia pastoris*, purified to homogeneity, and characterised by SDS PAGE, mass spectroscopy, and N-terminal sequencing (chapter 2). The FN fragments are structurally characterised by NMR (chapter 3) and functionally characterised in their ability to bind gelatin by gelatin affinity chromatography (chapter 4). Within chapter 4, the role of the glycosylation in gelatin binding is assessed, and direct binding to collagen is examined by NMR through the use of short collagen peptides. A motif within the FN fragments is explored for its ability to stimulate fibroblast migration (chapter 5). Finally, the role of the whole gelatin binding domain of FN, within the extracellular matrix, is discussed (chapter 6).

## 1.8. References


Aguirre, K.M., McCormick, R.J. and Schwarzbauer, J.E. (1994) Fibronectin self-association is mediated by complementary sites within the amino-terminal one-third of the molecule. *J Biol Chem*, **269**, 27863-27868.

Akhtar, S., Meek, K.M. and James, V. (1999) Immunolocalization of elastin, collagen type I and type III, fibronectin, and vitronectin in extracellular matrix components of normal and myxomatous mitral heart valve chordae tendineae. *Cardiovasc Pathol*, **8**, 203-211.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Watson, J. (2002) *Molecular Biology of the Cell*. Garland Science, New York, NY.

Arnaout, M.A. (2002) Integrin structure: new twists and turns in dynamic cell adhesion. *Immunol Rev*, **186**, 125-140.

Balbona, K., Tran, H., Godyna, S., Ingham, K.C., Strickland, D.K. and Argraves, W.S. (1992) Fibulin binds to itself and to the carboxyl-terminal heparin-binding region of fibronectin. *J Biol Chem*, **267**, 20120-20125.

Balian, G., Click, E.M., Crouch, E., Davidson, J.M. and Bornstein, P. (1979) Isolation of a collagen-binding fragment from fibronectin and cold-insoluble globulin. *J Biol Chem*, **254**, 1429-1432.

Banyai, L., Trexler, M., Koncz, S., Gyenes, M., Sipos, G. and Patthy, L. (1990) The collagen-binding site of type-II units of bovine seminal fluid protein PDC-109 and fibronectin. *Eur J Biochem*, **193**, 801-806.

Baron, M., Norman, D., Willis, A. and Campbell, I.D. (1990) Structure of the fibronectin type 1 module. *Nature*, **345**, 642-646.

Bella, J., Eaton, M., Brodsky, B. and Berman, H.M. (1994) Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science*, **266**, 75-81.

Berrier, A.L. and Yamada, K.M. (2007) Cell-matrix adhesion. *J Cell Physiol*.

Bork, P., Downing, A.K., Kieffer, B. and Campbell, I.D. (1996) Structure and distribution of modules in extracellular proteins. *Q Rev Biophys*, **29**, 119-167.

Bosman, F.T. and Stamenkovic, I. (2003) Functional structure and composition of the extracellular matrix. *J Pathol*, **200**, 423-428.

Brodsky, B. and Ramshaw, J.A. (1997) The collagen triple-helix structure. *Matrix Biol*, **15**, 545-554.

Bultmann, H., Santas, A.J. and Peters, D.M. (1998) Fibronectin fibrillogenesis involves the heparin II binding domain of fibronectin. *J Biol Chem*, **273**, 2601-2609.

Campbell, I.D. and Downing, A.K. (1998) NMR of modular proteins. *Nat Struct Biol*, **5 Suppl**, 496-499.

Campbell, I.D. and Spitzfaden, C. (1994) Building proteins with fibronectin type III modules. *Structure*, **2**, 333-337.

Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J.L., Fields, G.B., Visse, R. and Nagase, H. (2004) Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *Embo J*, **23**, 3020-3030.

Collier, I.E., Krasnov, P.A., Strongin, A.Y., Birkedal-Hansen, H. and Goldberg, G.I. (1992) 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.

Dickinson, C.D., Veerapandian, B., Dai, X.P., Hamlin, R.C., Xuong, N.H., Ruoslahti, E. and Ely, K.R. (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J Mol Biol*, **236**, 1079-1092.

Emsley, J., Knight, C.G., Farndale, R.W. and Barnes, M.J. (2004) Structure of the integrin alpha2beta1-binding collagen peptide. *J Mol Biol*, **335**, 1019-1028.

Engvall, E. and Ruoslahti, E. (1977) Binding of soluble form of fibroblast surface protein, fibronectin, to collagen. *Int J Cancer*, **20**, 1-5.

ffrench-Constant, C. (1995) Alternative splicing of fibronectin--many different proteins but few different functions. *Exp Cell Res*, **221**, 261-271.

Fraser, R.D., MacRae, T.P. and Suzuki, E. (1979) Chain conformation in the collagen molecule. *J Mol Biol*, **129**, 463-481.

Fukuda, M., Levery, S.B. and Hakomori, S. (1982) Carbohydrate structure of hamster plasma fibronectin. Evidence for chemical diversity between cellular and plasma fibronectins. *J Biol Chem*, **257**, 6856-6860.

Garcia-Pardo, A., Pearlstein, E. and Frangione, B. (1984) Primary structure of human plasma fibronectin--characterization of the 6,000 dalton C-terminal fragment containing the interchain disulfide bridges. *Biochem Biophys Res Commun*, **120**, 1015-1021.

Gelse, K., Poschl, E. and Aigner, T. (2003) Collagens - structure, function, and biosynthesis. *Adv Drug Deliv Rev*, **55**, 1531-1546.

George, E.L., Georges-Labouesse, E.N., Patel-King, R.S., Rayburn, H. and Hynes, R.O. (1993) Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. *Development*, **119**, 1079-1091.

Guidry, C., Miller, E.J. and Hook, M. (1990) A second fibronectin-binding region is present in collagen alpha chains. *J Biol Chem*, **265**, 19230-19236.

Hahn, L.H. and Yamada, K.M. (1979) Identification and isolation of a collagen-binding fragment of the adhesive glycoprotein fibronectin. *Proc Natl Acad Sci U S A*, **76**, 1160-1163.

Hayashi, M. and Yamada, K.M. (1983) Domain structure of the carboxyl-terminal half of human plasma fibronectin. *J Biol Chem*, **258**, 3332-3340.

Hocking, D.C., Sottile, J. and McKeown-Longo, P.J. (1994) Fibronectin's III-1 module contains a conformation-dependent binding site for the amino-terminal region of fibronectin. *J Biol Chem*, **269**, 19183-19187.

Hohenester, E. and Engel, J. (2002) Domain structure and organisation in extracellular matrix proteins. *Matrix Biol*, **21**, 115-128.

Hulmes, D.J. (2002) Building collagen molecules, fibrils, and suprafibrillar structures. *J Struct Biol*, **137**, 2-10.

Humphries, J.D., Byron, A. and Humphries, M.J. (2006) Integrin ligands at a glance. *J Cell Sci*, **119**, 3901-3903.

Hynes, R. (1990) *Fibronectins*. Springer-Verlag, Berlin, Germany.

Hynes, R.O. (2002) Integrins: bidirectional, allosteric signaling machines. *Cell*, **110**, 673-687.

Ingham, K.C., Brew, S.A., Huff, S. and Litvinovich, S.V. (1997) Cryptic self-association sites in type III modules of fibronectin. *J Biol Chem*, **272**, 1718-1724.

Ingham, K.C., Brew, S.A. and Isaacs, B.S. (1988) Interaction of fibronectin and its gelatin-binding domains with fluorescent-labeled chains of type I collagen. *J Biol Chem*, **263**, 4624-4628.

Ingham, K.C., Brew, S.A., Migliorini, M.M. and Busby, T.F. (1993) Binding of heparin by type III domains and peptides from the carboxy terminal hep-2 region of fibronectin. *Biochemistry*, **32**, 12548-12553.

Ingham, K.C., Brew, S.A. and Novokhatny, V.V. (1995) Influence of carbohydrate on structure, stability, and function of gelatin-binding fragments of fibronectin. *Arch Biochem Biophys*, **316**, 235-240.

Johansson, S. and Hook, M. (1980) Heparin enhances the rate of binding of fibronectin to collagen. *Biochem J*, **187**, 521-524.

Johnson, K.J., Sage, H., Briscoe, G. and Erickson, H.P. (1999) The compact conformation of fibronectin is determined by intramolecular ionic interactions. *J Biol Chem*, **274**, 15473-15479.

Katagiri, Y., Brew, S.A. and Ingham, K.C. (2003) All six modules of the gelatin-binding domain of fibronectin are required for full affinity. *J Biol Chem*, **278**, 11897-11902.

La Celle, P., Blumenstock, F.A., McKinley, C., Saba, T.M., Vincent, P.A. and Gray, V. (1990) Blood-borne collagenous debris complexes with plasma fibronectin after thermal injury. *Blood*, **75**, 470-478.

Larsen, M., Artym, V.V., Green, J.A. and Yamada, K.M. (2006) The matrix reorganized: extracellular matrix remodeling and integrin signaling. *Curr Opin Cell Biol*, **18**, 463-471.

Marini, J.C., Forlino, A., Cabral, W.A., Barnes, A.M., San Antonio, J.D., Milgrom, S., Hyland, J.C., Körkkö, J., Prockop, D.J., De Paepe, A., Coucke, P., Symoens, S., Glorieux, F.H., Roughley, P.J., Lund, A.M., Kuurila-Svahn, K., Hartikka, H., Cohn, D.H., Krakow, D., Mottes, M., Schwarze, U., Chen, D., Yang, K., Kuslich, C., Troendle, J., Dalgleish, R. and Byers, P.H. (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat*, **28,** 209-21.

Mao, Y. and Schwarzbauer, J.E. (2005) Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biol*, **24**, 389-399.

McDonald, J.A., Quade, B.J., Broekelmann, T.J., LaChance, R., Forsman, K., Hasegawa, E. and Akiyama, S. (1987) Fibronectin's cell-adhesive domain and an amino-terminal matrix assembly domain participate in its assembly into fibroblast pericellular matrix. *J Biol Chem*, **262**, 2957-2967.

McKeown-Longo, P.J. and Mosher, D.F. (1985) Interaction of the 70,000-mol-wt amino-terminal fragment of fibronectin with the matrix-assembly receptor of fibroblasts. *J Cell Biol*, **100**, 364-374.

Millard, C.J., Campbell, I.D. and Pickford, A.R. (2005) Gelatin binding to the $^{8}$Fn1-$^{9}$Fn1 module pair of human fibronectin requires site-specific N-glycosylation. *FEBS Lett*, **579**, 4529-4534.

Mould, A.P. and Humphries, M.J. (2004) Regulation of integrin function through conformational complexity: not simply a knee-jerk reaction? *Curr Opin Cell Biol*, **16**, 544-551.

Olden, K., Pratt, R.M. and Yamada, K.M. (1979) Role of carbohydrate in biological function of the adhesive glycoprotein fibronectin. *Proc Natl Acad Sci U S A*, **76**, 3343-3347.

Orgel, J.P., Irving, T.C., Miller, A. and Wess, T.J. (2006) Microfibrillar structure of type I collagen in situ. *Proc Natl Acad Sci U S A*, **103**, 9001-9005.

Orgel, J.P., Miller, A., Irving, T.C., Fischetti, R.F., Hammersley, A.P. and Wess, T.J. (2001) The in situ supermolecular structure of type I collagen. *Structure*, **9**, 1061-1069.

Pankov, R. and Yamada, K.M. (2002) Fibronectin at a glance. *J Cell Sci*, **115**, 3861-3863.

Paul, J.I. and Hynes, R.O. (1984) Multiple fibronectin subunits and their post-translational modifications. *J Biol Chem*, **259**, 13477-13487.

Pickford, A.R., Potts, J.R., Bright, J.R., Phan, I. and Campbell, I.D. (1997) Solution structure of a type 2 module from fibronectin: implications for the structure and function of the gelatin-binding domain. *Structure*, **5**, 359-370.

Pickford, A.R., Smith, S.P., Staunton, D., Boyd, J. and Campbell, I.D. (2001) The hairpin structure of the $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2 fragment from human fibronectin enhances gelatin binding. *Embo J*, **20**, 1519-1529.

Pierschbacher, M.D., Hayman, E.G. and Ruoslahti, E. (1981) Location of the cell-attachment site in fibronectin with monoclonal antibodies and proteolytic fragments of the molecule. *Cell*, **26**, 259-267.

Pierschbacher, M.D., Hayman, E.G. and Ruoslahti, E. (1985) The cell attachment determinant in fibronectin. *J Cell Biochem*, **28**, 115-126.

Pierschbacher, M.D. and Ruoslahti, E. (1984) Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature*, **309**, 30-33.

Potts, J.R. and Campbell, I.D. (1994) Fibronectin structure and assembly. *Curr Opin Cell Biol*, **6**, 648-655.

Potts, J.R. and Campbell, I.D. (1996) Structure and function of fibronectin modules. *Matrix Biol*, **15**, 313-320; discussion 321.

Rudino-Pinera, E., Ravelli, R.B., Sheldrick, G.M., Nanao, M.H., Korostelev, V.V., Werner, J.M., Schwarz-Linek, U., Potts, J.R. and Garman, E.F. (2007) The solution and crystal structures of a module pair from the Staphylococcus aureus-binding site of human fibronectin--a tale with a twist. *J Mol Biol*, **368**, 833-844.

Ruoslahti, E. (1996) RGD and other recognition sequences for integrins. *Annu Rev Cell Dev Biol*, **12**, 697-715.

Ruoslahti, E., Hayman, E.G., Pierschbacher, M. and Engvall, E. (1982) Fibronectin: purification, immunochemical properties, and biological activities. *Methods Enzymol*, **82 Pt A**, 803-831.

Ruoslahti, E. and Pierschbacher, M.D. (1987) New perspectives in cell adhesion: RGD and integrins. *Science*, **238**, 491-497.

Schor, S.L., Ellis, I., Banyard, J. and Schor, A.M. (1999) Motogenic activity of IGD-containing synthetic peptides. *J Cell Sci*, **112 ( Pt 22)**, 3879-3888.

Schwarz-Linek, U., Werner, J.M., Pickford, A.R., Gurusiddappa, S., Kim, J.H., Pilka, E.S., Briggs, J.A., Gough, T.S., Hook, M., Campbell, I.D. and Potts, J.R. (2003) Pathogenic bacteria attach to human fibronectin through a tandem beta-zipper. *Nature*, **423**, 177-181.

Schwarzbauer, J.E. (1991) Identification of the fibronectin sequences required for assembly of a fibrillar matrix. *J Cell Biol*, **113**, 1463-1473.

Schwarzbauer, J.E., Patel, R.S., Fonda, D. and Hynes, R.O. (1987) Multiple sites of alternative splicing of the rat fibronectin gene transcript. *Embo J*, **6**, 2573-2580.

Singer, A.J. and Clark, R.A. (1999) Cutaneous wound healing. *N Engl J Med*, **341**, 738-746.

Sottile, J. and Mosher, D.F. (1997) N-terminal type I modules required for fibronectin binding to fibroblasts and to fibronectin's III1 module. *Biochem J*, **323 ( Pt 1)**, 51-60.

Takasaki, S., Yamashita, K., Suzuki, K. and Kobata, A. (1980) Structural studies of the sugar chains of cold-insoluble globulin isolated from human plasma. *J Biochem (Tokyo)*, **88**, 1587-1594.

Tanzer, M.L. (2006) Current concepts of extracellular matrix. *J Orthop Sci*, **11**, 326-331.

Tordai, H. and Patthy, L. (1999) The gelatin-binding site of the second type-II domain of gelatinase A/MMP-2. *Eur J Biochem*, **259**, 513-518.

Traub, W. and Piez, K.A. (1971) The chemistry and structure of collagen. *Adv Protein Chem*, **25**, 243-352.

van der Rest, M. and Garrone, R. (1990) Collagens as multidomain proteins. *Biochimie*, **72**, 473-484.

Vitellaro-Zuccarello, L., Garbelli, R. and Rossi, V.D. (1992) Immunocytochemical localization of collagen types I, III, IV, and fibronectin in the human dermis. Modifications with ageing. *Cell Tissue Res*, **268**, 505-511.

Wierzbicka-Patynowski, I. and Schwarzbauer, J.E. (2003) The ins and outs of fibronectin matrix assembly. *J Cell Sci*, **116**, 3269-3276.

Xiong, J.P., Goodman, S.L. and Amin Arnaout, M. (2007) Purification, analysis, and crystal structure of integrins. *Methods Enzymol*, **426**, 307-336.

Xiong, J.P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D.L., Joachimiak, A., Goodman, S.L. and Arnaout, M.A. (2001) Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science*, **294**, 339-345.

Xu, J., Rodriguez, D., Petitclerc, E., Kim, J.J., Hangai, M., Moon, Y.S., Davis, G.E. and Brooks, P.C. (2001) Proteolytic exposure of a cryptic site within collagen type IV is required for angiogenesis and tumor growth in vivo. *J Cell Biol*, **154**, 1069-1079.

Yamada, K.M. (1983) Cell surface interactions with extracellular materials. *Annu Rev Biochem*, **52**, 761-799.

Zhu, B.C. and Laine, R.A. (1985) Polylactosamine glycosylation on human fetal placental fibronectin weakens the binding affinity of fibronectin to gelatin. *J Biol Chem*, **260**, 4041-4045.

# 2. Expression and purification

## 2.1. Protein expression systems

The ability to produce proteins of the extracellular matrix (ECM) by recombinant means has greatly accelerated understanding of the large variety of molecules which normally exist in an insoluble, cross-linked matrix. Recombinant expression methods have been particularly valuable for defining the structure and function of many ECM molecules, and in particular, fibronectin (FN) modules. A number of different heterologous overexpression systems have been developed to deal with the large size, heterogeneity and diverse translational modification present in FN. Fn1, Fn2, and Fn3 modules have all been produced previously, many in this laboratory, using a cell free system (Staunton et al., 2006), *Escherichia coli* (Vakonakis et al., 2007), *Pichia pastoris* (Bright et al., 2000), *Saccharomyces cerevisiae* (Baron et al., 1990), baculovirus-infected cells (Kishore et al., 1997), and mammalian cells (Houard et al., 2005).

Structural studies on fully folded, correctly modified protein are the obvious goal, so although some of the expression methods may be time expensive, the most valid results are only obtained from studies on the protein (fragments) in native conformations. Although the two expression systems described here are not the quickest or most straightforward, they are considered the most appropriate for the proteins under study; *P. pastoris* has been used extensively for the production of Fn1 modules, and a recently developed transient mammalian system has been adapted for the expression of extended FN constructs.

### 2.1.1. Expression of recombinant proteins in *Pichia pastoris*

*P. pastoris* has been used to produce recombinant proteins originating from bacteria (Tschopp et al., 1987), plants (Juge et al., 1996), viruses (Sugrue et al., 1997), and humans (Millard et al., 2005). *P. pastoris* is an attractive eukaryotic expression system because it is far easier to genetically manipulate and culture than mammalian cells. Most of the molecular manipulation techniques, first developed in *S. cerevisiae,* have been adapted and optimised for *P. pastoris* (Higgins and Cregg, 1998). *P. pastoris* is often preferred due to the availability of the alcohol oxidase (AOX1) gene promoter, which is highly induced with methanol, but highly repressed with glucose.

The protein of interest may be targeted for export from the cell by coexpression with a secretion signal. There are relatively few endogenous proteins secreted by *P. pastoris*, so that the recombinant material can be extracted with relative ease in the first purification step. The eukaryotic cell machinery results in the heterologous protein undergoing many of the post-translational modifications seen in higher eukaryotes. In successful cases, the resulting protein is fully folded, proteolytically processed, disulphide linked, and glycosylated in a similar way to the native protein from the original host. It is thus particularly suitable for expression of glycosylated extracellular proteins containing disulphide bonds, as are often found in the ECM.

*P. pastoris* is also a suitable host for the production of isotopically labelled proteins for study by NMR, where the simplicity of the cell culture media allows relatively cheap and efficient isotope incorporation. *P. pastoris* is now the second most widely

used expression system in the NMR community, following *E. coli* (Pickford and O'Leary, 2004).

Fragments of FN have been successfully expressed in both labelled and unlabelled forms. The usual procedure is to initially optimise the expression conditions in shake flasks, and then scale up for fermenter expression and isotope incorporation. High level protein expression in *P. pastoris* will typically yield 1-80 mg/litre depending on the size and nature of the construct although yields of greater than 2 g/litre have been reported (Cereghino and Cregg, 2000). Protein can be produced through shake flask expression, although due to poor aeration, the protein yield may be reduced threefold.

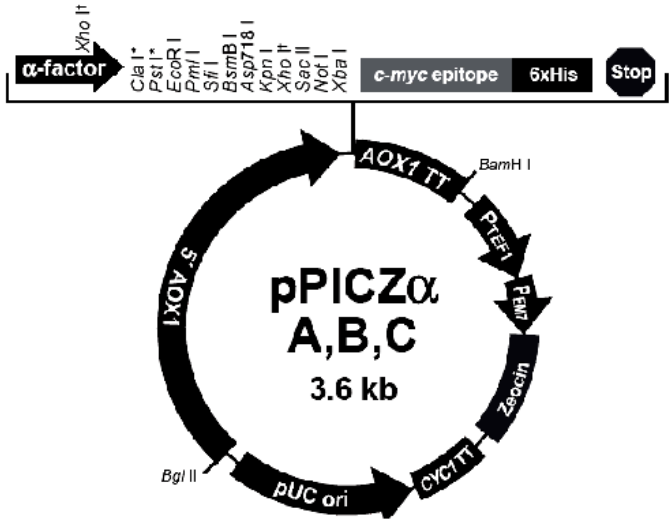



**Figure 2:1** The shuttle vector pPICZα A,B,C (Invitrogen). The multiple cloning site allows insertion of the gene of interest between the secretion signal and purification tags. This whole region is flanked by the AOX1 gene and its promoter which is inducible by the addition of methanol. The antibiotic resistance gene confers resistance to Zeocin™ in both prokaryotic and eukaryotic hosts.

The molecular biology involved in engineering *P. pastoris* strains to produce heterologous protein has become more straightforward because the reagents are now commercially available (Invitrogen) (Figure 2:1). Shuttle vector choice has increased

and protocols for *P. pastoris* DNA manipulation are now widely available. Here we used the vector pPICZα B which confers resistance to the antibiotic Zeocin™ in both *E. coli* and *P. pastoris* and carries the α-factor signal sequence for secretion of recombinant protein from *P. pastoris.* The α-factor signal sequence consists of a 19-amino acid signal (pre) sequence followed by a 66-residue (pro) sequence which targets the expressed protein to the culture medium, and is removed in two stages as the protein is processed through the endoplasmic reticulum. The *P. pastoris* strain X-33 recovers well from electroporation treatment, and is able to express large amounts of recombinant protein on induction with methanol.

### 2.1.2. Transient expression in mammalian HEK293T cells

The production of mammalian recombinant protein in high level expression systems such as *E. coli* and *P. pastoris* is adequate for small proteins with minor modifications but limited for the synthesis, folding, and post-translational modification of larger constructs. In order to improve the quality of the protein, expression would ideally be performed in native cell types, and for this reason a variety of mammalian expression systems have been developed. Although there has been much success, the systems all have their drawbacks; Chinese hamster ovary (CHO) cells take a long time to establish stable clones that produce sufficient amount of protein (Cockett et al., 1990), and human embryonic kidney (HEK293EBNA) cells require specialised and expensive equipment (Meissner et al., 2001). Therefore, a transient expression system using human embryonic kidney (HEK293T) cells has been developed which focuses on being simple, cost-effective and, most importantly, is able to produce correctly expressed mammalian protein in native cells (Aricescu et al., 2006).

The major advantage of transient expression is the ability to rapidly screen many new constructs without the necessity to create stable cell lines. The gene of interest is inserted into a modified lentiviral (pLEXm) vector, and incubated along with the cells and a transfection reagent polyethylenimine (PEI) which causes the cells to be more permeable, facilitating the take up of foreign DNA. The protein is excreted into the medium, ready for harvesting 3-4 days later; this can also be performed in parallel in a 24-well plate. The system can be scaled up for large scale expression in roller bottles and the whole procedure, from target cloning to crystallisation, can be completed in two weeks (Aricescu et al., 2006).

A disadvantage of the transient expression system is that new cells have to be transfected with new DNA and cells will only continue secreting proteins for 7-10 days. Large amounts of DNA are required, typically 50 μg DNA per 15 cm tissue-culture dish (approximate surface area 176 $cm^2$), but the effort of purifying so much DNA is rewarded with yields of between 1 and 40 mg of recombinant protein (per litre of culture). In addition, the ability of the cells to translate foreign DNA can allow for more complex protocols, such as co-transfection, which can result in the successful expression of protein complexes.

## 2.2. Methods

### 2.2.1. PCR amplification

The gene fragment of interest was amplified from template dsDNA of FN (provided by Dr. A.R. Pickford) using the high fidelity thermophilic Deep Vent® DNA Polymerase (all enzymes from New England Biolabs (NEB), unless otherwise stated). Outer primers and mutagenesis primers (Sigma Genosys) were used to construct mutants in a two-step overlap-extension polymerase chain reaction (PCR) (Figure 2:2). Melting of the duplex DNA (95°C for 30 s), annealing of primers (54-68°C for 30 s), and extension along the template (72°C for 1 min) were performed 26 times in each reaction before analysis by agarose gel electrophoresis.

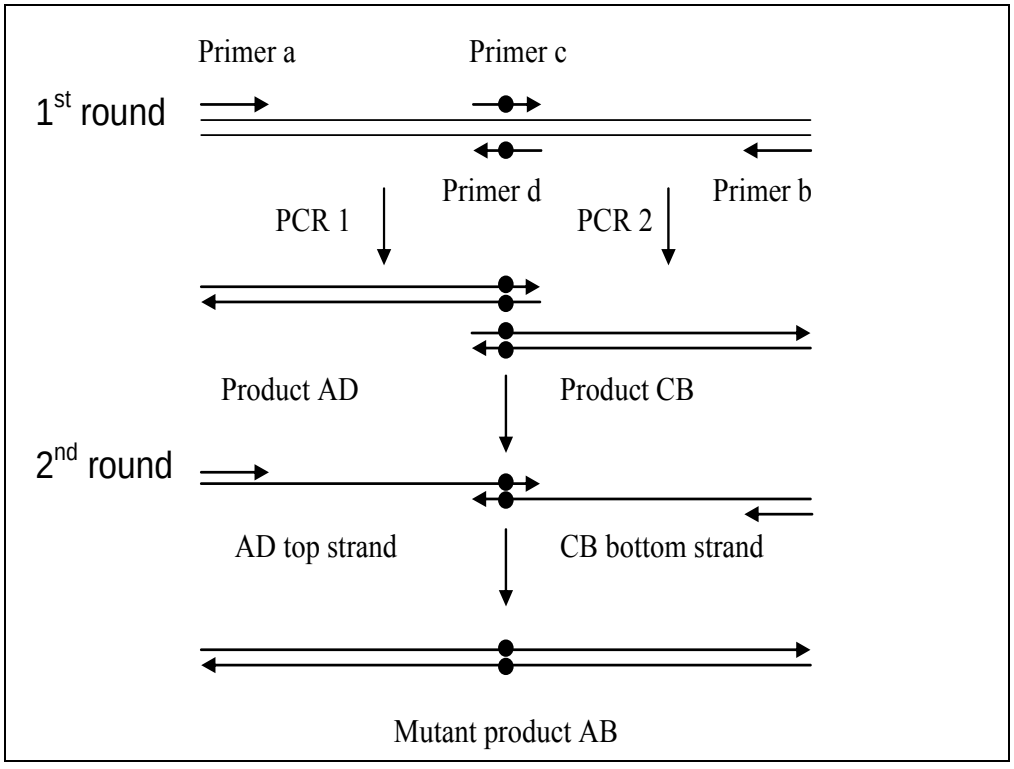



**Figure 2:2** Site directed mutagenesis by overlap extension PCR. 1st round: **Primer a** and **Primer d** amplify the upstream DNA and **Primer b** and **Primer c** amplify the downstream DNA from the template dsDNA. The template and primers are removed. 2nd round: the PCR products from the first round are allowed to anneal and extend before outer **Primer a** and **Primer b** amplify the mutant product. The DNA is transcribed in a 5' to 3' direction.

### 2.2.2. Primer design

Primers were designed to be between 18 and 30 base pairs in length with a GC content between 40 and 60% (Figure 2:3). The primers were adjusted to overlap with the gene of interest with a melting temperature ($T_m$) greater than 60°C, as calculated using the following formula:

$$T_m = 2°C*(A+T) + 4°C*(G+C)$$

The primers were designed to create restriction sites at both ends of the gene to aid ligation into the cloning vector. Two to six base pairs were added 5' to the restriction sites to increase the efficiency of restriction enzyme cleavage. Primers making use of the *XhoI* restriction site also included the Kex2 signal cleavage site 5' of the gene to allow for correct proteolytic processing during expression in *P. pastoris*. Those primers predicted to have significant secondary structure or pairs of primers able to dimerise were rejected and redesigned.

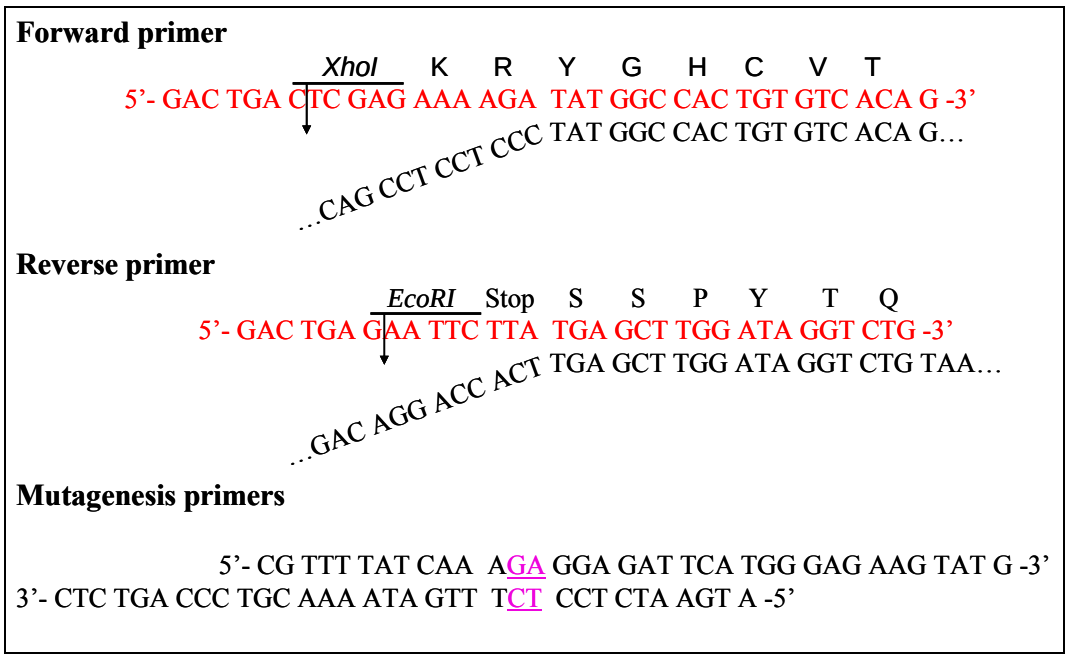



**Figure 2:3** The forward and reverse primers (red) for amplifying the Hex DNA by PCR. The melting temperature of the primer to the complementary part of the gene (black) is approximately 60°C. The internal primers for mutation Ile to Arg within the $^{9}$Fn1 module are shown (purple).

### 2.2.3. Vector construction and gene fragment insertion

The pPICZλ vector was constructed and gene fragments inserted into this vector as outlined in Figure 2:4 and Figure 2:5 (see Appendix A). Briefly, vector pPICZλ was produced by transplanting the 54 bp *EcoRI-XbaI* double cut fragment from the multiple cloning site of LITMUS 28i (New England Biolabs) into the corresponding restriction sites of pPICZα B (Invitrogen). Restriction enzymes were used to digest the substrate DNA to completion (1 U/μg DNA) in the recommended buffer. 5'-Dephosphorylation of double stranded vector DNA was catalysed by Antarctic Phosphatase (5 U/μg DNA) for 15 mins at 37°C, and ligation of 5' phosphate and 3' hydroxyl termini was catalysed by T4 DNA Ligase (1 U/pmolDNA termini) for 2 hours at 16°C. After each step, the enzyme was heat inactivated at 65°C for 20 mins, to avoid inhibition by the previous step, and the presence of DNA product was confirmed by agarose gel electrophoresis.

### 2.2.4. Transformation of *E. coli*

The pPICZλ vectors containing FN fragments were transformed into *E. coli* strain DH5α (Invitrogen) by incubation with foreign plasmid DNA followed by heat shock, then plated out on low-salt (<5 g/l [NaCl]) LB-Zeocin™ (Invitrogen) plates, and incubated overnight at 37°C. Single colonies showing Zeocin™ resistance were selected, grown up in 2 ml low-salt (<5 g/l [NaCl]) LB-Zeocin™ overnight at 37°C, before purifying plasmid DNA by miniprep (HiSpeed Plasmid Miniprep Kit - Qiagen). The presence of the FN insert was initially identified by an *EcoRI-XhoI* double digest, and subsequently confirmed by plasmid sequencing, before the preparation of 100-200 μg plasmid DNA (HiSpeed Plasmid Midiprep Kit - Qiagen).

**Vector construction: pPICZλ**

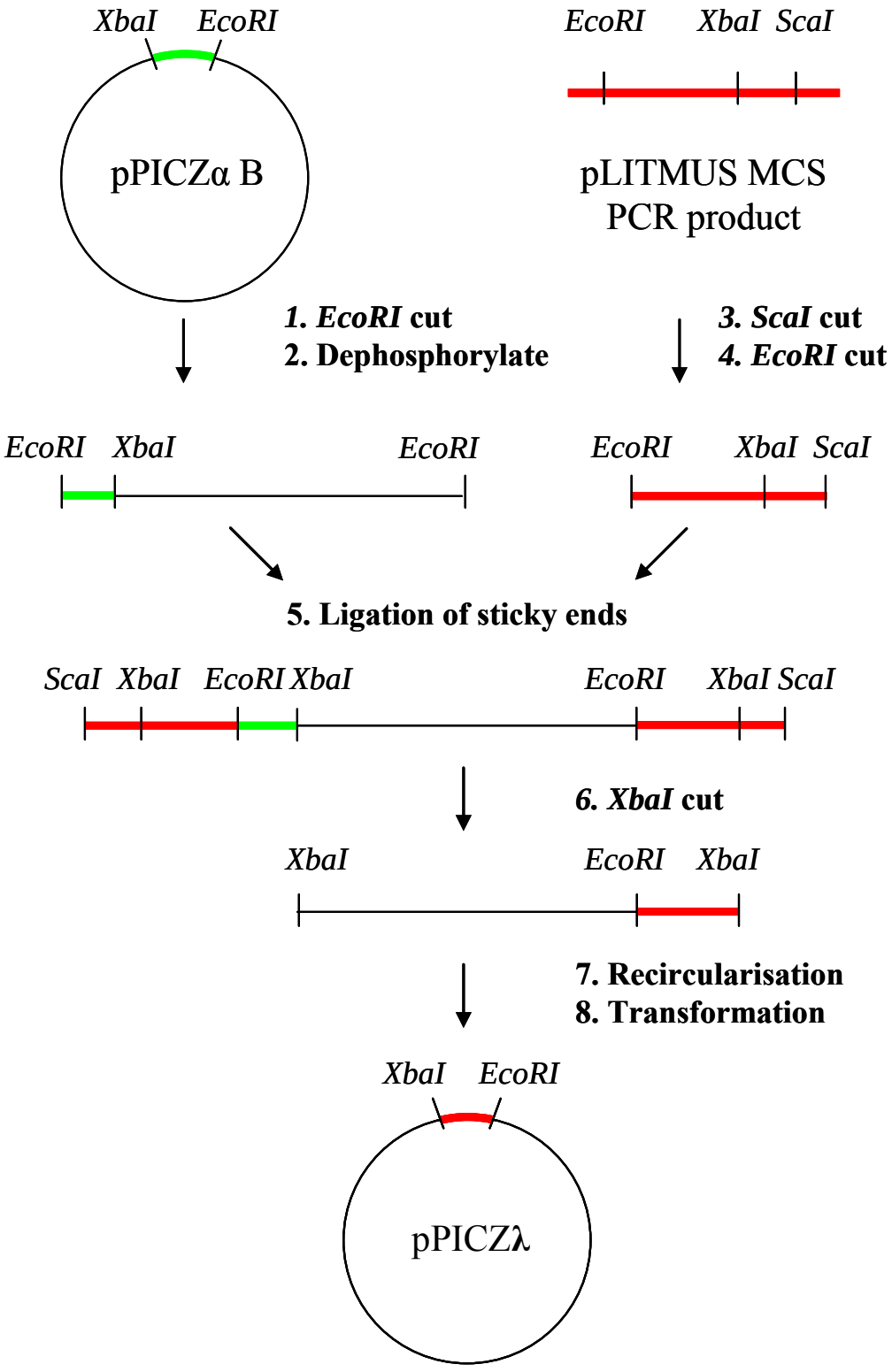



**Figure 2:4** Construction of pPICZλ. pPICZα B (Invitrogen) is linearised to allow ligation of the PCR amplified products (red) to either end of the vector. Digest by *XbaI* creates sticky ends, which allows recircularisation and insertion of the pLITMUS multiple cloning site (MCS).

**Typical Gene Fragment Insertion**

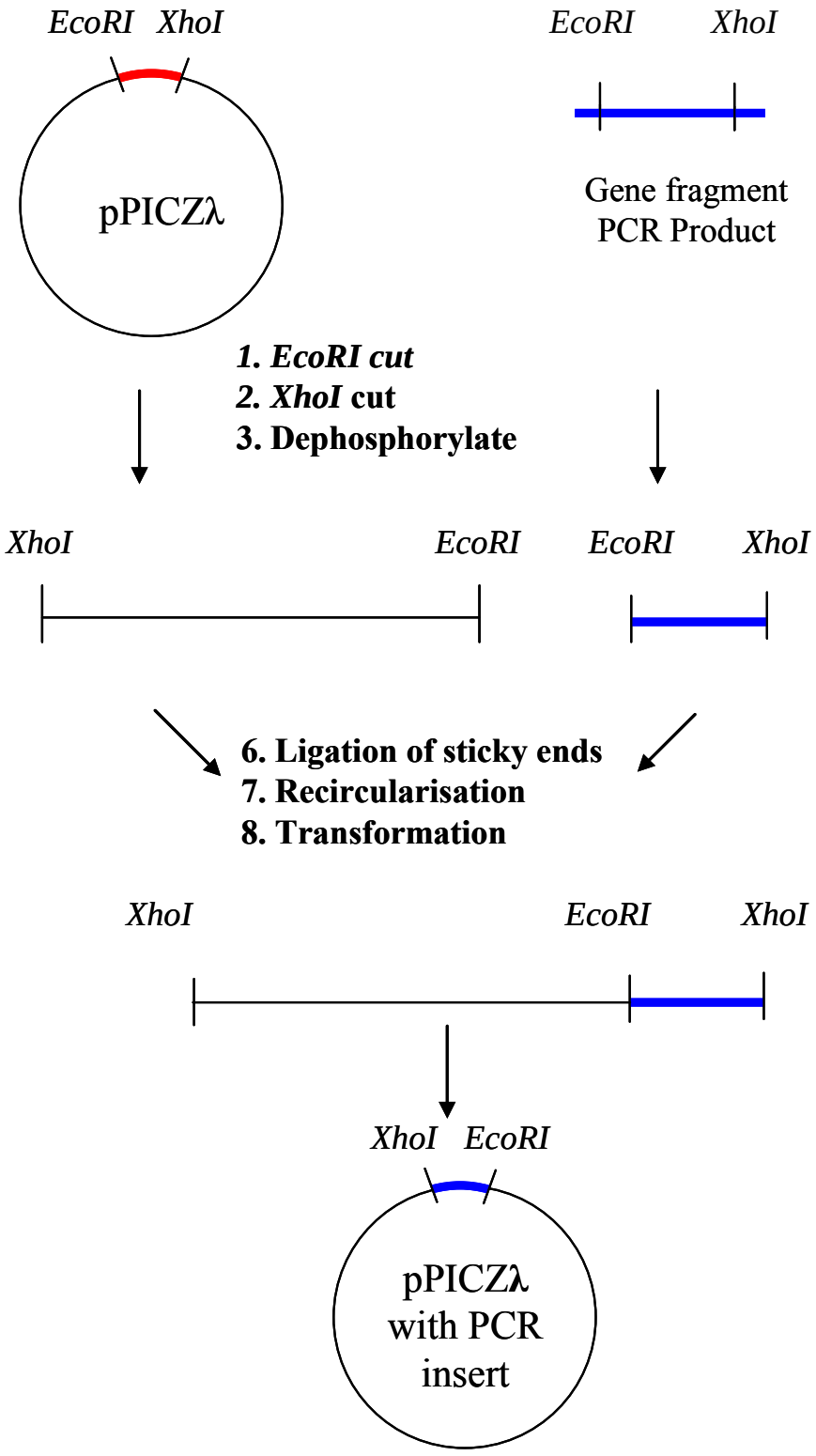



**Figure 2:5** Insertion of PCR amplified FN gene fragment into vector pPICZλ. The multiple cloning site (MCS) of pPICZλ was excised by double restriction enzyme digest (red) and replaced with the PCR amplified gene fragment (blue).

### 2.2.5. Preparation of *P. pastoris* cells for electroporation

A single colony of *P. pastoris*, strain X-33 (Invitrogen), was picked from a freshly streaked plate (yeast extract (1%), peptone (2%), dextrose (2%) (= YPD), agar (2%)), and used to inoculate a starter culture of 5 ml YPD. The culture was grown overnight at 30°C in YPD before diluting 1 ml of this starter into 500 ml of fresh YPD medium in a 2 litre flask. The cells were grown overnight to an $OD_{600}$ = 1.3-1.5 and harvested by centrifugation at 1500g for 5 mins at 4 °C. In a stepwise fashion, the pellet was washed and harvested (1500g for 5 mins at 4°C) three more times, twice in reducing volumes of ice-cold (0°C) sterile water (500 ml and 250 ml) and once in 20 ml of ice-cold (0°C) 1 M sorbitol. The cells were finally resuspended in 1ml of ice-cold (0°C) 1 M sorbitol and used immediately for electroporation.

### 2.2.6. Preparation of DNA for transformation into *P. pastoris*

Typically 1-10 µg linearised plasmid DNA was used for each transformation. The plasmid was digested with *SacI* for 2 hours at 37°C before inactivation by incubation at 65°C for 20 mins. The DNA was precipitated using 1/10 volume 3 M sodium acetate and 2.5 volumes of 100% ethanol. After centrifuging, the DNA pellet was washed with 70% ethanol, air dried, and resuspended in 10 µl sterile water.

### 2.2.7. Transformation of *P. pastoris*

The DNA was incubated with 60 µl freshly prepared electrocompetent *P. pastoris* strain X-33 cells for 5 mins and transferred to an ice-cold 0.2 cm electroporation cuvette. The cells were pulsed (25 µF, 200 Ω, 2000 V) for 5 ms, immediately resuspended in 1 ml ice-cold 1 M sorbitol, and allowed to recover at 30°C without

shaking for 1 hour. The transformants were plated on 100 μg/ml Zeocin™ plates, and after 2-4 days, larger colonies were plated onto higher concentrations of Zeocin™. Colonies growing well on high concentrations of Zeocin™ were screened for recombinant protein expression.

### 2.2.8. Small scale protein expression screens

A single colony from a freshly streaked high concentration Zeocin™ plate was used to inoculate 50 ml glycerol containing media (glycerol provided the initial carbon source for growth) (5 ml 1 M potassium phosphate pH 6.0, 3.4 g yeast nitrogen base, 10 g ammonium sulphate, 10 ml glycerol, 0.4 mg biotin to 50 ml with sterile water (=BMG)). The culture was grown in a 250 ml baffled flask with shaking at 30°C for 48 hours (or until the culture reached an $OD_{600}$ = 0.5-1.0), before centrifuging at 3000g for 5 mins, and resuspending the cells in 50ml methanol containing media (BMG replacing glycerol with 1% methanol (=BMM)). Upon the exhaustion of the glycerol from the media the sole carbon source became methanol which also served as the inducer from protein expression. The cells were grown for a further 72 hours at 30°C before centrifuging 1.5 ml of supernatant at 10,000g for 1 min for protein analysis. The supernatant was concentrated by the addition of 10 μl StrataClean™ Resin (Stratagene) to collect the protein, the beads were boiled for 5 mins with sample buffer, and then the slurry loaded directly onto a reducing sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel. Colonies containing and expressing the gene of interest were confirmed by the presence of protein bands of the correct molecular weight.

### 2.2.9. *P. pastoris* fermentation

Unlabelled and uniformly $^{15}$N-labelled protein was expressed by fermentation in a 1 litre table-top glass fermenter (Electrolab Ltd) with variable speed impeller, air supply and oxygen probe; the temperature was controlled via a heat mat and cold finger; base (KOH) and methanol (50%) were fed as required. The fermentation was carried out using previously published protocols (Bright et al., 2000), (www.invitrogen.com) (see Appendix D, p189). Briefly, 500 ml basal salt media (13.4 ml phosphoric acid, 0.46 g calcium sulphate, 7.5 g magnesium sulphate heptahydrate, and potassium hydroxide to pH 3.0 (=BSM)), 5 ml glycerol (carbon source), and 5 g ammonium sulphate (nitrogen source), was inoculated with a 50 ml overnight starter culture, and grown at 30°C for 24 hours with constant aeration. The carbon source was then switched to a limiting drip feed of methanol as judged by the regularity of dissolved oxygen (DO) spikes in the aeration trace. The dissolved oxygen concentration was held constant (<20%) by monitoring the impeller speed, and the pH was kept at 3.0 by the addition of base. Although the protein was most likely expressed unfolded at that pH, degradation by proteases was kept to a minimum, and was also further reduced by the addition of antifoam.

### 2.2.10. Harvesting the protein supernatant

The culture was harvested after 72-90 hours (after the start of the methanol feed) as indicated by the decreasing regularity of the DO spikes and the start of a slight pH increase. The culture was centrifuged at 12,000g at 4°C for 20 mins and the supernatant collected and stored at 4°C. The cells were washed in 500 ml 100 mM HCl, centrifuged at 12,000g at 4°C for 20 mins, and the supernatant added to that

already collected. The culture supernatant was filtered, diluted to 2.5 litres with sterile water and adjusted to pH 3.0 with 5 M HCl.

### 2.2.11. Cation exchange chromatography

The crude protein product was partially purified by loading at 20 ml/min to a SP-Sepharose Fast Flow (Amersham Pharmacia Biotech) column that had been equilibrated previously with buffer A (20 mM sodium citrate pH 3.0). The column was washed with two volumes of buffer A before the bound protein was eluted with a gradient of 0-100% buffer B (2 M NaCl, 20 mM sodium citrate pH 3.0) at 20 ml/min. The absorbance of the eluent was monitored at 280 nm, and those fractions with significant $A_{280}$ were analysed for expressed protein by SDS-PAGE.

### 2.2.12. Trimming of high mannose oligosaccharides

The high mannose sugar chains attached at Asn399 and/or Asn497 and/or Asn511 in *P. pastoris*-secreted proteins were trimmed with endoglycosidase (Endo) $H_f$ (Sticht et al., 1998). This enzyme cleaves the chitobiose core of high-mannose sugars to leave a single N-acetylglucosamine (GlcNAc) bound to the Asp residue via an N-glycosidic linkage. Fractions from cation-exchange chromatography were adjusted to pH 5.5 with 1 M tri-potassium phosphate and incubated with 1000 U Endo $H_f$ per ml of sample (New England Biolabs) for 1 hour at 37ºC.

### 2.2.13. Preparative gelatin affinity chromatography

Deglycosylated fractions from the cation exchange were further purified by affinity chromatography with gelatin-Sepharose 4B (Amersham Pharmacia Biotech) at 4ºC. The affinity column was equilibrated with buffer C (PBS pH 7.4) before loading the sample at 2 ml/min. The column was washed with one volume of buffer C and the bound protein eluted with a gradient of 0-100% buffer D (4 M Urea, PBS pH 7.4) over 5 column volumes (CV) at 2 ml/min. Reducing SDS-PAGE was used to analyse samples with significant $A_{280,}$ and the main peaks of protein were pooled separately.

### 2.2.14. Reverse-phase high performance liquid chromatography (RP-HPLC)

Final purification to homogeneity was achieved by RP-HPLC on a semi-preparative C4 column (Phenomenex Jupiter 300 Å 10 μm) with a gradient of 30-40% acetonitrile and 0.1% trifluoroacetic acid at 3 ml/min (see Figure 2:10 and Figure 2:11). Fractions with significant $A_{280}$ were freeze-dried for analysis by electrospray ionization mass spectrometry (ESI-MS).

### 2.2.15. Transient protein expression

The gelatin binding domain, termed Hex, (six modules, see Figure 2:6) gene fragment with C-terminal $His_6$ tag was cloned into vector pHLsec (gift from Dr. A.R. Aricescu) and used to transform *E. coli* strain DH5α. Positive colonies were selected on carbenicillin and the plasmid extracted from 500 ml overnight bacterial culture. DNA was purified (Endotoxin-Free Plasmid Mega Kit - Qiagen) and used to transfect

HEK293T cells with the transfection reagent polyethylenimine (PEI). The transfection protocol, growth, and harvesting was performed by Dr. A.R. Aricescu.

### 2.2.16. Purification of transiently expressed protein

The supernatant was passed through a 22 μm filter (Millipore) to remove cellular debris before adopting a two step purification procedure to achieve final purity. First, the protein (Hex) was passed down a Talon® Resin purification column (Clontech), pre-equilibrated with 50 mM sodium phosphate pH 7.0, 300 mM NaCl, and retained due to the $His_6$ tag. The protein was washed with 5 mM imidazole, 50 mM sodium phosphate, 300 mM NaCl pH 7.0 and eluted with 100 mM imidazole 50 mM sodium phosphate, 300 mM NaCl pH 7.0 in 5 ml fractions. The fractions were assessed by SDS-PAGE, those containing protein of the correct molecular weight were pooled, and the protein concentrated in a Amicon® Ultra-4 centrifugal filter device 30K NMWL (nominal molecular weight limit) (Millipore) to 2 ml. The protein was washed into 10 mM Tris, 10 mM NaCl pH 7.0 at 1.5 ml/min on a Sephadex® G-75 Superfine gel filtration column as a final purification step.

### 2.2.17. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein analysis by was performed using NuPAGE® 10% Bis-Tris pre-cast gels (Invitrogen). Protein samples were diluted 1:4 into sample loading buffer (1 M Tris-Cl pH 6.8, SDS (6%), Glycerol (30%), bromophenol blue (0.06%) and β-mercaptoethanol (2.5%) for reducing gels), heat denatured at 95°C for 5 mins, before loading 15 μl into each lane and running at 200V for 35 mins in MES SDS running buffer (Invitrogen). Gels were stained (methanol (10%), acetic acid (40%), Coomassie blue (0.2%)) for 15 mins with shaking and destained for 1 hour (methanol

(10%), acetic acid (40%)) with shaking. Protein molecular weights were judged against Mark 12 standard molecular weight markers (Invitrogen).

### 2.2.18. Western blotting

Protein was analysed by western blot detection with two primary antibodies; anti-FN (gelatin binding domain, clone IST-10, Chemicon), and anti-$His_6$ (Amersham). Protein was transferred from a pre-run SDS-PAGE gel to a PVDF (polyvinylidene difluoride) membrane at 30 V for 2 hours in tris-glycine buffer (12 mM tris base, 96 mM glycine, 20% methanol). The membrane was placed in blocking buffer (5 g Marvel, 100 µl Tween-20, 100 ml PBS) and incubated with shaking for 10 mins at 25°C. The buffer was discarded and replaced with the new blocking buffer before incubating the membrane with shaking for 10 mins at 25°C (repeated twice). The membrane was placed into blocking buffer with added primary antibody and incubated at 4°C overnight. Antibody concentrations were determined empirically (anti-FN dilution 1:750, anti-$His_6$ dilution 1:3000).

Detection of binding of the primary antibody was determined through a secondary antibody linked to a reporter enzyme (horseradish peroxidase (HRP) conjugated anti-mouse IgG). The antibody/blocking buffer was discarded from the membrane, replaced with wash buffer (1 ml Tween-20, 500 ml PBS), and incubated with shaking for 10 mins at 25°C (repeated three times). The secondary antibody was diluted 1:5000 into wash buffer and incubated with shaking for 1 hour at 25°C. The membrane was washed twice with wash buffer with shaking for 10 mins at 25°C before chemiluminescent detection of the HRP on photographic film (Amersham).

### 2.2.19. The 2100 Bioanalyzer System

Further protein analysis was performed on the 2100 Bioanalyzer System (Agilent) to detect the protein molecular weight. The protein was denatured and reduced (as with SDS-PAGE) before being injected into a linear non-crosslinked polymer, and separated on the basis of size. The system was used as an independent method to confirm the SDS-PAGE results and provided resolution comparable to a 1-dimensional 10% gel.

### 2.2.20. Electrospray ionization mass spectrometry (ESI-MS)

The identity and purity of the recombinant proteins was confirmed by ESI-MS using the Open Access Facility at the Chemistry Research Laboratory, Oxford University.

### 2.2.21. N-terminal sequencing

The identity of the protein fragments was confirmed by N-terminal sequencing as performed by Mr Tony Willis of the MRC Immunochemistry Unit on Applied Biosystems Procise 494A.

## 2.3. Results

### 2.3.1. Design of constructs

Construct design is a critical part of any expression strategy as failure to express is often linked to errors in defining module and domain boundaries. Modular constructs of FN were designed as shown in Figure 2:6 and Table 1.

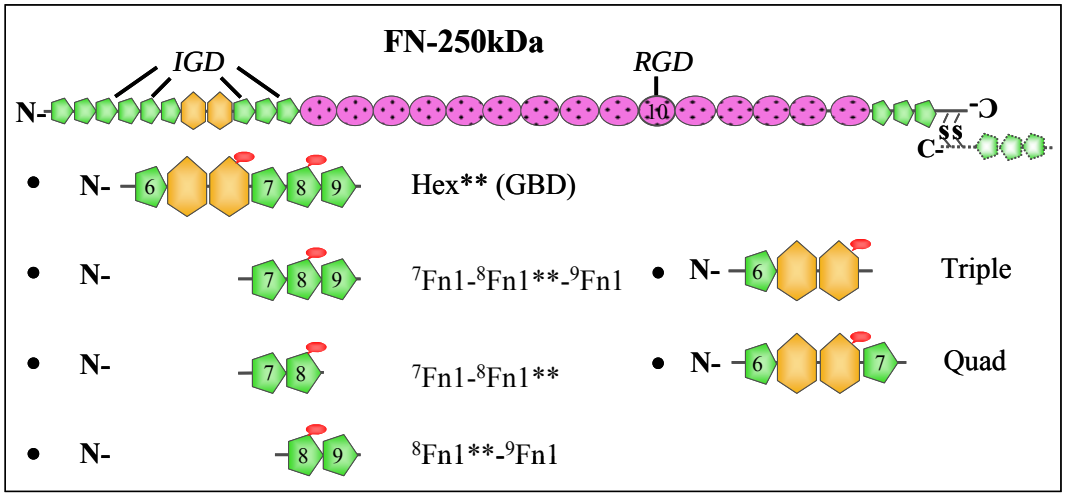



**Figure 2:6** A single monomer of dimeric FN (above) with FN constructs expressed in *P. pastoris* (below). Modules; Fn1 (green), Fn2 (orange), Fn3 (purple), with the position of the sugar residues shown in red ($^{2}$Fn2 and $^{8}$Fn1). ** indicates double mutation after the introduction of mutations R503K and N497Q. The Quad and Triple constructs were gifts from Dr. A.R. Pickford.

| | N-term | C-term | Glycosylation | Theoretical MW |
|---|---|---|---|---|
| **Hex**** | Y274 | S577 | 2x GlcNAc: N399, N511 | 34975 |
| **Quad** | Y274 | D484 | 1x GlcNAc: N399 | 24037 |
| **Triple** | Y274 | A433 | 1x GlcNAc: N399 | 18016 |
| **$^{7}$Fn1-$^{8}$Fn1**** | A434 | V527 | 1x GlcNAc: N511 | 11123 |
| **$^{8}$Fn1**-$^{9}$Fn1** | D485 | S577 | 1x GlcNAc: N511 | 11074 |
| **$^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1** | A434 | S577 | 1x GlcNAc: N511 | 16972 |

**Table 1** Construct design for the six fragment lengths after the introduction of mutations R503K and N497Q. Further mutations are mentioned later in the text. The theoretical molecular weight for the oxidised protein is indicated. Affinity tags (such as $His_6$) are not included in the molecular weight. Each module contains two disulphide bonds. Hex** = $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1, Quad = $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1, Triple = $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2.

The $^{9}$Fn1 C-terminus (Q572) was initially designed to form part of the final β-sheet and therefore to have some secondary structure to aid NMR studies. The C-terminus was extended to S577 for the biological assays to avoid omitting any functional tail residues. The $^{8}$Fn1 C-terminus of the $^{7}$Fn1-$^{8}$Fn1 construct was judged to be V527 from the initial structural data from the $^{8}$Fn1-$^{9}$Fn1 module pair. The N-termini of the constructs followed previous designs; residue Y274 in $^{6}$Fn1 constructs (Pickford et al., 2001), A434 in $^{7}$Fn1 constructs (Baron et al., 1990), and D485 in the $^{8}$Fn1 constructs (Millard et al., 2005), (see Appendix B).

The wild type $^{8}$Fn1-$^{9}$Fn1 had been shown to be proteolytically cleaved by the *Pichia Pastoris* endoprotease KEX2 between residues R503 and H504. The R503K mutation was therefore introduced to alleviate this proteolysis problem (Millard et al., 2005). This $^{8}$Fn1*-$^{9}$Fn1 (R503K) module pair expressed in *P. pastoris* was found to be heterogeneously glycosylated at two sites; at least one of the sites was glycosylated and the different fractions could be separated by increasing the retention time on a C4 RP-HPLC column. We previously concluded that glycosylation of Asn511 was critical for optimal binding to gelatin whereas glycosylation of Asn497 had no significant effect on binding activity (Millard et al., 2005). To increase the yield of homogeneous protein, and to simplify the purification procedure, Asn497 was mutated to remove the "non-essential" glycosylation site. The mutation N497Q was introduced into all constructs used in this study (with the exception of $^{8}$Fn1*-$^{9}$Fn1 which was used to assign $^{8}$Fn1**-$^{9}$Fn1) as it was judged to be a suitable conservative mutation to remove the Asn497 glycosylation site and is referred to as $^{8}$Fn1**-$^{9}$Fn1 (N497, R503K). The mutation was judged to have not affected the gelatin binding affinity of the $^{8}$Fn1**-$^{9}$Fn1 construct.

### 2.3.2. Gene fragment amplification and insertion into pPICZλ

Vector design and construction was carried out in a stepwise manner, initial work focusing on engineering a convenient vector capable of allowing the insertion of the FN gene fragments in one easy step. Upon completion of the new vector, the one-step insertion strategy was implemented, and PCR-amplified fragments were spliced into the vector. PCR amplified FN fragment dsDNA (containing the R503K, N497Q) were ligated into the *XhoI-EcoRI* double cut pPICZλ vector and transformed into *E. coli* strain DH5α. Selection on Zeocin™ plates indicated positive colonies and sequencing confirmed the correct orientation of the gene within the vector.

### 2.3.3. Transformation of *P. pastoris*

The *SacI* cut plasmid DNA was transformed into *P. pastoris* by electroporation and high expression colonies were selected on increasing concentrations of Zeocin™. Initially, protein was produced in small scale expression screens before scaling up to fermentation. SDS-PAGE gave an indication of potential expression levels and confirmed the approximate molecular weight of the protein.

### 2.3.4. *P. pastoris* fermentation

Each FN fragment was expressed in a 500 ml fermentation vessel where cells were induced to secrete protein for 5 days in methanol containing media. Initially the culture was grown at 30°C with glycerol as the carbon source. Once the biomass had reached $OD_{260}$=1 (around 24 hours), the carbon source was switched to methanol, which also acted to induce recombinant protein expression. The fermentation trace (Figure 2:7) showed that the culture deoxygenated the media (maintained at ~20% $O_2$

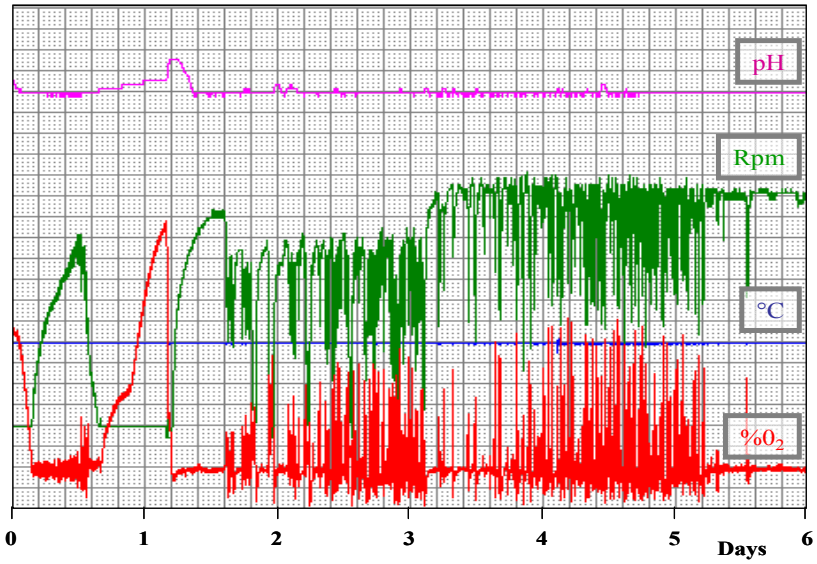



**Figure 2:7** A typical *P. pastoris* fermentation trace. The computer controlled impeller (green) maintains the dissolved oxygen (red) at 20% through a feedback mechanism. The temperature (blue) is kept at 30°C by a heating mat and cold finger. The pH (purple) is maintained at 3.0 by the addition of base to the media and acidification by the growing culture.

by the impeller) until the glycerol was exhausted. Depletion of carbon source caused the pH (originally set to 3.0) to rise. A growth limiting methanol drip feed enabled the culture to be grown at a controlled rate; too little methanol limiting the growth, too much being toxic to the cells. The depletion of each methanol drop resulted in a "spike" in the dissolved oxygen (DO) trace before the next drop arrived. The fermentation was harvested when the DO spikes began to reduce in frequency and the pH started to rise, indicating that the culture reached stationary phase due to nutrient depletion and accumulation of toxic products (Figure 2:8).

### 2.3.5. Protein expression and purification

After harvest the cells were pelleted by centrifugation at 10,000g for 20 min at 4ºC, washed with sterile water, and the protein-containing supernatant was filtered to remove cellular debris. The unlabelled and $^{15}$N-labelled constructs were expressed at

levels of approximately 40-80 mg/l (dependant on the construct) from fermentations of *P. pastoris*, as judged by SDS-PAGE of crude fermentation supernatant (data not shown). "Wild type" and mutant, unlabelled and $^{15}$N-labelled $^{8}$Fn1*-$^{9}$Fn1, $^{8}$Fn1**-$^{9}$Fn1, $^{7}$Fn1-$^{8}$Fn1**, and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 constructs were purified by the stepwise application of cation-exchange and RP-HPLC, whereas unlabelled and $^{15}$N-labelled Hex were purified by cation-exchange and gelatin affinity chromatography because of the poor resolution of this large fragment on the RP-HPLC column.

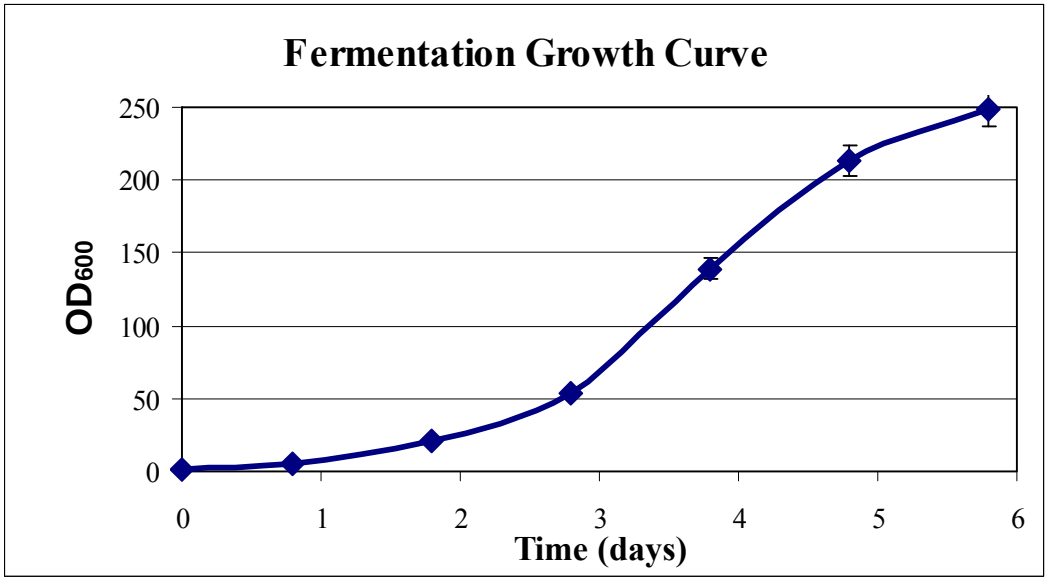



**Figure 2:8** The *P. pastoris* growth profile. The recombinant material is harvested after 6 days to ensure the quality of the protein. The cell mass and protein yield would continue to increase beyond six days but would increase the chance of protease attack. The doubling time is around 2 hours for a healthy culture grown in basal salts medium with methanol as the carbon source and ammonium sulphate as the nitrogen source.

### 2.3.6. Cation exchange chromatography

The cation-exchange chromatography was conducted at pH 3.0, well below the theoretical pI of all constructs (calculated by ProtParam), to ensure that the protein bound to the negatively charged SP sepharose matrix (Figure 2:9). A 2 M NaCl gradient was used to elute the protein and the absorbance of each fraction was

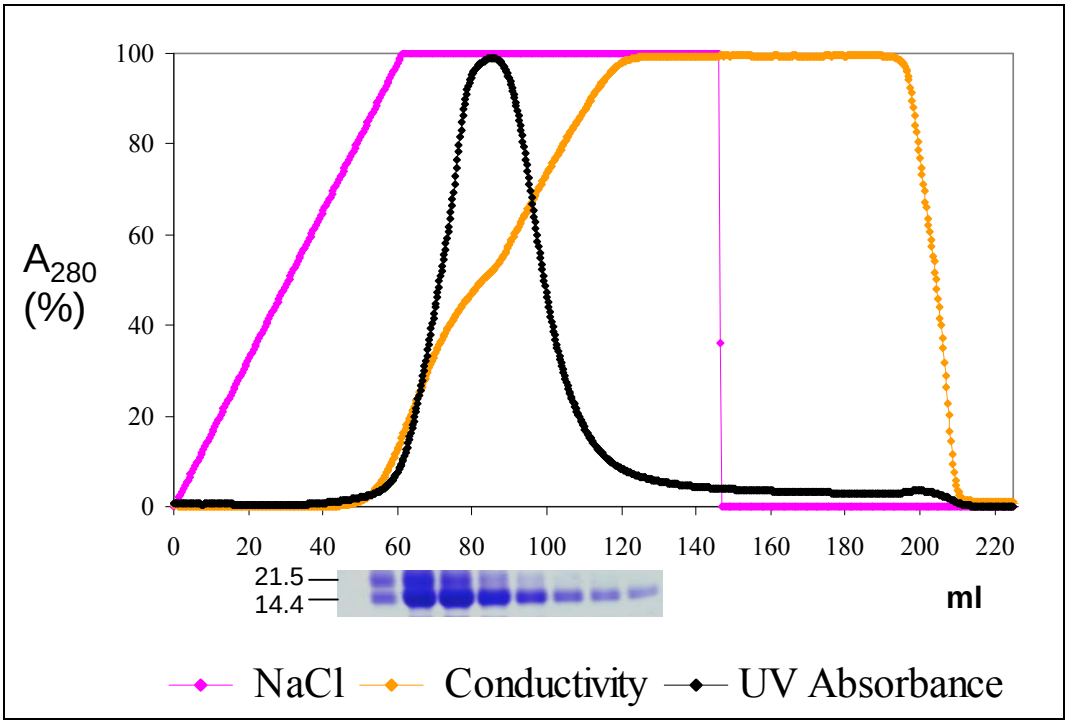



**Figure 2:9** Cation-exchange chromatography of crude $^{8}$Fn1**-$^{9}$Fn1 culture supernatant from the $^{15}$N-labelled $^{8}$Fn1**-$^{9}$Fn1 fermentation. Assuming a void volume of 60 ml (column volume), the protein eluted around 0.6-0.8 M NaCl (2 M NaCl applied to the column). The fractions around the $A_{280}$ elution peak were analysed on a 10% SDS PAGE gel (reducing) to confirm presence of glycosylated protein (~14kDa), (inset). The profile was similar for all *P. pastoris*-expressed FN fragments.

monitored at $A_{280}$. The fragments were detected in the 10 ml fractions taken between 50 and 150 ml as judged by SDS-PAGE of glycosylated samples (Figure 2:9, inset). Fractions from 60 to 120 ml were pooled in preparation for the next purification step. This initial step in the purification appeared to be of poor resolution; however, it also served as a concentration step, reducing the protein volume from 500 to 60 ml. The pH was adjusted to 5.5 and the high mannose sugar chain was removed with Endo $H_f$.

### 2.3.7. Reverse-phase high performance liquid chromatography

The cation exchange-purified fragments (with the exception of the Hex) were purified by RP-HPLC (Figure 2:10). The major peak was selected for analysis by ESI-MS and

NMR. The purity of each variant was improved (to > 95%) by lyophilisation and a second pass through the RP-HPLC column (Figure 2:11).

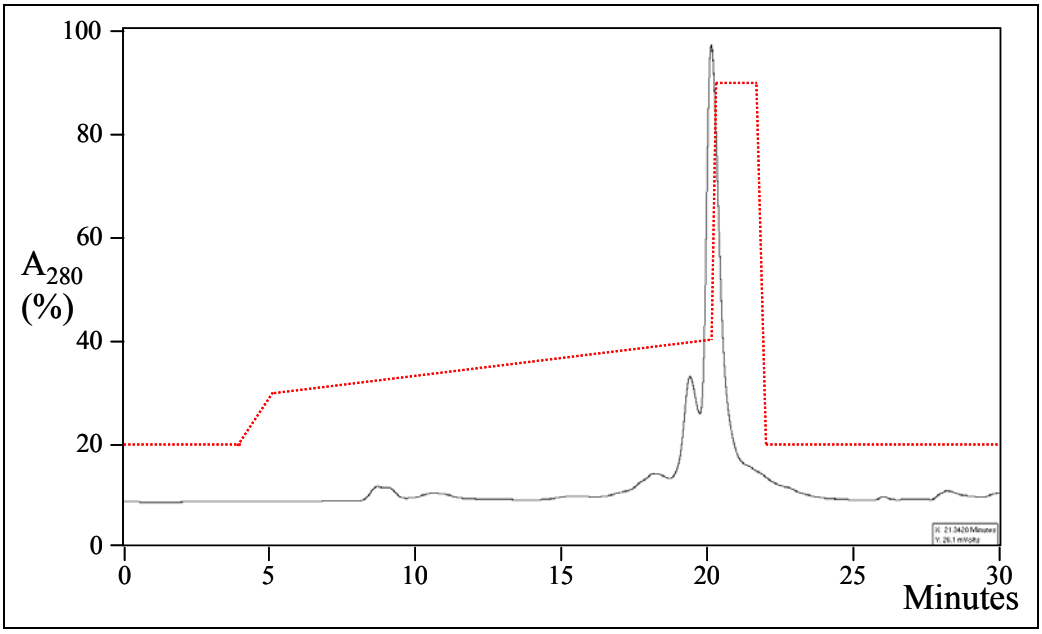



**Figure 2:10** RP-HPLC separation of fractions from cation-exchange purified $^{15}$N-labelled $^{8}$Fn1*-$^{9}$Fn1. The various peaks are separated by their differing hydrophobicity.

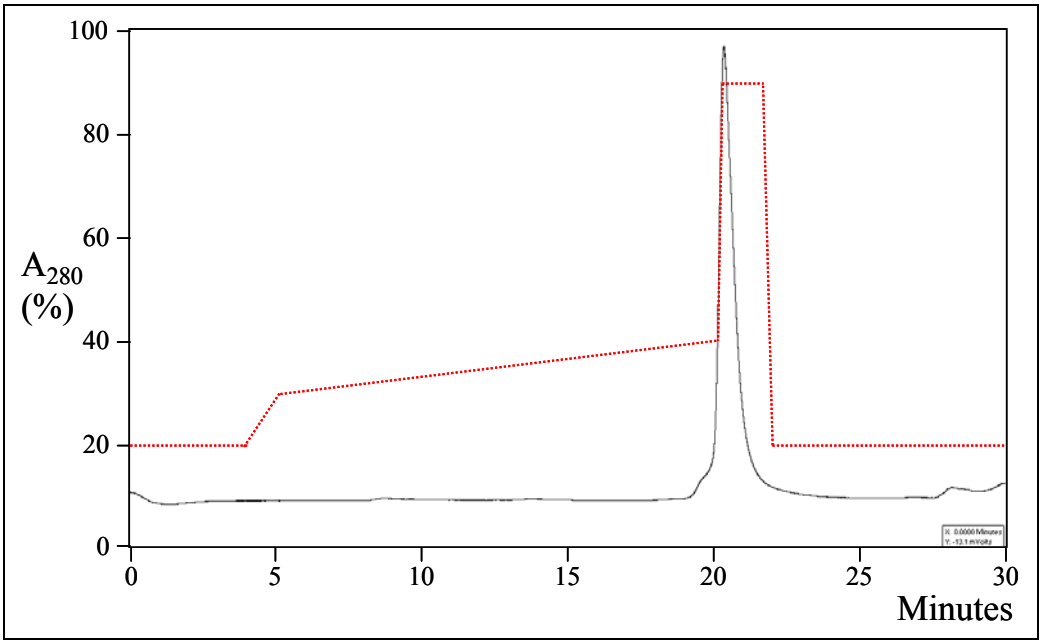



**Figure 2:11** A single peak of $^{8}$Fn1*-$^{9}$Fn1 is achieved after a second pass through the RP-HPLC column. This single peak is suggestive of high purity and was typical of all constructs purified in this way. The purity was further confirmed by NMR.

### 2.3.8. Affinity chromatography

This step was designed to separate out the recombinant Hex protein from endogenously secreted *P. pastoris* proteins, on the assumption that none of the proteins produced by yeast were able to bind to the gelatin affinity column. To elute the protein, a 4 M urea gradient was applied to act as a chaotrope to disrupt the FN-gelatin interaction (Figure 2:12). This purification technique was used as a final polishing step for the Hex, as purification by RP-HPLC was unsuccessful.

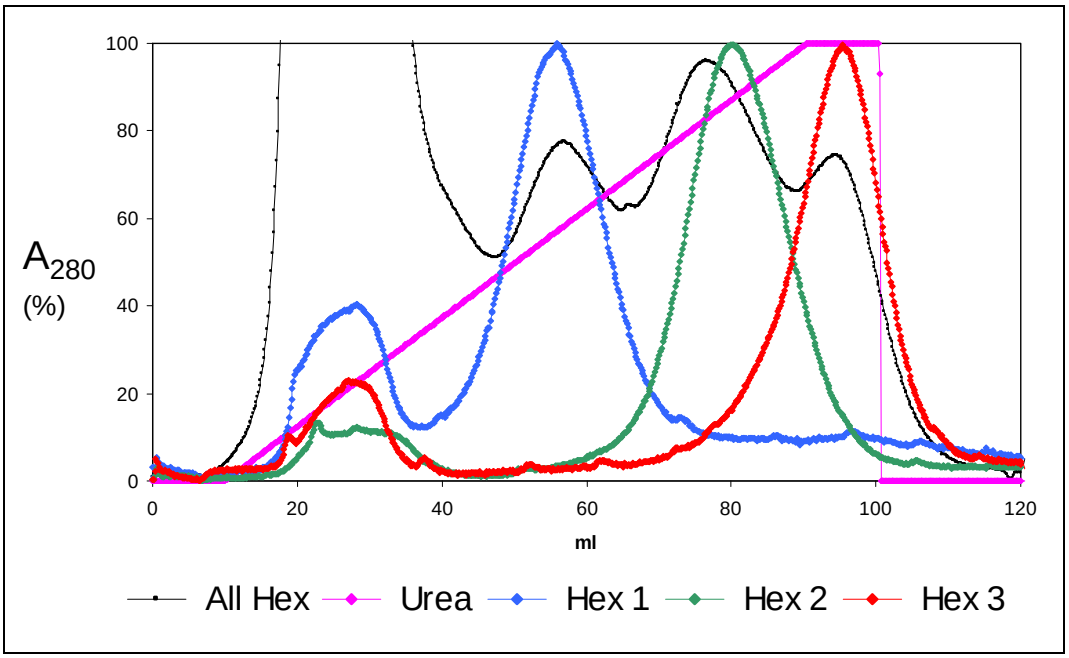



**Figure 2:12** Gelatin-affinity chromatography of cation-exchange purified Hex (black). The first peak centred on 30 ml is a non-binding impurity leaving the column. The Hex is separated into three major peaks: the first and second peaks are weaker binding (55 ml and 80 ml), but the third is a much tighter binding form which elutes between 90 and 100 ml. The three binding forms elute at their original urea concentration when re-run on the column (blue, green and red).

The Hex was detected in three main peaks, with the tightest binding form only eluting from the column at a urea concentration of around 2.6 M (assuming a void volume of 20 ml). Initially the multiple forms were thought to result from methionine oxidation and attempts were made to reduce this oxidation. Two different approaches were employed: firstly cells were grown to a much higher biomass before induction and

protein was collected after only 28 hours. Secondly, the amount of trace metals added to the fermentation was reduced. A combination of the two appeared to reduce the multiple oxidation states and increase the yield of the tightest binding form relative to the other two weaker binding forms. The $O_2$ concentration was kept at the lowest recommended level (20%) in order to avoid Met oxidation without inhibiting growth.

Each of the binding forms was taken in turn, desalted, re-run on the gelatin column and shown to elute at their original urea concentration indicating that there was little or no interchange between forms. All three were judged to be of the correct molecular weight by non-reducing SDS-PAGE (data not shown), but under reducing conditions the two weaker binding fragments were shown to be cleavage products of the full length Hex (Figure 2:13), held together only by disulphide bonds under non-reducing conditions.

The two weaker binding Hex fragments appeared to be proteolytic fragments, presumably cleaved by proteases in the fermentation media. N-terminal sequencing suggested the smaller band had the same N-terminus as the $^{8}$Fn1**-$^{9}$Fn1 (data not shown) module pair and the larger was assumed to be the Quad as the only other N-terminus detected was that of the $^{6}$Fn1. The full length Hex was present in all fractions, which may have been misfolded, wrongly disulphide linked protein, methionine oxidised protein, or deglycosylated material which was more easily eluted from the gelatin column by urea. The two weak-binding forms were not investigated further. The tightest binding form of the Hex was collected and was considered >95% pure from its 2100 bioanalyzer trace (Figure 2:14A).

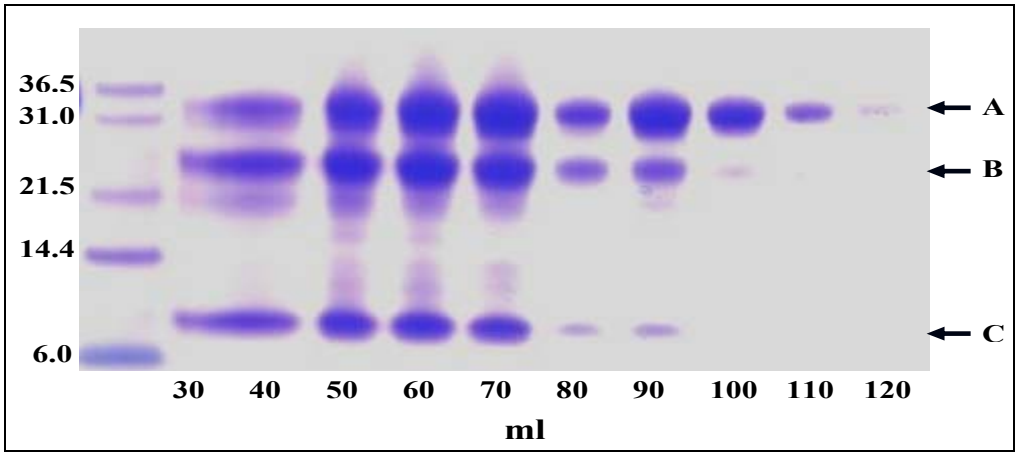



**Figure 2:13** A reducing SDS-PAGE gel of 10 ml gelatin affinity chromatography fractions of the Hex FN fragment (compare with Figure 2:12). The extent of proteolysis is observed when the disulphide bonds are reduced to show the Hex fragment (A) and two smaller proteolytic fragments (B and C).

### 2.3.9. Mass spectrometry

The oxidised molecular weight of the Hex by ESI-MS was calculated as 35,015.5 Da which is within 40 Da of that predicted (Figure 2:14B). The observed molecular weights of the Triple and the Quad were also 42 Da larger than expected, an observation previously noted by Dr. A.P. Pagett (personal communication). The extra

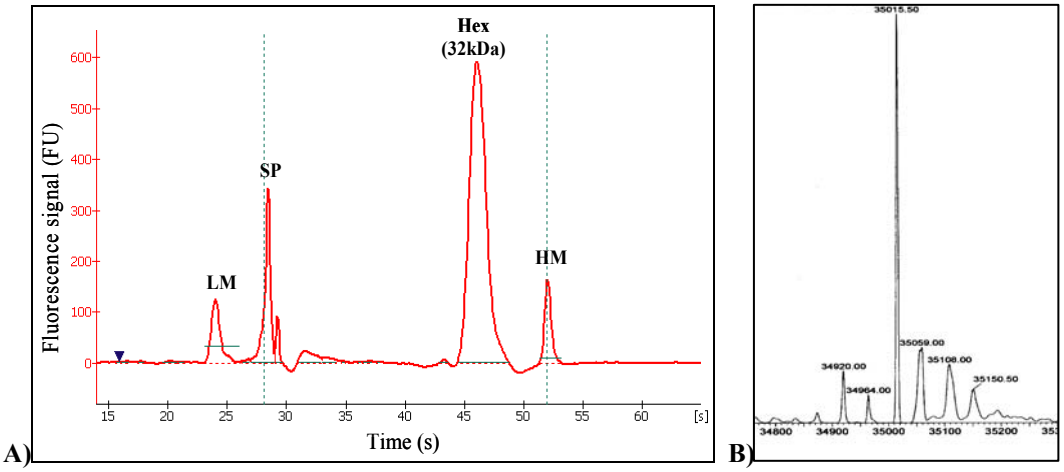



**Figure 2:14** A, capillary electrophoresis of the Hex fragment under reducing conditions that shows the protein as a single pure species. The system peak (SP), low (LM) and high (HM) molecular weight markers are common to all runs. The protein size is determined relative to protein standards (6-53 kDa) and calculated to be 32,000 Da (predicted MW 34,980 Da) which is within the observed error for samples in this system. B, molecular mass of the Hex fragment (35015.5 Da) as measured by ESI-MS.

mass could be due to the addition of an acetyl group (+42 Da), or carboxylation of Asp or Glu residues (+44 Da). N-terminal acetylation of Ser and Met residues of proteins produced in *P. pastoris* has been reported previously (Moerschell et al., 1990), (Johnson et al., 2003). The molecular weights of $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1, $^{7}$Fn1-$^{8}$Fn1** and $^{8}$Fn1**-$^{9}$Fn1 were within 1 Da of that expected.

### 2.3.10. N-terminal sequencing

All fragments were N-terminal sequenced and found to be correct: $^{6}$Fn1 constructs - YGH§VTDS, $^{7}$Fn1 constructs - AHEEI§TT and $^{8}$Fn1 constructs - DQ§IVDDI (where "§" is non-derivatised Cys) (data not shown).

### 2.3.11. Western blotting

The two primary antibodies (anti-FN and anti-$His_6$) were used to further confirm the identity of the FN fragments. The anti-FN antibody was specific to the gelatin binding domain and recognised the Hex and larger FN fragments (70 kDa and 90 kDa expressed by transient expression in HEK293T cells by Dr. D. Staunton and Dr. A.R. Aricescu). This antibody did not recognise recombinant fragments smaller than the Hex (Triple, Quad, $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1, $^{7}$Fn1-$^{8}$Fn1**, or $^{8}$Fn1**-$^{9}$Fn1) which suggested that the antibody recognised an epitope which is dependent on the quaternary structure of the domain. The two bands of the glycosylated Hex were presumably due to different length sugars as they resolved into a single band after treatment with Endo $H_f$ (Figure 2:15). The 70 kDa and 90 kDa fragments showed variation in sugar length, but also significant degradation, as a band in the 70 kDa fragment was of the same size as the glycosylated Hex.

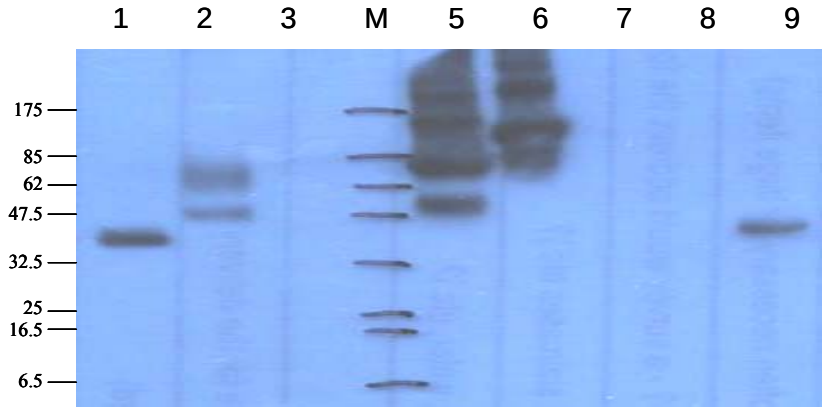



**Figure 2:15** Western blots of FN fragments detected with anti-FN antibody. Lanes: 1, Hex - deglycosylated, 2, Hex - glycosylated, 3, $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1, M, prestained markers (with molecular weights indicated on the left), 5, 70 kDa, 6, 90 kDa, 7, Triple, 8, $^{8}$Fn1**-$^{9}$Fn1, 9, Hex - deglycosylated (repeat of lane 1). The marker bands have been enhanced for clarity.

The anti-$His_6$ antibody recognised the $His_6$ tag on $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1, $^{7}$Fn1-$^{8}$Fn1** and $^{8}$Fn1**-$^{9}$Fn1 (where a $His_6$ tag was present) confirming that the C-terminus of the FN fragments was intact (data not shown).

### 2.3.12. Transient mammalian protein expression

Small-scale transfection was performed with several Hex DNA constructs (with different length linkers between the end of the gene and the $His_6$ tag); these were screened and the best expression clones were analysed by western blot analysis (data not shown). The best expression clone (Hex construct containing a short GGS linker between the C-terminus of the $^{9}$Fn1 and the C-terminal $His_6$ tag) was scaled up for large-scale transfection in roller bottles using 1 mg DNA for a 500 ml culture (performed by Dr. A.R. Aricescu). The cellular debris was removed by centrifugation, filtered through a 22 μm membrane, before loading onto cobalt-coated Talon® beads.

### 2.3.13. Protein purification of transiently expressed Hex

The small scale expression SDS-PAGE (data not shown) indicated that the yield of the best expressing construct was good but when scaled up expression levels were only 5 mg/litre (compared to the expected yield of 40 mg/litre from the small scale expression screen). The protein was passed through the Talon® resin column and then gel filtrated (Figure 2:16) as a final polishing step to achieve a purity of 95% as judged by SDS PAGE.

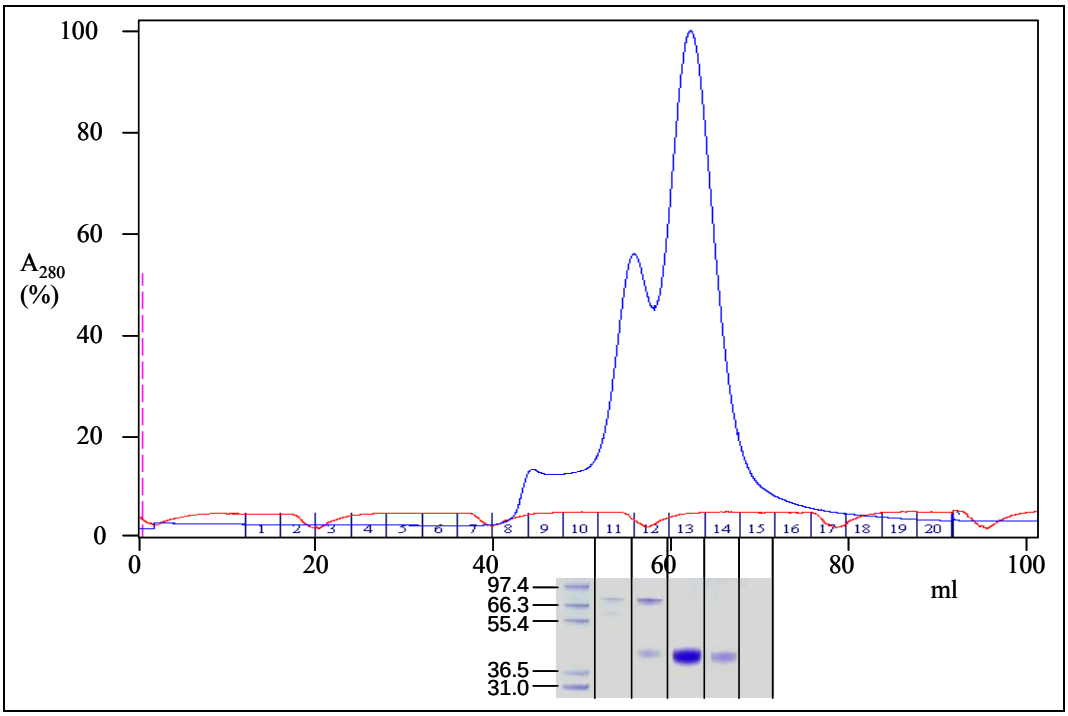



**Figure 2:16** Gel filtration profile of Hex obtained on a superdex gel filtration 75 column. Fractions were analysed for protein content by SDS PAGE gel and fractions 13 and 14 were judged to be the correct molecular weight to be pooled ready for crystallisation trials (sugars uncleaved).

## 2.4. Discussion

### 2.4.1. Vector design and cloning

The design and construction of the new pPICZλ shuttle vector streamlined the insertion and amplification of gene products to a single ligation step. Electroporation was a quick and effective step for insertion of the vector into *P. pastoris* cells where the natural process of homologous recombination allowed the gene to insert into the genomic DNA. Dose dependent selection on Zeocin™ indicated the highest expression clones, and SDS-PAGE analysis of the small scale expression confirmed the presence of protein. The process from gene amplification to isolation of high protein expression clone typically took 3-4 weeks.

The design and transient expression of protein in mammalian cells was an even swifter process as stable cell lines were not required. Target cloning to crystallisation trials was completed in two weeks.

### 2.4.2. Protein expression, purification, and analysis

Fermentation, harvesting, and purification of *P. pastoris* expressed protein took a further 1-2 weeks. The integrity of the protein from both *P. pastoris* and HEK293T cells was analysed. With the exception of the Hex, all the constructs expressed > 40 mg/litre, and were purified without being noticeably attacked by proteases. The Hex, Triple, and Quad all seemed to be 40-44 Da larger that predicted which could be due to acetylation on the N-terminus or within the first three modules of the protein. The other proteins were within 1 Da of the predicted mass.

In all cases, the proteins showed no signs of degradation over 2 weeks of storage at 4°C, and were shown to be stable at reasonably high protein concentrations (typically 1-2 mM). The proteins were considered suitable for NMR and gelatin binding functional studies, which allowed further analysis of their structure and biological properties.

## 2.5. References


Aricescu, A.R., Lu, W. and Jones, E.Y. (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr*, **62**, 1243-1250.

Baron, M., Norman, D., Willis, A. and Campbell, I.D. (1990) Structure of the fibronectin type 1 module. *Nature*, **345**, 642-646.

Bright, J.R., Pickford, A.R., Potts, J.R. and Campbell, I.D. (2000) Preparation of isotopically labeled recombinant fragments of fibronectin for functional and structural study by heteronuclear nuclear magnetic resonance spectroscopy. *Methods Mol Biol*, **139**, 59-69.

Cereghino, J.L. and Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev*, **24**, 45-66.

Cockett, M.I., Bebbington, C.R. and Yarranton, G.T. (1990) High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. *Biotechnology (N Y)*, **8**, 662-667.

Higgins, D.R. and Cregg, J.M. (1998) Introduction to Pichia pastoris. *Methods Mol Biol*, **103**, 1-15.

Houard, X., Germain, S., Gervais, M., Michaud, A., van den Brule, F., Foidart, J.M., Noel, A., Monnot, C. and Corvol, P. (2005) Migration-stimulating factor displays HEXXH-dependent catalytic activity important for promoting tumor cell migration. *Int J Cancer*, **116**, 378-384.

Johnson, S.K., Zhang, W., Smith, L.A., Hywood-Potter, K.J., Todd Swanson, S., Schlegel, V.L. and Meagher, M.M. (2003) Scale-up of the fermentation and purification of the recombinant heavy chain fragment C of botulinum neurotoxin serotype F, expressed in Pichia pastoris. *Protein Expr Purif*, **32**, 1-9.

Juge, N., Andersen, J.S., Tull, D., Roepstorff, P. and Svensson, B. (1996) Overexpression, purification, and characterization of recombinant barley alpha-amylases 1 and 2 secreted by the methylotrophic yeast Pichia pastoris. *Protein Expr Purif*, **8**, 204-214.

Kishore, R., Samuel, M., Khan, M.Y., Hand, J., Frenz, D.A. and Newman, S.A. (1997) Interaction of the NH2-terminal domain of fibronectin with heparin. Role of the omega-loops of the type I modules. *J Biol Chem*, **272**, 17078-17085.

Meissner, P., Pick, H., Kulangara, A., Chatellard, P., Friedrich, K. and Wurm, F.M. (2001) Transient gene expression: recombinant protein production with suspension-adapted HEK293-EBNA cells. *Biotechnol Bioeng*, **75**, 197-203.

Millard, C.J., Campbell, I.D. and Pickford, A.R. (2005) Gelatin binding to the $^{8}$Fn1-$^{9}$Fn1 module pair of human fibronectin requires site-specific N-glycosylation. *FEBS Lett*, **579**, 4529-4534.

Moerschell, R.P., Hosokawa, Y., Tsunasawa, S. and Sherman, F. (1990) The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine in vivo. Processing of altered iso-1-cytochromes c created by oligonucleotide transformation. *J Biol Chem*, **265**, 19638-19643.

Pickford, A.R. and O'Leary, J.M. (2004) Isotopic labeling of recombinant proteins from the methylotrophic yeast Pichia pastoris. *Methods Mol Biol*, **278**, 17-33.

Pickford, A.R., Smith, S.P., Staunton, D., Boyd, J. and Campbell, I.D. (2001) The hairpin structure of the $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2 fragment from human fibronectin enhances gelatin binding. *Embo J*, **20**, 1519-1529.

Staunton, D., Schlinkert, R., Zanetti, G., Colebrook, S.A. and Campbell, I.D. (2006) Cell-free expression and selective isotope labelling in protein NMR. *Magn Reson Chem*, **44 Spec No**, S2-9.

Sticht, H., Pickford, A.R., Potts, J.R. and Campbell, I.D. (1998) Solution structure of the glycosylated second type 2 module of fibronectin. *J Mol Biol*, **276**, 177-187.

Sugrue, R.J., Fu, J., Howe, J. and Chan, Y.C. (1997) Expression of the dengue virus structural proteins in Pichia pastoris leads to the generation of virus-like particles. *J Gen Virol*, **78 ( Pt 8)**, 1861-1866.

Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A. and Gingeras, T.R. (1987) Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. *Nucleic Acids Res*, **15**, 3859-3876.

Vakonakis, I., Staunton, D., Rooney, L.M. and Campbell, I.D. (2007) Interdomain association in fibronectin: insight into cryptic sites and fibrillogenesis. *Embo J*, **26**, 2575-2583.

# 3. Structural characterisation of FN fragments

## 3.1. Protein structure at atomic resolution

In recent years, nuclear magnetic resonance (NMR) and crystallography have led the way in determining three-dimensional protein structure at atomic resolution. These two technologies have often been used as complementary methods in characterising folded and partially folded states as well as measuring protein dynamics and determining structural interfaces. Both techniques have their strengths and weaknesses; crystallography can be used to solve the structure of huge complexes such as the ribosome (Schuwirth et al., 2005) and the bluetongue virus core (Grimes et al., 1998), but obtaining crystals is often a lengthy process and proteins may be forced into a non-native state. NMR can be used to verify or challenge these crystal structures by studying the proteins in solution where pH, temperature, and salt concentration can be altered with ease. Protein dynamics and ligand binding can also be measured by NMR.

Until recently, *ab initio* structure determination of proteins greater than ~30 kDa by NMR had been limited by the complexity and overlap of the data (Campbell and Downing, 1998). New advances in instrumentation, labelling methods, and spin physics have pushed the limits of structure determination by NMR to greater than 50 kDa (Sprangers et al., 2007). While these pioneering studies have increased the sensitivity of NMR to gain as much information as possible from the data, advances in NMR processing have allowed key pieces of information to be extracted from even larger datasets.

### 3.1.1. Protein NMR

NMR has proved invaluable in the study of FN modules; initial structure determination (Baron et al., 1990), module orientation (Hashimoto et al., 2000), ligand binding (Schwarz-Linek et al., 2003), and confirmation of crystal structures (Rudino-Pinera et al., 2007). Structural assignment of large proteins is reliant on complete data from many two- and three-dimensional experiments, whereas much information about protein folding can be deduced from a quick and simple one-dimensional or two-dimensional experiment. Here, both approaches are used; more data are needed for the assignment of the $^{8}$Fn1-$^{9}$Fn1 module pair, less data (but equally important) are needed to verify the correct folding of FN fragments.

### 3.1.2. The basic NMR experiment

In order to understand and apply NMR, it is convenient to discuss the phenomenon with reference to atomic nuclei as everyday objects (a classical picture), rather than through complicated mathematical equations (in a quantum mechanical sense). The only accurate way of describing the behaviour of atomic nuclei is by quantum mechanics, but here, a more simple explanation will suffice.

When a nucleus with non-zero spin (such as $^{1}$H, $^{13}$C, $^{15}$N etc.) is placed in a magnetic field, its nuclear dipole moment acts like a bar magnet that can align with or against the magnetic field. For spin ½ there are two quantised energy levels and movement between them can only occur if the atoms absorb or emit a photon of energy (Figure 3:1). At equilibrium, the populations of the states are not quite equal, with more of the nuclei in the lower energy state. If electromagnetic radiation is applied to the

system at the appropriate frequency, the population levels are disturbed and net absorption causes excitation to the higher energy state. Subsequent relaxation back to equilibrium can be analysed to give information about motions and interactions within the sample.

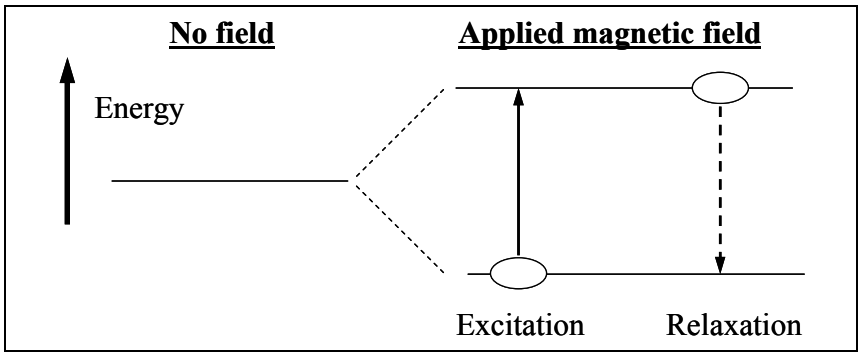



**Figure 3:1** Placing a spin ½ nucleus in a magnetic field causes the energy state to be split. The absorption of electromagnetic radiation results in excitation and through the release of energy the population will relax back to equilibrium (adapted from (Rattle, 1995) p. 8).

A spinning charged nucleus in a magnetic field will precess about the direction of the field at a rate called its Larmor frequency. This depends on the type of nucleus and is directly proportional to the strength of the applied field. In the case of a large number of nuclei, there is a finite resultant magnetisation in the z direction (called $M_z$), since the populations of the states are not quite equal (Figure 3:2).

If now a second, varying field ($B_1$) is applied to the nuclei, perpendicular to the applied static field (Bo), the resultant magnetisation ($M_z$) can be tilted away from the z axis into the xy plane, resulting in an $M_{xy}$ component. This radio frequency pulse is defined by its flip angle (duration and strength of the pulse) and the direction from which it is applied in the rotating frame, (e.g. 90°,+y). Over time the magnetisation returns back to equilibrium by a process called relaxation and the rate of relaxation

can be measured by the detector coil in an NMR magnet. The signal received by the detector is called free induction decay (FID), which by means of a mathematical integration known as the Fourier Transform (FT), allows the individual components in the sample to be viewed as a spectrum. The precise external chemical environment of the nucleus is reported as a difference in chemical shift from a standard.

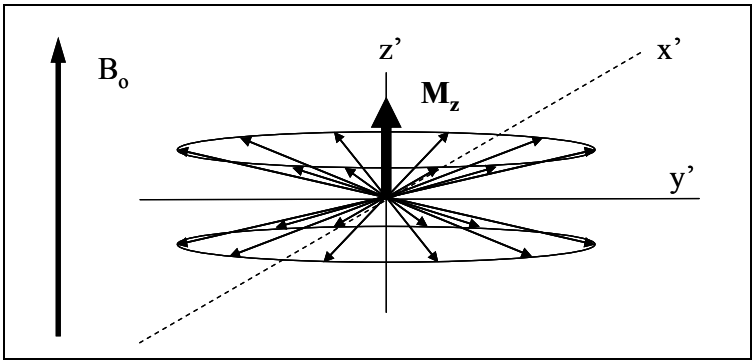



**Figure 3:2** The resultant magnetisation of nuclei in a magnetic field ($B_o$). Each individual nucleus has a rotating magnetic moment component in the xy plane but due to the large number of nuclei this is averaged out to zero net magnetisation in this plane. The equilibrium state of a large number of nuclei causes a small resultant magnetisation in the z plane (adapted from (Rattle, 1995) p. 14).

### 3.1.3. One-dimensional $^1$H NMR

One-dimensional $^1$H NMR is ideal for looking at small molecules but resolution becomes a problem with larger systems due to signal overlap. There is little chance of assigning a 10-40 kDa protein from a one-dimensional proton spectrum (due to peak overlap) but this can be used in preparatory analysis of a protein sample. A good dispersion of peaks generally indicates a folded protein. In very general terms, the far left of the spectrum contains resonances from aromatic and NH protons, the middle from aliphatic protons, and the far right methyl protons. One-dimensional spectra can also be used to judge the signal to noise ratio of the protein sample (which partly depends on concentration) and the purity of the sample. One-dimensional NMR was used here routinely to check the structural integrity of unlabelled fragments.

### 3.1.4. Two-dimensional NMR

A one-dimensional spectrum contains all of the chemical shift data from a sample but often in an inaccessible form. In order to assign the peaks in a spectrum, the resonances must be "spread out" into at least one other dimension. There are a number of NMR experiments in which peak overlap can be resolved:

### 3.1.5. COSY, TOCSY, and NOESY

Phase information from one group of spins can be passed on to another group of spins in a process called coherence transfer. Transfer can only occur between spins that are coupled in one of two ways; nuclei that are connected through up to three covalent bonds (scalar or through-bond coupling), or nuclei that are within 5Å of each other (through-space coupling). The COSY and TOCSY experiments produce two-dimensional connectivity maps of through-bond couplings, whereas NOESY experiments detect through-space connectivities. The presence of an off-diagonal peak (crosspeak) in the two-dimensional map demonstrates spin coupling between the two nuclei. The information from these experiments can be combined to give information about the environment of each amino acid and can give an indication of distance restraints between residues in space. The experiments can also serve as a starting point in structural assignment.

### 3.1.6. [$^{1}$H-$^{15}$N]-HSQC

Coherence can be transferred from protons to other nuclei and of particular use in protein NMR are $^{13}$C and $^{15}$N. For the heteronuclear single quantum coherence (HSQC) experiment, proteins must be isotopically labelled with these nuclei before

coherence can be transferred (although transfer can occur to some degree with natural abundance $^{13}C/^{15}N$), and the result is to achieve very good resolution of $^{1}H$ to $^{13}C$ or $^{15}N$ couplings. In the $^{15}N$-edited HSQC ([$^{1}H$-$^{15}N$]-HSQC) (Figure 3:3), the protons are excited, magnetisation is transferred to nitrogen, and then back to the protons. The result is a dispersed spectrum displaying all the crosspeaks between the HN protons and the backbone NH of each amino acid, as well as the $NH_2$ groups of the side chains of Asn and Gln, and of the aromatic HN protons of Trp and His.

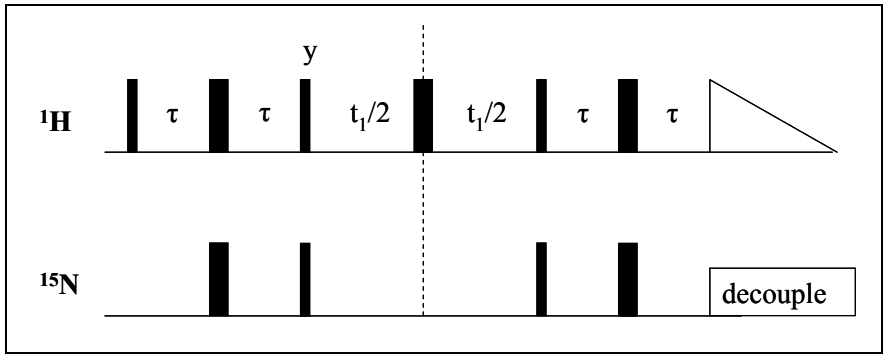



**Figure 3:3** Pulse sequence for the basic [$^{1}H$-$^{15}N$]-HSQC. The experiment is designed to transfer magnetisation between $^{1}H$ and $^{15}N$ and proton detection is used to maximise sensitivity. 90° pulses are indicated by thin bars and 180° pulses are indicated by thick bars (adapted from (Cavanagh, 2007) p. 537).

### 3.1.7. Heteronuclear {$^{1}H$-}$^{15}N$ NOE

There are a number of experiments for measuring the mobility of residues in a protein in solution and the heteronuclear {$^{1}H$-}$^{15}N$ NOE experiment gives a quick indication of residue mobility. Two experiments are recorded, one with (NOE) and one without (NONOE) $^{1}H$ saturation during the recycle delay, with the resulting spectra resembling the [$^{1}H$-$^{15}N$]-HSQC. The difference in peak intensity at each crosspeak gives an indication to the mobility of each residue.

### 3.1.8. Three-dimensional NMR

The assignment of proteins greater than 50 residues is greatly aided by "spreading out" the resonances into a third dimension, for example using $^{15}N$ labelled protein. Experiments such as three-dimensional gradient-enhanced [$^{1}H$-$^{15}N$]-TOCSY-HSQC and [$^{1}H$-$^{15}N$]-NOESY-HSQC are needed in order to fully assign a $^{15}N$ labelled protein. Triple resonance pulse sequences are available for doubly labelled proteins ($^{13}C$ and $^{15}N$) and these can help further in assignment.

### 3.1.9. NMR assignment

The first step towards solving the structure of a protein by NMR is to obtain sequence-specific resonance assignments. Once assigned, the information content of the NMR data becomes much more valuable. The information is not only useful in providing structural restraints in model building, but also for observing mobility of residues in the native protein, and calculating ligand binding constants.

There are four main steps in assigning the $^{1}H$ NMR lines of all hydrogen atoms in small proteins: initially the protein is suspended in $D_2O$ and the spin systems of non-labile protons of individual amino acid residues are identified using the $^{1}H$-$^{1}H$ through-bond connectivities. Secondly, the protein is resuspended in $H_2O$ to complete the amino acid spin system assignments of labile protons (typical amino acid spin systems are shown in Figure 3:4). Sequential $^{1}H$ spin systems are then identified from sequential NOE connectivities with the aim of constructing peptide segments that are unique, and finally the segments are fitted together using the protein amino acid sequence. A standard set of experiments recorded in $H_2O$ and $D_2O$ can be used to

achieve complete sequential assignment (i.e., [$^{1}$H-$^{15}$N]-HSQC, [$^{1}$H-$^{15}$N]-TOCSY-HSQC, [$^{1}$H-$^{15}$N]-NOESY-HSQC, [$^{1}$H-$^{1}$H]-DQF-COSY, [$^{1}$H-$^{1}$H]-TOCSY and [$^{1}$H-$^{1}$H]-NOESY (Pickford et al., 2001).

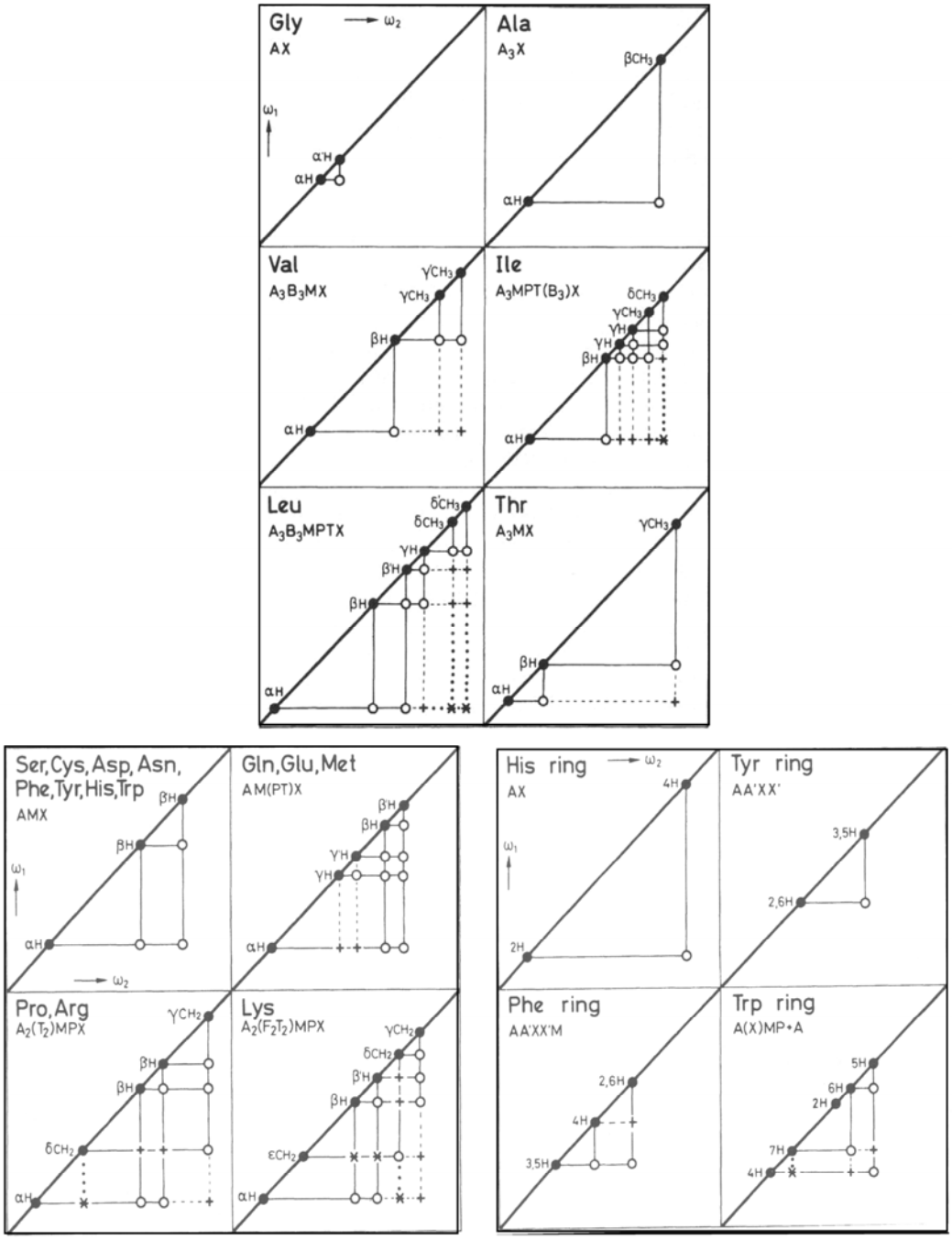



**Figure 3:4** Schematic representations of the two-dimensional proton spin systems for the common amino acids. The positions of the diagonal are represented by black circles. Open circles and solid lines indicate signals visible in a COSY spectrum, while crosses and dotted lines indicate additional signals seen using techniques such as TOCSY, which can be observed by changing the mixtime of the experiment. $\omega_1$ and $\omega_2$ = $^{1}$H chemical shift (ppm) (adapted from (Wuthrich, 1986) p. 136-137).

Sequential assignment can be greatly aided by knowledge of the secondary structure of the protein. By comparison of the $^{8}$Fn1-$^{9}$Fn1 primary sequence with other Fn1 modules, the module is predicted to be arranged as antiparallel β-sheets, although the exact position of turns and the module linker remain to be confirmed. Sequential assignment can therefore take advantage of common characteristics of antiparallel β sheet by tracing the short $^{1}$H-$^{1}$H contacts indicated in Figure 3:5.

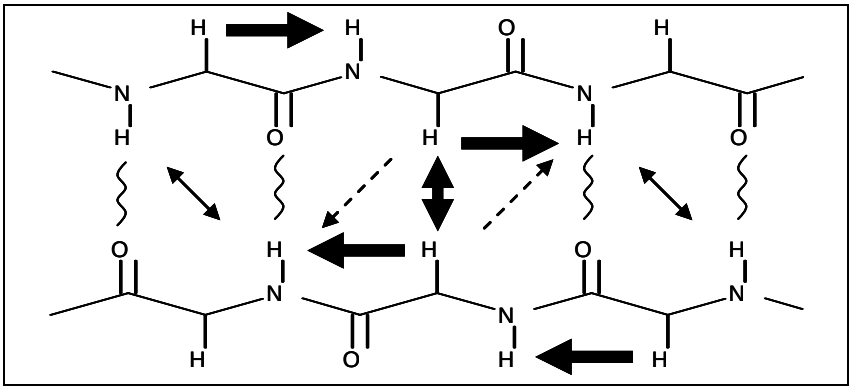



**Figure 3:5** Short sequential and short interstrand $^{1}$H-$^{1}$H distances in antiparallel β-sheets. Sequential intrastrand distances ($\alpha H_i$-$NH_j$) are indicated by thick horizontal arrows. Short interstrand distances are indicated by the thick double headed vertical arrow ($\alpha H_i$-$\alpha H_j$), thin double headed arrows ($NH_i$-$NH_j$) and dotted arrows ($\alpha H_i$-$NH_j$). All backbone distances indicated are between 3-4Å. Wavy lines indicate interstrand hydrogen bonds (adapted from (Wuthrich, 1986) p.126).

In this study, NMR was used to confirm the structure of all of the FN fragments expressed in *P. pastoris*. The secondary structure of the $^{8}$Fn1*-$^{9}$Fn1 module pair was characterised by NMR.

## 3.2. Methods

### 3.2.1. NMR spectroscopy

All NMR experiments were recorded in-house on spectrometers operating at 500, 600, 750, and 950 MHz. Samples were prepared by dissolving the [*u*-$^{15}$N]-FN fragments to a final concentration of 1.0 mM in either 95% $H_2O$/5% $D_2O$ or 99.9% $D_2O$, each containing 150 mM NaCl, 20 mM $KH_2PO_4$, 1.0 mM 1,4-dioxane, and were adjusted to pH 6.6 using NaOH. Spectra used to assign $^{8}$Fn1*-$^{9}$Fn1 were acquired at 45°C on the 600 MHz spectrometer after judging that these conditions gave the best signal. Further spectra of $^{8}$Fn1*-$^{9}$Fn1 and all other NMR data were collected at 25°C. One-dimensional and two-dimensional NMR spectra were used to confirm the correct folding of FN fragments.

### 3.2.2. NMR processing and analysis

Data was manipulated by first converting the data into a UNIX® compatible file and processed with NMRPipe (Delaglio et al., 1995), using NMRDraw as the graphical interface. Files generated by Charlie Taylor and Dr. I. Vakonakis were used as data processing templates. Once optimised, the data was converted to an NMRview file format for further analysis.

### 3.2.3. Data conversion

The number of points and the apodization size (point at which the window function tends to zero) was defined. The spectrometer operating frequency was entered, and the spectrum was referenced to the internal standard 1,4-dioxane at 3.743 ppm, with

indirect referencing in the $^{15}N$ dimension using a $^{15}N/^{1}H$ frequency ratio of 0.101329118 (Wishart et al., 1995).

### 3.2.4. Data processing

The following functions were applied to optimise the presentation of the data; the FID was aligned with the horizontal axis to improve the appearance of the baseline. A window function was applied (sine multiplication (SP)) to achieve the best ratio of spectral resolution to signal/noise, the FID was zero-filled to increase the digital resolution, and finally Fourier transformed to produce the final spectrum. The quality of the spectrum was assessed in NMRdraw.

### 3.2.5. Data analysis

The programme NMRView5.0 (Bruce Johnson, One Moon Scientific, Inc.) was used for data analysis including the assignment of $^{8}$Fn1*-$^{9}$Fn1. The $^{8}$Fn1*-$^{9}$Fn1 module pair was assigned using the method of spin system identification followed by sequential assignment. These assignments were cross-referenced with preliminary assignments of the $^{8}$Fn1*-$^{9}$Fn1 module pair recorded at 25°C as conducted by Dr. A.R. Pickford.

## 3.3. Results

### 3.3.1. NMR analysis of $^{8}$Fn1**-$^{9}$Fn1

The singly glycosylated $^{8}$Fn1*-$^{9}$Fn1 (R503K) module pair (N-linked at Asn511) was expressed and purified in *P. pastoris* as previously described (Millard et al., 2005). This glycoform was assumed to be most similar in structure to $^{8}$Fn1**-$^{9}$Fn1 (N497Q, R503K) and this was confirmed by comparing their [$^{1}$H-$^{15}$N]-HSQC spectra. Indeed, the overlay of the crosspeaks at 15°C was very good (apart from the mutated N497 residue), but as the temperature was raised, the crosspeaks of $^{8}$Fn1**-$^{9}$Fn1 began to lose signal (Figure 3:6).

In addition to temperature, the quality of the spectra was affected by the field strength applied. The quality of the NMR spectra from both proteins was vastly improved by using a lower field magnet (compare Figure 3:6 and Figure 3:7A). Peaks observed at lower field (500MHz) were lost at higher field (750MHz). Certain crosspeaks were completely lost from the $^{8}$Fn1**-$^{9}$Fn1 spectrum in a process called exchange broadening. Exchange broadening had a less pronounced effect on the quality of the $^{8}$Fn1*-$^{9}$Fn1 spectrum and more crosspeaks were visible at 500MHz.

It had been hoped that the two proteins would behave identically in solution, but the mutation introduced in $^{8}$Fn1**-$^{9}$Fn1 to increase the yield of homogeneous protein, produced lower quality spectra (either through aggregation or exchange broadening). The $^{8}$Fn1*-$^{9}$Fn1 module pair construct produced high quality spectra and data was collected on this protein for sequential assignment.

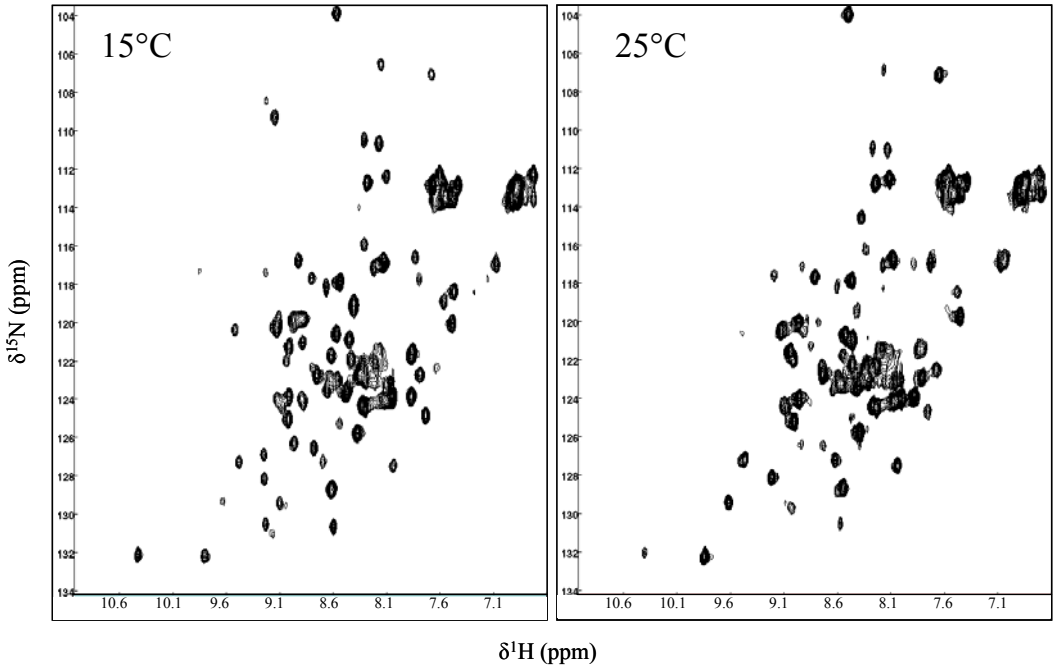



**Figure 3:6** Direct comparison of $^{8}$Fn1**-$^{9}$Fn1 (N497Q, R503K) [$^{1}$H-$^{15}$N]-HSQC spectra at 15°C and 25°C using identical experimental parameters for the 750 MHz spectrometer. The increase in temperature causes crosspeaks to be lost at 25°C.

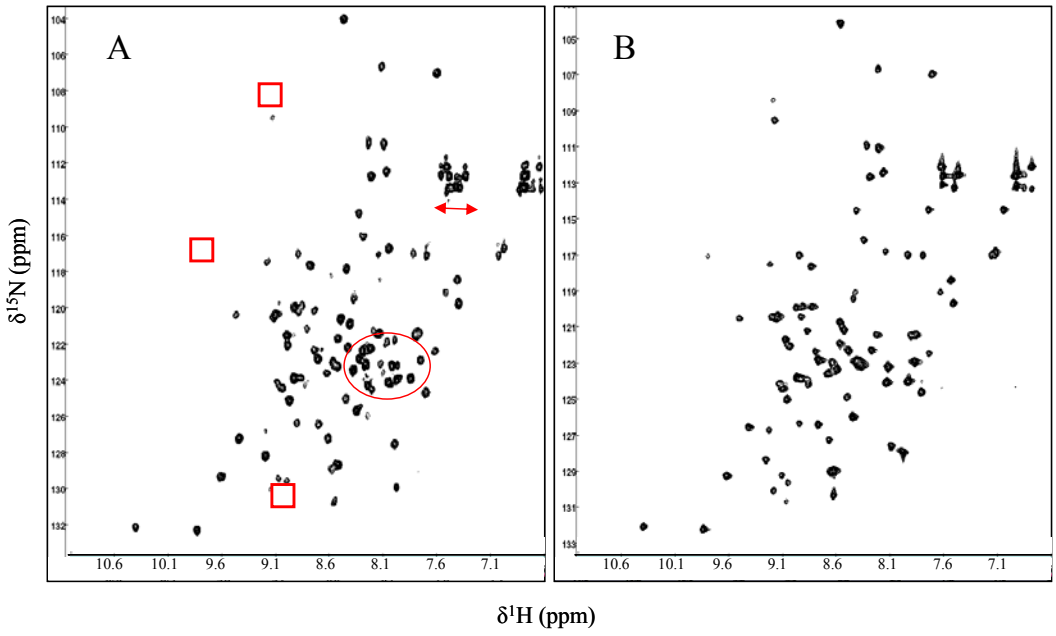



**Figure 3:7** [$^{1}$H-$^{15}$N]-HSQC spectra at 25°C. A, $^{8}$Fn1**-$^{9}$Fn1 (N497Q, R503K) and B, $^{8}$Fn1*-$^{9}$Fn1 (R503K) on the 500 MHz spectrometer (four times more scans were needed on the 500 MHz compared to the 750 MHz spectrometer for a good signal to noise ratio). The $^{8}$Fn1**-$^{9}$Fn1 and $^{8}$Fn1*-$^{9}$Fn1 module pair spectra differ very slightly; mutated Asn497 side chain crosspeaks disappear (red arrows), extra crosspeaks are present due to the C-terminal extension (majority of new peaks circled in red), and several crosspeaks disappear due to exchange processes (missing peaks highlighted with red squares).

### 3.3.2. Sequential assignment of $^{8}$Fn1*-$^{9}$Fn1

The two-dimensional and three-dimensional data of the $^{8}$Fn1*-$^{9}$Fn1 (R503K) module pair was well dispersed with little peak overlap allowing for rapid sequential assignment at 45°C (as the signal was best at this temperature). In $D_2O$ solution the labile protons were exchanged with deuterium and so the $^{1}$H NMR spectrum contained only carbon-bound hydrogen atoms. Initially the ring protons of the four aromatic residues and unique spin systems of Thr, Val, and the single GlcNAc were assigned as far as possible from the TOCSY (Figure 3:8) and COSY (Figure 3:9A) data using the random coil $^{1}$H chemical shifts as a guide (pH 7.0, 35°C), (Wuthrich, 1986). Then the regions of $^{1}$H chemical shifts for Leu, Pro, Arg, and Lys, the AMX spins (Ser, Cys, Asp, Asn, His, Phe, Tyr, and Trp), and Gln, Glu, and Met were identified. Finally crosspeaks for Gly, Ala, and Ile were tentatively assigned although many of the crosspeaks could not be confirmed due to peak overlap. The aromatic spin systems were linked to the non-liable backbone proton through crosspeaks in the NOESY data by extending to the already identified spin systems (Figure 3:9B).

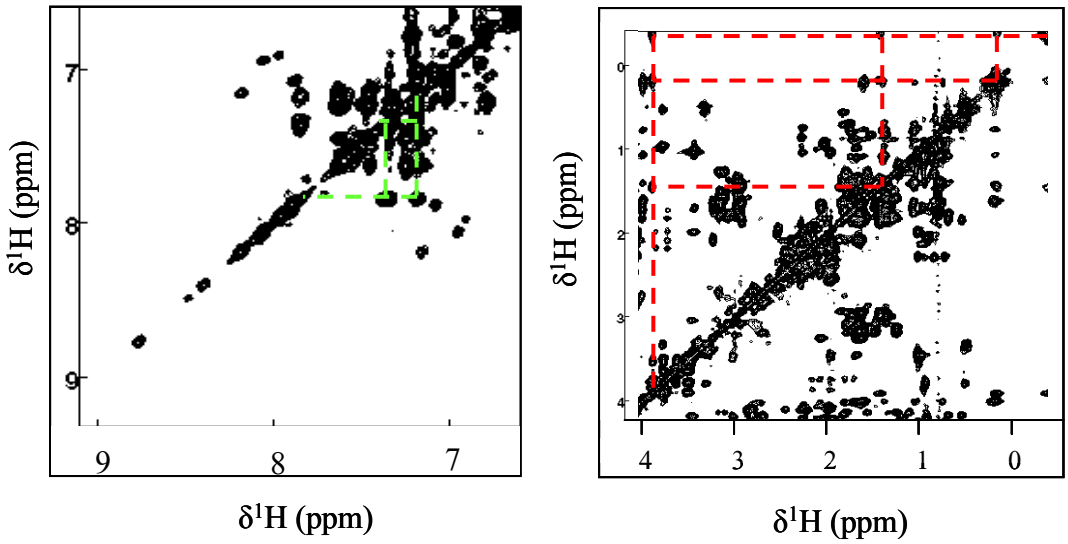



**Figure 3:8** Selected areas from the $D_2O$ [$^{1}$H-$^{1}$H]-TOCSY spectrum of $^{8}$Fn1*-$^{9}$Fn1. The spin system corresponding to the aromatic protons of Phe515 is highlighted on the left (green) and the spin system corresponding to Val489 is highlighted on the right (red).

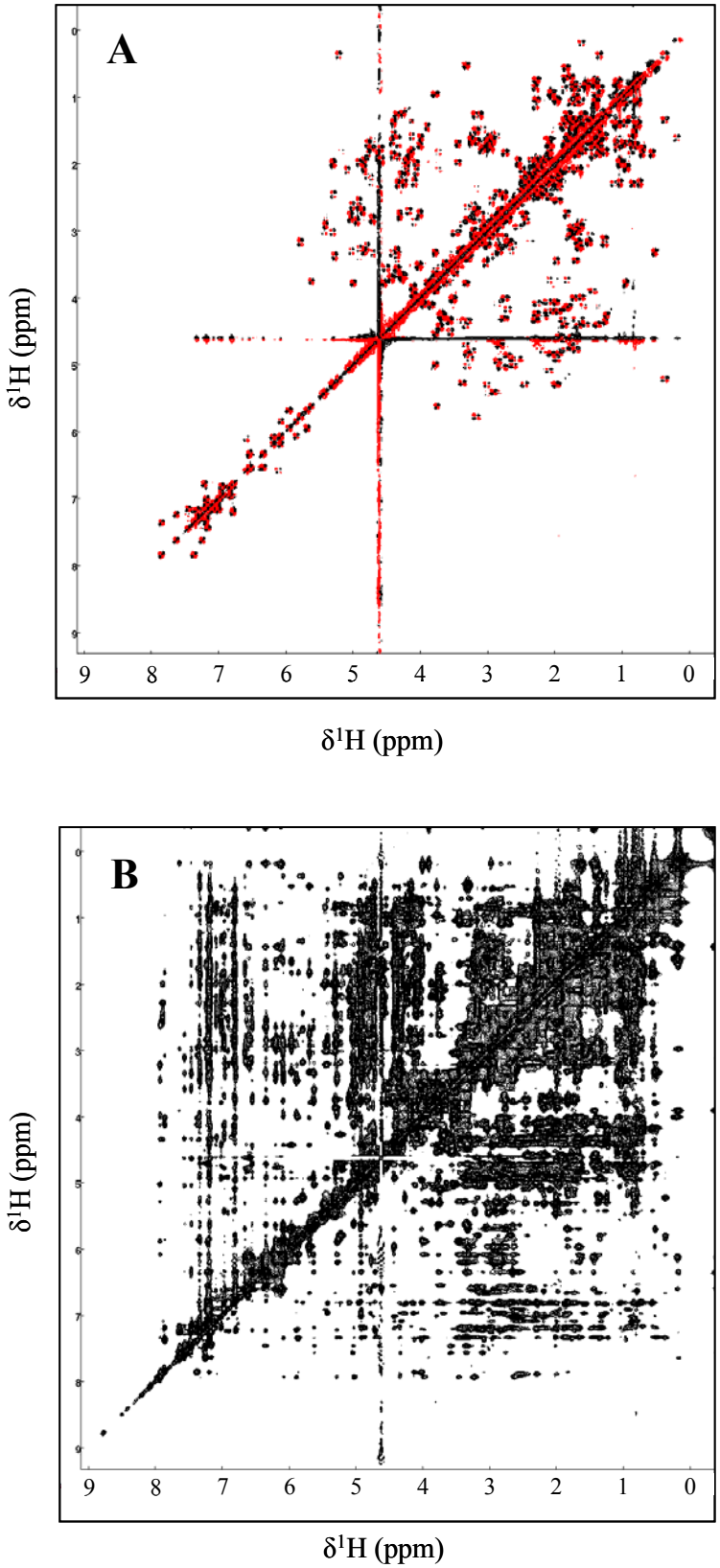



**Figure 3:9** $D_2O$ spectra of $^{8}$Fn1*-$^{9}$Fn1; A, $^{1}$H-COSY and B, 100 ms mixtime $^{1}$H-NOESY. The COSY experiment presents a map of the complete scalar coupling network in the macromolecular structure. The NOESY experiment contains through-space couplings between nuclei separated by less than 5 Å.

$H_2O$ $^{1}H$ NMR spectra were collected to extend the spin system assignment. This confirmed the assignments and allowed the Arg and Pro residues to be uniquely identified with the addition of labile protons (Pro is identified through the absence of labile protons). Three-dimensional data was then used to sequentially step through the sequence and resolve areas of crosspeak overlap. The secondary structure of the protein was predicted to be β-sheet in nature by comparison to other Fn1 modules and this information aided the "walk" through the sequence. By the identification of short sequential intra- and inter-strand backbone $^{1}H$-$^{1}H$ crosspeaks the β-sheets were sequentially assigned. Interstrand NOEs between protons across the β-sheets were also recorded to confirm the contacts proposed by the secondary structure predictions.

A temperature titration allowed the assignments at 45°C to be transferred to the 25°C [$^{1}H$-$^{15}N$]-HSQC spectrum and were used to further Dr. A.R. Pickford's partial assignment of the protein. In total, 73 of the backbone resonances were assigned in the 25°C [$^{1}H$-$^{15}N$]-HSQC, with the addition of three Trp sidechain, two of the Asn sidechain, and four of the Gln sidechain resonances (Figure 3:10). Residues which remained unassigned were largely in loop regions and included the D to E loop of the $^{9}$Fn1 which was thought to have little structural character. Further experiments used to characterise the flexibility of the protein were conducted at 25°C to prolong the lifetime of the sample. Once assigned, direct comparison of the $^{8}$Fn1**-$^{9}$Fn1 spectrum and the $^{8}$Fn1*-$^{9}$Fn1 spectrum allowed assignment of the $^{8}$Fn1**-$^{9}$Fn1 crosspeaks. The two proteins were considered share the same overall fold due to the similarity of the two HSQC spectra. After this comparison, a C-terminal $His_6$ tag was engineered onto the $^{8}$Fn1**-$^{9}$Fn1 module pair to allow identification of the module

pair by western blot. This addition did not significantly affect the NMR spectra other than by obvious addition of new peaks from C-terminal residues (data not shown).

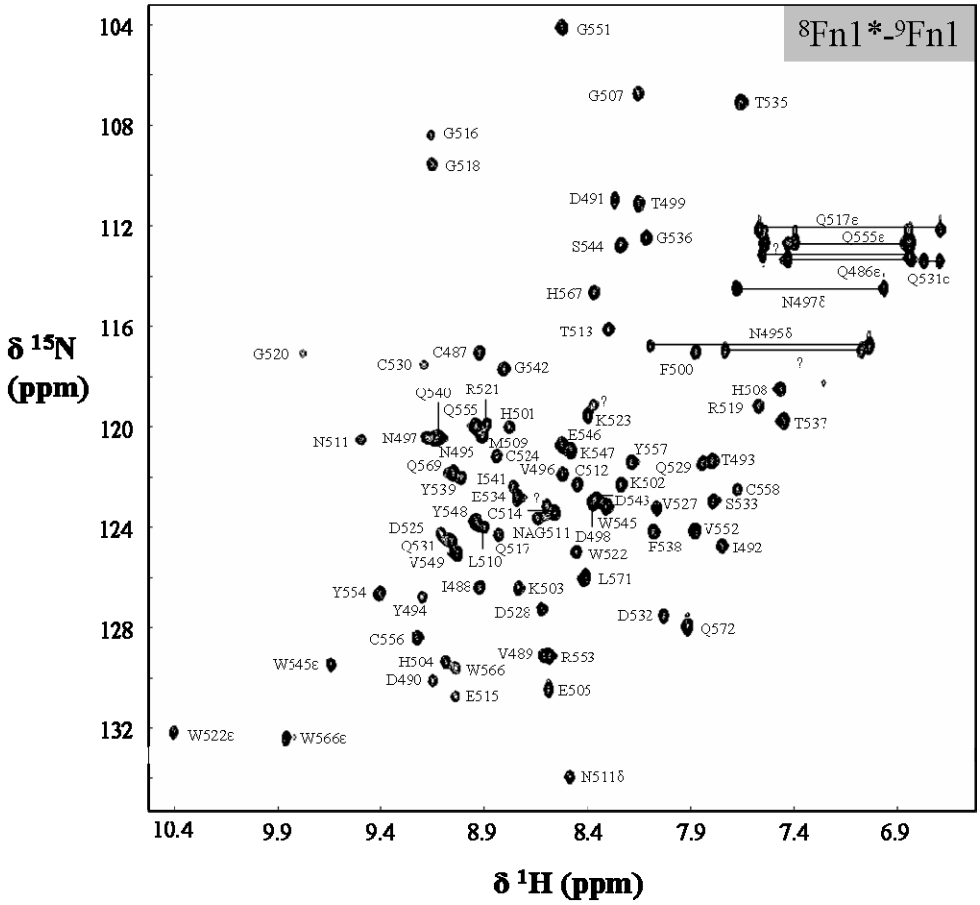



**Figure 3:10** The 90% assigned [$^{1}$H-$^{15}$N]-HSQC spectrum of the $^{8}$Fn1*-$^{9}$Fn1 module pair recorded on the 750MHz spectrometer at 25°C.

### 3.3.3. Secondary structure of $^{8}$Fn1*-$^{9}$Fn1

The secondary structure of $^{8}$Fn1*-$^{9}$Fn1 was mapped using inter-strand NOEs. Both modules were found to have a typical Fn1 module fold (Figure 3:11) and the observed hydrogen bonding pattern agrees well with existing structures (Potts et al., 1999), (Rudino-Pinera et al., 2007). The $^{8}$Fn1**-$^{9}$Fn1 module pair was judged to be folded in an identical manner due to the similarity of the [$^{1}$H-$^{15}$N]-HSQC spectra.

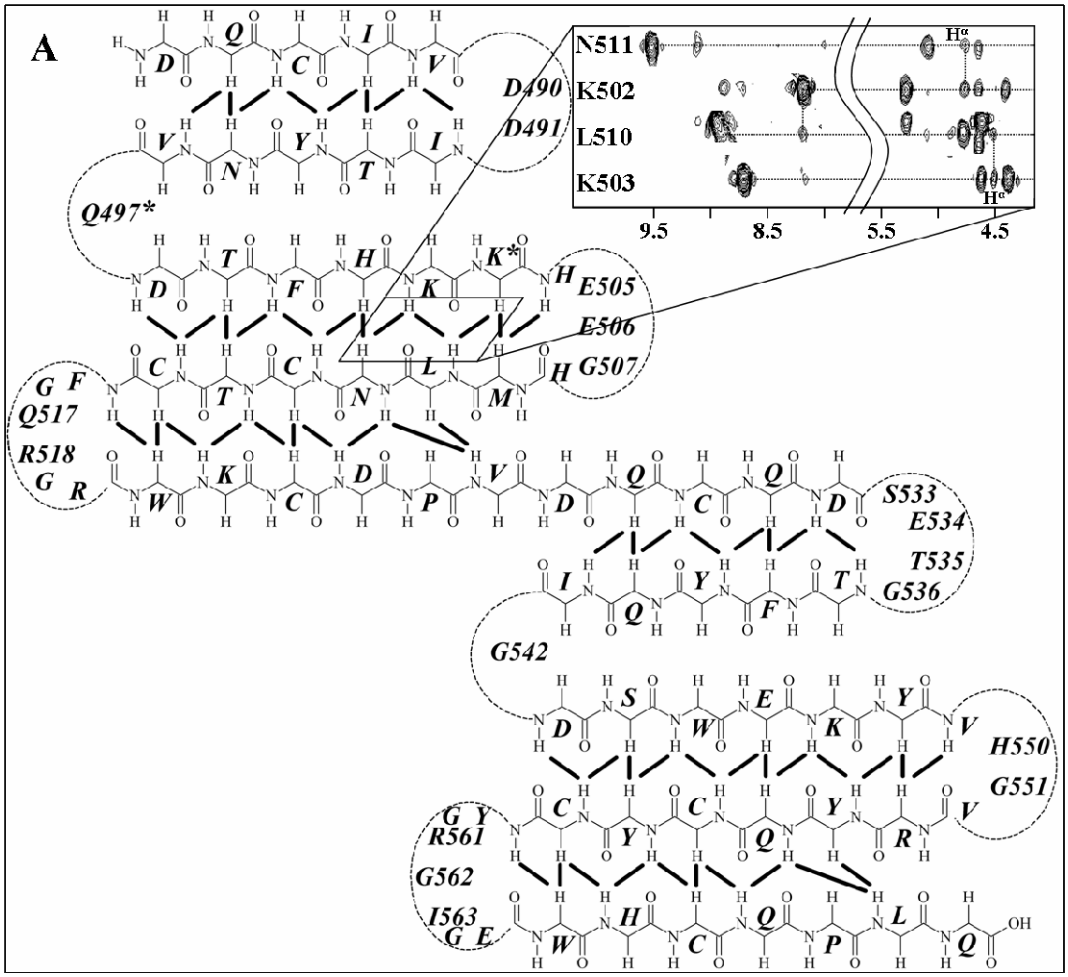



**Figure 3:11** Secondary structure of the $^{8}$Fn1*-$^{9}$Fn1 module pair. Each module adopts a typical Fn1 fold, with a two-stranded antiparallel β-sheet folded over a three-stranded β-sheet. Interstrand NOEs between protons are indicated by bold lines within the β-sheets. The 3D-NOESY strips (inset) show the interstrand $N_H$-$H_\alpha$ and $N_H$-$N_H$ NOEs between four of the residues in the $^{8}$F1 module. The stars indicate the position of the mutated residues N497Q and R503K.

### 3.3.4. NMR analysis of $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1

The one-dimensional $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 spectra were well dispersed, indicating that the fragments were fully folded (Figure 3:12D). Some of the peaks could be tentatively assigned by comparison with the one-dimensional spectra of $^{7}$Fn1 (Dr. M. Baron, D. Phil thesis) and $^{8}$Fn1**-$^{9}$Fn1, especially up-field where the methyl peaks were well dispersed. Observing the proteins by one-dimensional NMR was sufficient to confirm the folding of the FN fragments.

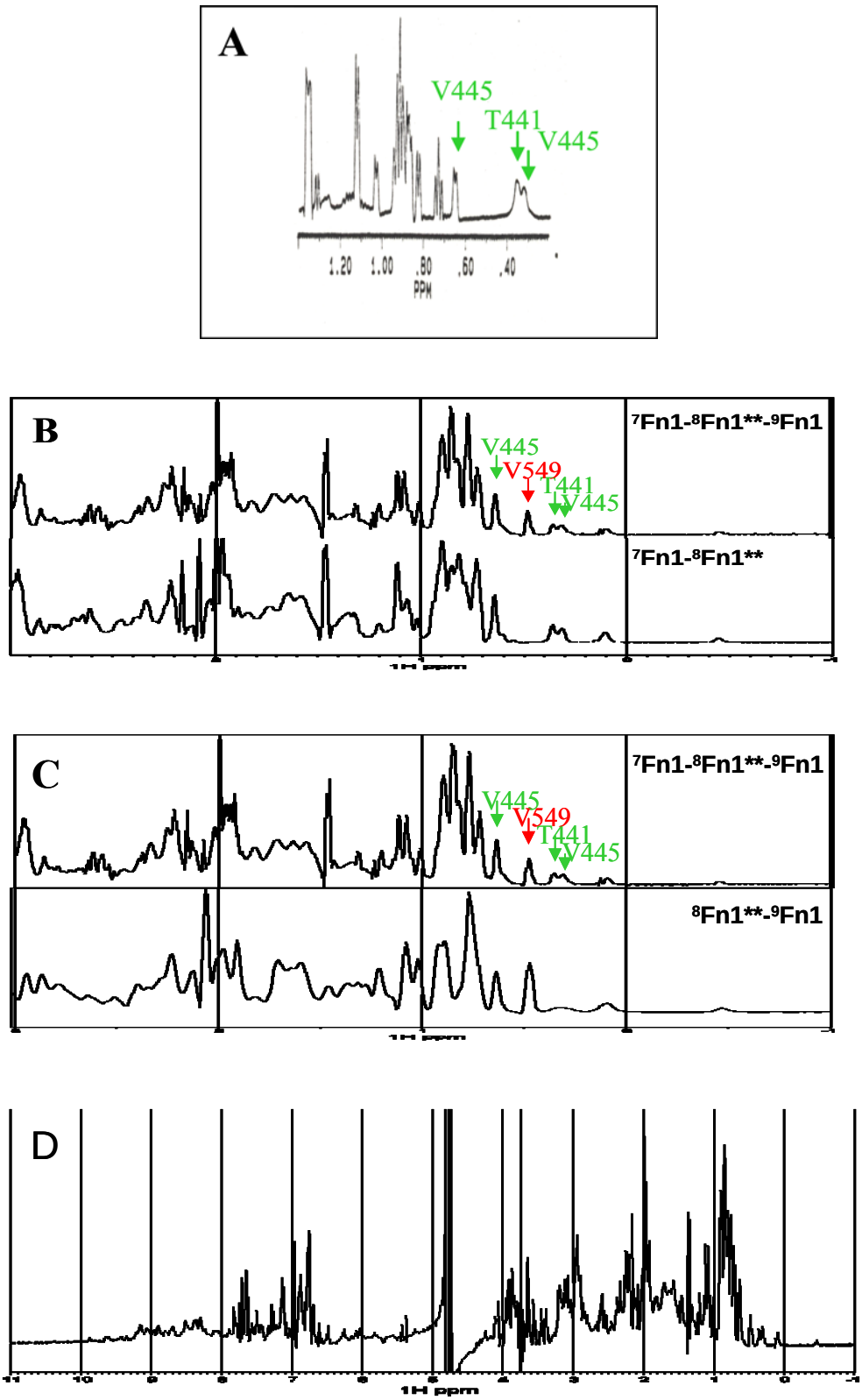



**Figure 3:12** One-dimensional spectra of FN fragments. A, methyl peaks of the $^{7}$Fn1 previously assigned (Baron, M., D. Phil thesis). B, methyl peaks of $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 and $^{7}$Fn1-$^{8}$Fn1**. C, methyl peaks of $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 and $^{8}$Fn1**-$^{9}$Fn1. D, the full one-dimensional spectrum of $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1. The $^{7}$Fn1 residues are labelled in green (T441 and V445) and the $^{9}$Fn1 residue is labelled in red (V549).

### 3.3.5. NMR characterisation of the Hex at 950MHz

Study of the Hex was greatly aided by the highest field NMR spectrometer in the world. The 950 MHz spectrometer successfully collected a well resolved two-dimensional spectrum of 150 μM singly labelled ($^{15}$N)-Hex (Figure 3:13). The dispersed crosspeaks of the [$^{1}$H-$^{15}$N]-HSQC spectrum indicated that the sample contained some monomeric folded protein (aggregation would cause a reduction in dispersion of crosspeaks). The spectrum of the Hex was not assigned at this stage due to low expression levels in *P. pastoris* and the large size of the protein.

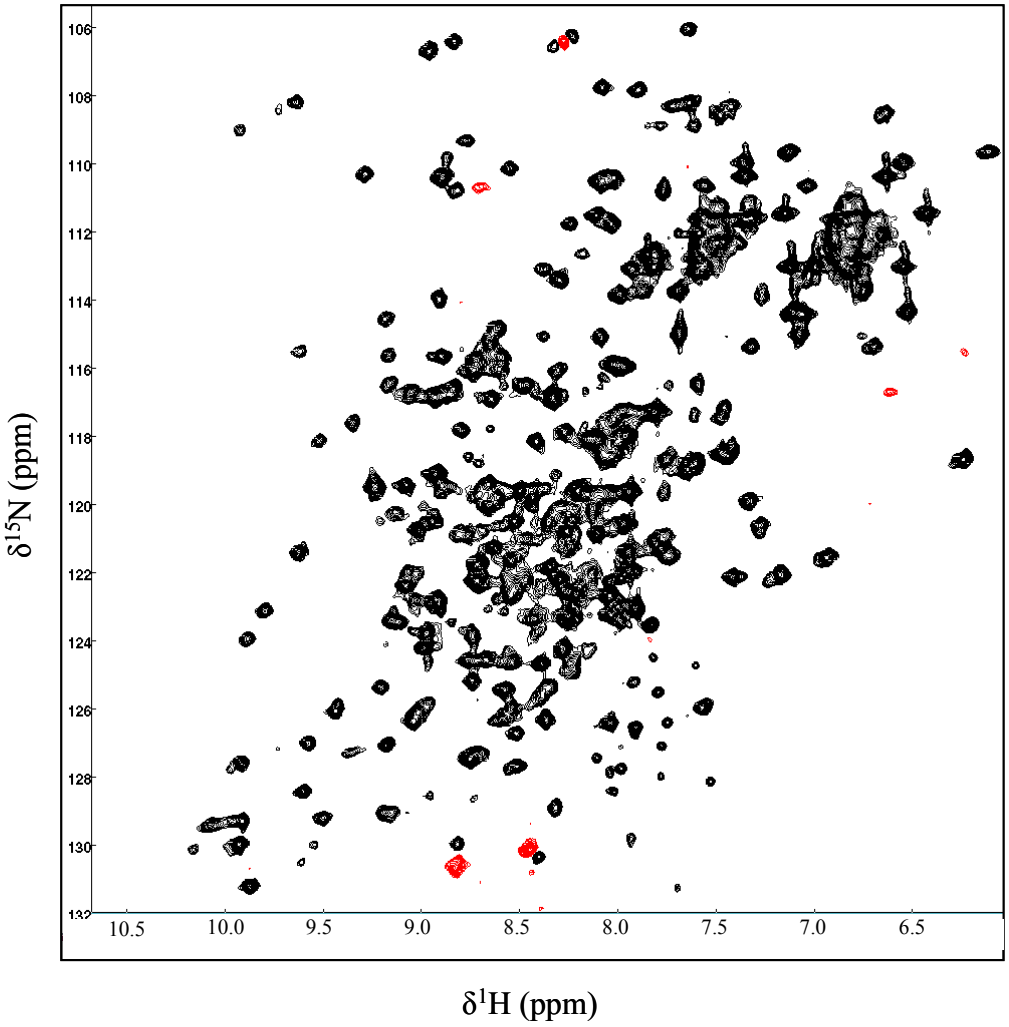



**Figure 3:13** The Hex [$^{1}$H-$^{15}$N]-HSQC spectra. The peaks are well dispersed which indicates a folded protein. A single crosspeak at 12.04, 108.80 is not shown and is caused by a residue (possibly histidine) in a very unusual environment. Negative peaks are shown in red.

### 3.3.6. Crystallisation trials of the Hex

In order to characterise the Hex at atomic resolution, the protein produced in both *P. pastoris* and mammalian HEK293T cells was put into crystal trials, for X-ray crystallography. The protein was concentrated to 5.8 mg/ml and tested under 384 conditions (4x 96 wells), at 4°C, 12°C, and 25°C, both in house and at the "Oxford Protein Production Facility" (The Henry Wellcome Building for Genomic Medicine, Oxford). Initial screens produced promising results (Figure 3:14) but crystals were not forthcoming in subsequent narrower screens. The glycosylation pattern on the Hex in mammalian cell expression was not analysed.



**Figure 3:14** A potential crystal from Hex expressed in HEK293T cells. Screen conditions were: 35% (v/v) 2-methyl-2,4-pentanediol (MPD), 100mM Na/K phosphate pH 7.0 (Hampton Research).

## 3.4. Discussion

### 3.4.1. The recombinant fibronectin fragments are folded

NMR was successfully used to characterise the entire range of *P. pastoris*-expressed gelatin binding domain fragments used in this study at atomic resolution. The large dispersion of crosspeaks in one-dimensional and two-dimensional NMR spectra confirmed that the *P. pastoris* recombinantly expressed FN fragments were fully folded. All of the fragments could be concentrated (0.5-1mM), without aggregation, which speeded up the experiments and reduced the need for excessive equipment time. Around half of the crosspeaks in the Hex [$^{1}$H-$^{15}$N]-HSQC spectrum could be resolved using the 22.3T (950 MHz $^{1}$H Larmor frequency) superconducting magnet.

### 3.4.2. The N497Q mutation subtly alters the behaviour of the protein

NMR spectra of $^{8}$Fn1*-$^{9}$Fn1 and $^{8}$Fn1**-$^{9}$Fn1 showed significant similarity to suggest that the single point mutation had made little difference to the structure of the protein. The quality of the NMR spectra was vastly improved for both proteins by using a lower magnetic field which relieved exchange broadening effects seen at higher field. Exchange broadening remained a problem to a greater extent (disappearance of crosspeaks) at 500MHz in the $^{8}$Fn1**-$^{9}$Fn1 spectrum which indicated a subtle difference in the exchange regime of this mutant in comparison to the $^{8}$Fn1*-$^{9}$Fn1.

### 3.4.3. The $^{8}$Fn1**-$^{9}$Fn1 module pair contains typical Fn1 folds

The $^{8}$Fn1*-$^{9}$Fn1 module pair was assigned at 45°C by sequential spin system assignment and the secondary structure was mapped using inter-strand NOEs.

Several solution structures of Fn1 module pairs have already been calculated by NMR (Potts et al., 1999; Rudino-Pinera et al., 2007). In this study, both modules of $^{8}$Fn1*-$^{9}$Fn1 were found to have a typical Fn1 module fold. The observed hydrogen bonding pattern agreed well with existing structures (see Fig 3.11). Assignments were transferred to the 25°C [$^{1}$H-$^{15}$N]-HSQC spectrum by temperature titration, and by comparison, used to assign the $^{8}$Fn1**-$^{9}$Fn1 25°C [$^{1}$H-$^{15}$N]-HSQC spectrum for later studies (chapter 4 and chapter 5).

### 3.4.4. Further study following NMR assignment

The assignment of the [$^{1}$H-$^{15}$N]-HSQC spectrum allowed further investigation into the structure and function of the module pair. The NMR assignment aided work with peptide titrations with collagen peptides (discussed in chapter 4) and allowed further investigation of the effect of single amino acid mutations made to the protein (discussed in chapter 5).

## 3.5. References


Baron, M., Norman, D., Willis, A. and Campbell, I.D. (1990) Structure of the fibronectin type 1 module. *Nature*, **345**, 642-646.

Campbell, I.D. and Downing, A.K. (1998) NMR of modular proteins. *Nat Struct Biol*, **5 Suppl**, 496-499.

Cavanagh, J. (2007) *Protein NMR Spectroscopy*. Elsevier Academic Press London.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, **6**, 277-293.

Grimes, J.M., Burroughs, J.N., Gouet, P., Diprose, J.M., Malby, R., Zientara, S., Mertens, P.P. and Stuart, D.I. (1998) The atomic structure of the bluetongue virus core. *Nature*, **395**, 470-478.

Hashimoto, Y., Smith, S.P., Pickford, A.R., Bocquier, A.A., Campbell, I.D. and Werner, J.M. (2000) The relative orientation of the fibronectin $^{6}$Fn1-$^{1}$Fn2 module pair: a 15N NMR relaxation study. *J Biomol NMR*, **17**, 203-214.

Ingham, K.C., Brew, S.A. and Novokhatny, V.V. (1995) Influence of carbohydrate on structure, stability, and function of gelatin-binding fragments of fibronectin. *Arch Biochem Biophys*, **316**, 235-240.

Millard, C.J., Campbell, I.D. and Pickford, A.R. (2005) Gelatin binding to the $^{8}$Fn1-$^{9}$Fn1 module pair of human fibronectin requires site-specific N-glycosylation. *FEBS Lett*, **579**, 4529-4534.

Pickford, A.R., Smith, S.P., Staunton, D., Boyd, J. and Campbell, I.D. (2001) The hairpin structure of the $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2 fragment from human fibronectin enhances gelatin binding. *Embo J*, **20**, 1519-1529.

Potts, J.R., Bright, J.R., Bolton, D., Pickford, A.R. and Campbell, I.D. (1999) Solution structure of the N-terminal F1 module pair from human fibronectin. *Biochemistry*, **38**, 8304-8312.

Rattle, H. (1995) *An NMR primer for life scientists*. Partnership Press, Hants.

Rudino-Pinera, E., Ravelli, R.B., Sheldrick, G.M., Nanao, M.H., Korostelev, V.V., Werner, J.M., Schwarz-Linek, U., Potts, J.R. and Garman, E.F. (2007) The solution and crystal structures of a module pair from the Staphylococcus aureus-binding site of human fibronectin--a tale with a twist. *J Mol Biol*, **368**, 833-844.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M. and Cate, J.H. (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science*, **310**, 827-834.

Schwarz-Linek, U., Werner, J.M., Pickford, A.R., Gurusiddappa, S., Kim, J.H., Pilka, E.S., Briggs, J.A., Gough, T.S., Hook, M., Campbell, I.D. and Potts, J.R. (2003) Pathogenic bacteria attach to human fibronectin through a tandem beta-zipper. *Nature*, **423**, 177-181.

Sprangers, R., Velyvis, A. and Kay, L.E. (2007) Solution NMR of supramolecular complexes: providing new insights into function. *Nat Methods*, **4**, 697-703.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR*, **6**, 135-140.

Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley, New York.

# 4. The interaction between collagen and FN

## 4.1. Experimental design

The nature of the interaction between collagen and the collagen/gelatin binding domain of FN has been studied extensively for many years. Investigation of the binding properties would ideally involve native collagen, but this has proved very difficult because the triple helical molecule self-associates under physiological conditions to form insoluble fibrils (Kadler et al., 1996). Despite many years of experimentation, progress has been slow, and results have been very difficult to interpret. As an alternative, gelatin (denatured collagen) has been used in binding studies, and is used here. Binding studies using gelatin have far outnumbered those involving native collagen, although it is important to remember its limitations. Gelatin is much easier to work with, and interpretation of results is more straightforward, not least because of the tighter interaction between FN and gelatin (Engvall et al., 1981). The general goal is to increase our understanding of how FN and native collagen bind together in the extracellular matrix.

### 4.1.1. Binding properties of collagen and gelatin to Fibronectin

Since gelatin is denatured collagen, any differences in binding properties have to be due to the protein's quaternary structure. The preference of FN to bind to the denatured form of collagen indicates that the triple helical structure of intact collagen negatively affects the binding (Engvall et al., 1978). This negativity can be partly reversed by raising the temperature. While FN has been shown to bind rapidly and

quantitatively to gelatin at 4°C and 37°C, FN does not bind to collagen at 4°C, and will only bind slowly with preincubation at 37°C (Engvall et al., 1981). This suggests that the triple helical structure of collagen may unwind partially at higher temperatures and reveal the FN binding site. This relaxation has been shown to occur near physiological temperatures by ultra-slow scanning calorimetry and isothermal circular dichroism (Leikina et al., 2002) suggesting the possibility of local micro-unfolding of the least stable domains.

### 4.1.2. Nature of the binding interaction

The nature of the interaction between FN and collagen is still very unclear due to the number of possible binding sites on collagen. Initially a single binding site for FN was proposed on type I collagen (α1(I)-CB7) (Kleinman et al., 1978). A subsequent study suggested there were at least two binding sites (Guidry et al., 1990), and most recently fourteen possible binding sites were isolated by titrating fluorescently-labelled fragments of type I collagen against the 42-kDa fragment of FN (Ingham et al., 2002). These studies have been unable to confirm whether the binding site for FN is present on a single collagen polypeptide chain or lies across the strands of the triple helix.

Despite the molecular detail of the interaction being unclear, the general mode of binding is better understood, and has been investigated by disrupting the interaction with various agents. The interaction has been proposed to be predominately electrostatic as binding of FN to gelatin-agarose is strongly inhibited by increasing NaCl concentration (Vuento et al., 1982). Increasing the salt concentration to

physiological levels causes a three-fold increase in the dissociation constant between gelatin and the 42kDa fragment from FN (Forastieri and Ingham, 1985). Therefore low ionic strength fosters non-specific electrostatic interactions between these binding partners. On the other hand, the $K_d$ of the interaction of the denatured form of collagen and FN is largely unaffected when raising the NaCl further from physiological concentrations to 1M NaCl, suggesting that the primary interaction with gelatin is not electrostatic in nature (Forastieri and Ingham, 1985).

An alternative hypothesis has been put forward to suggest that the interaction between gelatin and FN is dominated by a hydrophobic contribution, as it has been shown that binding may be disrupted with a variety of chaotropic agents including urea (Engvall and Ruoslahti, 1977) and anionic detergents (Vuento et al., 1982). Again, Forastieri and Ingham argue that this is not the case, and that in their fluorescence polarization assay dissociation is accompanied by an increase in entropy (not expected for a hydrophobic interaction). They suggest this loss in binding affinity is due to a general unfolding of the proteins rather than disruption of hydrophobic interactions (Forastieri and Ingham, 1985). FN can be eluted from gelatin-agarose by simply lowering the pH below 6.0 in the absence of chaotropes (Miekka et al., 1982), (Forastieri and Ingham, 1985). Although it is unclear which of the electrostatic, hydrophobic, or hydrogen bonds are the major contributor, it is reasonable to expect that all play some part in this interaction.

Determination of the nature of FN binding to collagen is only really of interest if the interaction is biologically significant. Although investigation *in vivo* has lagged

behind *in vitro* studies, there is a growing body of evidence to support an interaction within the ECM. A recent study suggested that FN-collagen interactions are critical for the ability of FN to enhance cell processes such as contractility (Sottile et al., 2007). They showed that FN was able to enhance cell migration when added to FN null myofibroblast cells, but by deleting the GBD from FN, the protein was defective in its ability to promote cell migration. In addition, the naturally occurring bacterial FN-binding protein SFS from parasitic bacterium *Streptococcus equi* subsp. *equi*, which had previously been shown to inhibit binding between FN and collagen in a concentration dependent manner (Lindmark and Guss, 1999), was able to block FN-induced cell migration in this assay (Sottile et al., 2007). These results support the biological relevance of the FN-collagen interaction and indicate that the interaction of the two *in vitro* is not simply fortuitous, but that FN-directed collagen deposition is one mechanism by which FN stimulates cell migration.

### 4.1.3. Gelatin affinity chromatography

Gelatin-sepharose chromatography is a suitable tool to investigate the gelatin binding properties of FN. Isolation of the GBD from native and recombinant sources has allowed the binding properties of the six modules in the domain to be investigated (Katagiri et al., 2003). FN fragments bind as they are passed down a gelatin column and are eluted with an increasing concentration of urea. Describing those fragments eluting last from the column as having the highest affinity is not strictly correct, but rather, the interaction between the two binding partners shows the most resistance to the disruption by the chaotrope (Forastieri and Ingham, 1985). It is possible that release from the column may be due to a single cooperative unfolding transition of the FN fragment in the presence of the denaturant (Isaacs et al., 1989) rather than a

disruption at the binding interface. Nevertheless, this technique can be used to compare the ability of different FN fragments to bind to gelatin.

### 4.1.4. Surface plasmon resonance (SPR)

Surface Plasmon Resonance (SPR) is a powerful technique for probing the interaction between biological molecules and their binding partners without the need for labelling (Van Der Merwe, 2001). Binding of a ligand to its receptor is detected by observing the reflected light from a gold coated slide and the refractive index near a sensor surface can be used to quantify kinetic constants, equilibrium constants, and evaluate the binding properties of recombinant protein against its native counterpart.

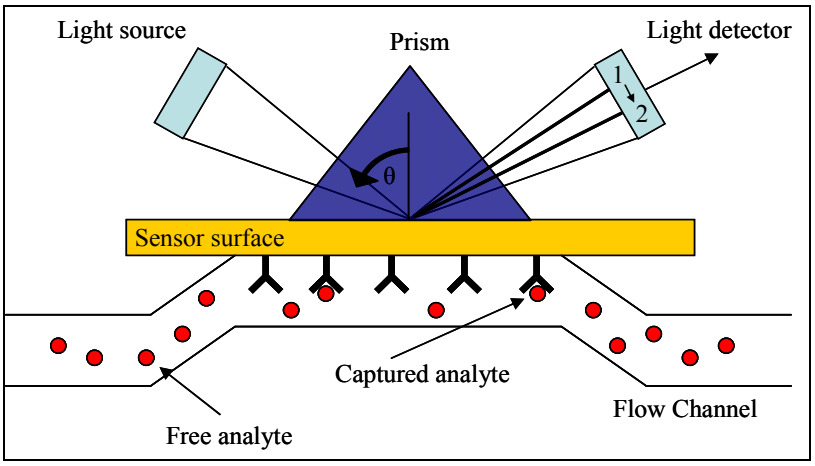



**Figure 4:1** A schematic representation of the SPR instrument. A convergent light beam is focused onto the underside of the silver/gold surface and is reflected into a detector array. Analyte is flowed over the chip surface at a defined flow rate and set temperature. The two dark lines in the reflected beam indicate the light intensity drop following the resonance phenomenon at time t1 (before binding) and t2 (after binding), (diagram modified from www.chbr.noaa.gov).

In order to detect an interaction, one binding partner (ligand) is covalently linked to a chip surface, while the other binding partner (analyte) is run continuously over the exposed surface. As the analyte binds to the ligand, the amount of bound material starts to accumulate, and the refractive index increases (Figure 4:1). The refractive

index is plotted in real time against the change in absorbance units (RU). These values are subtracted from the background response (a control cell without ligand bound) before analysing the sensorgram further (Figure 4:2).

### 4.1.5. Analysis of surface plasmon resonance data

SPR data can be used to calculate both kinetic and equilibrium constants through analysis of the association and dissociation results. In the first instance, the kinetics of a reaction are assumed to follow pseudo first order kinetics. The immobilised ligand (L) and the analyte (A) bind together. During the association phase the complex [LA] increases as a function of time:

$$d[LA]/dt = k_a.[L].[A] - K_d.[LA]$$

where $k_a$ is the association rate constant which describes the rate of complex formation, and $K_d$ is the dissociation rate constant. Fitting a simple 1:1 Langmuir binding model to the data can be used to calculate the $K_d$ ($K_d = k_{off}/k_{on}$).

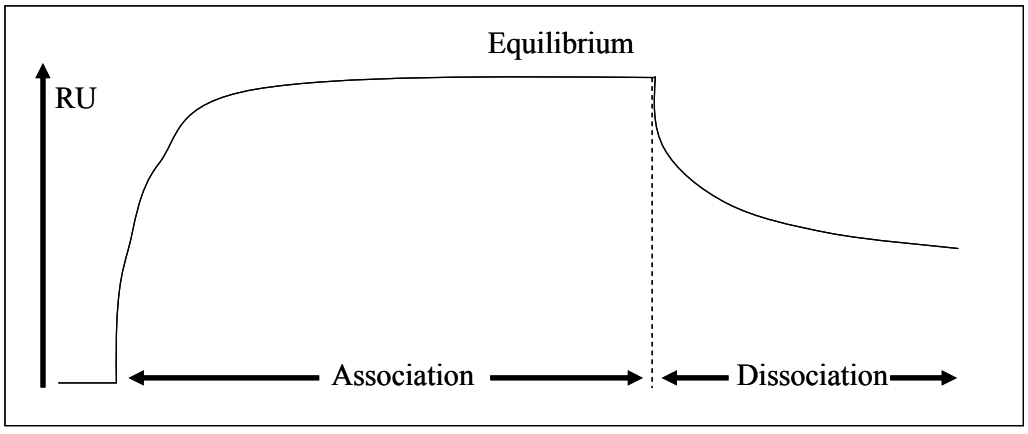



**Figure 4:2** A typical sensorgram showing the association phase and dissociation phase of the experiment. The analyte is passed through the flow channel, interacts with the bound ligand, and builds up on the chip surface. Buffer flow continues in the absence of analyte and the analyte is washed from the chip surface in the dissociation phase.

Data collected over a longer timescale can be used to calculate an equilibrium constant for binding as the analyte and ligand reach a state of equilibrium. The analyte is flowed over the chip surface at different concentrations and in each experiment the RU at equilibrium is used to calculate an equilibrium constant by means of a Scatchard plot.

$$\text{Bound} = C^A * Max / C^A + K_d$$

where Bound is measured in RU, '*Max*' is the maximum response (RU), $C^A$ is the concentration of injected analyte, and $K_d$ is the dissociation constant (Van Der Merwe, 2001). The Scatchard plot is obtained by plotting Bound/$C^A$ against Bound and the (negative) gradient will give an estimate of the $K_d$.

### 4.1.6. NMR titration

NMR is a powerful technique for detecting protein-protein interactions and mapping out binding interfaces at the atomic level. Perturbation of single residues can be mapped and used to calculate binding constants. The observed effects in NMR depend on the rates of exchange between different environments; either protein-protein interactions on different molecules, or different conformations of the same molecule. In a binding event, a nucleus exchanges between two magnetically distinct sites. If the off rate is much faster than the difference between the resonance frequencies, the exchange rate is 'fast' and a single averaged resonance is observed. If the nucleus exchanges between two magnetically distinct sites at a slower rate than the difference between the resonance frequencies, then the exchange rate is 'slow' and two separate resonances are observed for the 'bound' and 'free' states. Intermediate exchange will result in a broadening of resonances (Figure 4:3).

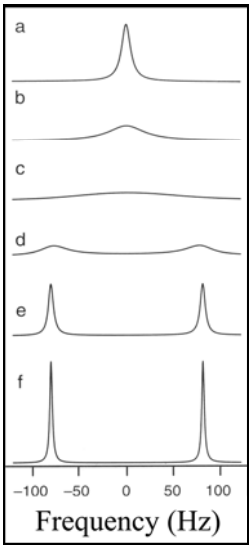



**Figure 4:3** NMR chemical exchange for a two site system. A, fast exchange, B-E, intermediate exchange, F, slow exchange (adapted from (Cavanagh, 2007), p400).

### 4.1.7. Binding analysis through synthetic peptide NMR titration

Ligand/protein binding modes can be determined by monitoring NMR resonances as the concentration of one of the binding partners is increased. Initial work to characterise the binding interaction between short synthetic collagen peptides and the first three modules of the gelatin binding domain suggested a $K_d$ of 6mM (Pagett et al., 2005). Subsequent experiments could progress to using triple-helical peptides, and eventually full collagen molecules, although obtaining these purified fragments and then interpreting the data would be a challenging project. In order to gain more of an insight, single stranded synthetic collagen peptides are sufficient for identification of further binding interfaces on the rest of the gelatin binding domain, and can provide initial data for modelling.

## 4.2. Methods

### 4.2.1. Surface plasmon resonance

Surface Plasmon Resonance (SPR) experiments were performed on a BIAcore 2000 instrument (BIAcore AB). Initially the ligands were prepared; purified type I collagen from human placenta (Sigma) was suspended in 10 mM acetate buffer pH 4.0 at 10 μg/ml, 50 μg/ml, and 100 μg/ml. Purified type I collagen (Sigma) was denatured by boiling at 95°C for 1 hour (gelatin), and suspended in 10 mM acetate buffer pH 4.0 at 10 μg/ml, 50 μg/ml, and 100 μg/ml. Collagen and gelatin (at these three concentrations) were coupled to different lanes of two separate CM5 sensorchips by covalent coupling using the amine coupling method as described in the BIAapplications handbook (BIAcore AB). The fourth cell in each sensorchip was derivatised for amine coupling in the absence of protein and used as a control cell. The 50 μg/ml collagen α1(1) chain cell (response level of ~5500 resonance units (RU)) and the 50 μg/ml gelatin cell (~8000 RU) were judged to give the best data for the equilibrium analysis of FN fragment binding.

Binding experiments were carried out in HBS-EP running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) Surfactant P20) at a flow rate of 1.0 μg/min. To compare the binding of the fragments, 400μl of five dilutions (0.5-8.0 μM) of the recombinant Hex protein and native proteolytic GBD fragment (Sigma) were run at 1.0 μg/min. The amount of fragment bound to the sensorchip was monitored by change in RU. The sensorchip data were analysed using global fitting protocols in the BIAevaluation 3.0 programme, subtracting any background binding from the control cell, and a $K_d$ for binding was estimated from a Scatchard plot.

### 4.2.2. Analytical gelatin affinity chromatography

Protein fragments were purified to at least 95% homogeneity by HPLC before loading onto a gelatin affinity column. The chitobiose core of high mannose sugars had been previously cleaved with either Endo Hf (NEB) or the whole sugar chain removed with PNGase F (NEB). The gelatin-sepharose 4B (Amersham Pharmacia Biotech) was initially packed to 3 cm depth in a XK 26/20 column (internal diameter 26mm), and for the second set of experiments, 6 cm depth in a XK16/20 column (internal diameter 16 mm). The columns were equilibrated with PBS pH 7.4 before loading the sample at 1ml/min. The column was washed with one volume of PBS pH 7.4 and the bound protein eluted with a gradient of 0-100% 4 M Urea (PBS pH 7.4) over 5 column volumes (CV) at 1.0 ml/min. The $A_{280}$ profile was used to characterise the protein elution.

### 4.2.3. NMR collagen peptide titration

Four different synthetic collagen peptides were obtained commercially for titration with $^{8}$Fn1*-$^{9}$Fn1 (R503K) (Table 2). Peptide 1 and peptide 2 (EZBiolab, Indiana, USA) sequences are taken N-terminal and C-terminal to the collagenase cut site (PQG--IA) of collagen α1(I) chain (Figure 4:4). Peptide 3 and peptide 4 (GLBiochem, Shanghai) sequences are taken from the N-telopeptide and the 260 position in the collagen α1(I) chain respectively (UniProtKB/Swiss-Prot: P02452, CO1A1_HUMAN). A single Tyr residue was introduced at the C-terminus of each peptide to allow for concentration measurements at $A_{280}$ on a UV spectrometer.

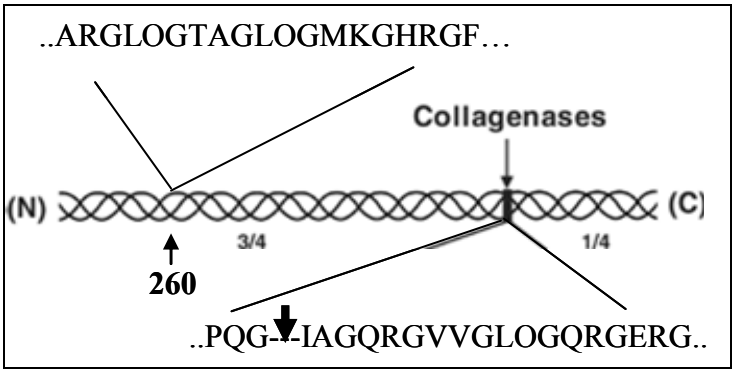



**Figure 4:4** Schematic diagram of type I collagen showing the approximate position of the collagenase cut site. Collagenases MMP-1 and MMP-3 cleave at a single residue in the collagen chain (large arrow). The 260 site is marked. The N-telopeptide region precedes the N-terminus shown (O = Hyp) (adapted from (Chung et al., 2004)).

The [$^{1}$H-$^{15}$N]-HSQC titrations were performed at 25°C on the 950 MHz spectrometer with 64 scans, 1K data points, and 360 increments to achieve good resolution in both proton and nitrogen dimensions. The protein concentration was fixed at 0.1 mM and peptide concentration was adjusted from 4:1 to 0.5:1 by removing a fraction of the sample and replacing that fraction with protein only. The intensity of the non-complexed (free) crosspeak was measured as the peptide/protein ratio was adjusted after each titration step.

| Peptide | Sequence | Position in collagen sequence |
|---|---|---|
| 1 | GPAGAOGTOGPQGIAG-(Y) | 941 - 956 |
| 2 | GQRGVVGLOGQRGERGFOGLOG-(Y) | 956 - 977 |
| 3 | GLGGNFAPQPQLSYGYDEKSTGGISVPGPM-(Y) | 154 - 181 |
| 4 | ARGLOGTAGLOGMKGHRGFSGLDGAK-(Y) | 252 - 277 |

**Table 2** Synthetic collagen peptides. Each was titrated separately with $^{8}$Fn1*-$^{9}$Fn1 (R503K) and the effect measured in the [$^{1}$H-$^{15}$N]-HSQC spectrum (O = Hyp, (Y) = Tyr residue added to allow concentration measurements).

### 4.2.4. NMR chemical shift mapping

Four residue crosspeaks showing significant shifts in the [$^{1}$H-$^{15}$N]-HSQC spectrum after peptide addition were selected. All the observed perturbations were in the 'slow' NMR exchange regime. The intensity of the "free" crosspeak (before peptide addition) at different peptide ratios was recorded at each titration point (relative to 0mM peptide) and these relative intensities fitted (using OriginPro 7.5) to the equation:

$$y = 1 - d_{max}*((P_{tot}+x+K_d)-sqrt((P_{tot}+x+K_d)^2-4*P_{tot}*x))/(2*P_{tot})$$

where the following initial values were used: $d_{max}$ = 1 (maximum relative crosspeak intensity), $P_{tot}$ = 0.1 mM (protein concentration), [with $P_{tot}$ and $K_d$ shared], $K_d$ = 0.5 mM (affinity constant), x independent (peptide concentration) and y dependent (relative crosspeak intensity). To calculate the $K_d$, the data were iterated with $d_{max}$ and $K_d$ allowed to vary until convergence (no reduction in the Chi-sqr value).

Chemical shift changes in the complexed (bound) spectrum upon binding, which could be identified by proximity (at saturation), were quantified using the weighted chemical shift change per residue, Δδ [HN,N]:

$$\Delta\delta\ [HN,N] = ([\Delta\delta_{HN}W_{HN}]^2 + [\Delta\delta_{N}W_{N}]^2)^{1/2}$$

where $W_{HN}$ and $W_N$ are weighting factors for the HN and N shifts, respectively ($W_{HN}$=1;$W_N$ = 0.154; (Ayed et al., 2001)), and $\Delta\delta = \delta_{bound} - \delta_{free}$. The crosspeak shifts were plotted onto the $^{8}$Fn1*-$^{9}$Fn1 schematic diagram (drawn by Dr. A.R. Pickford based on a typical Fn1 module pair) using the same colour scheme as used in the weighted shift map. Those peaks that could not be identified by proximity and those peaks that disappeared in the "bound" spectrum were indicated in the results.

## 4.3. Results

### 4.3.1. Characterisation of Hex by surface plasmon resonance

SPR was used to compare the binding properties of recombinantly expressed Hex against proteolytically derived gelatin binding domain (GBD) to ensure that the Hex was behaving in the same way as native protein (the dispersed crosspeaks of the [$^{1}$H-$^{15}$N]-HSQC spectra had already shown that the Hex was folded). Initially the GBD and Hex (analytes) were assessed for their ability to bind to gelatin (ligand) at 37°C. The system was allowed to equilibrate at several different analyte concentrations and the results analysed by the means of a Scatchard plot.

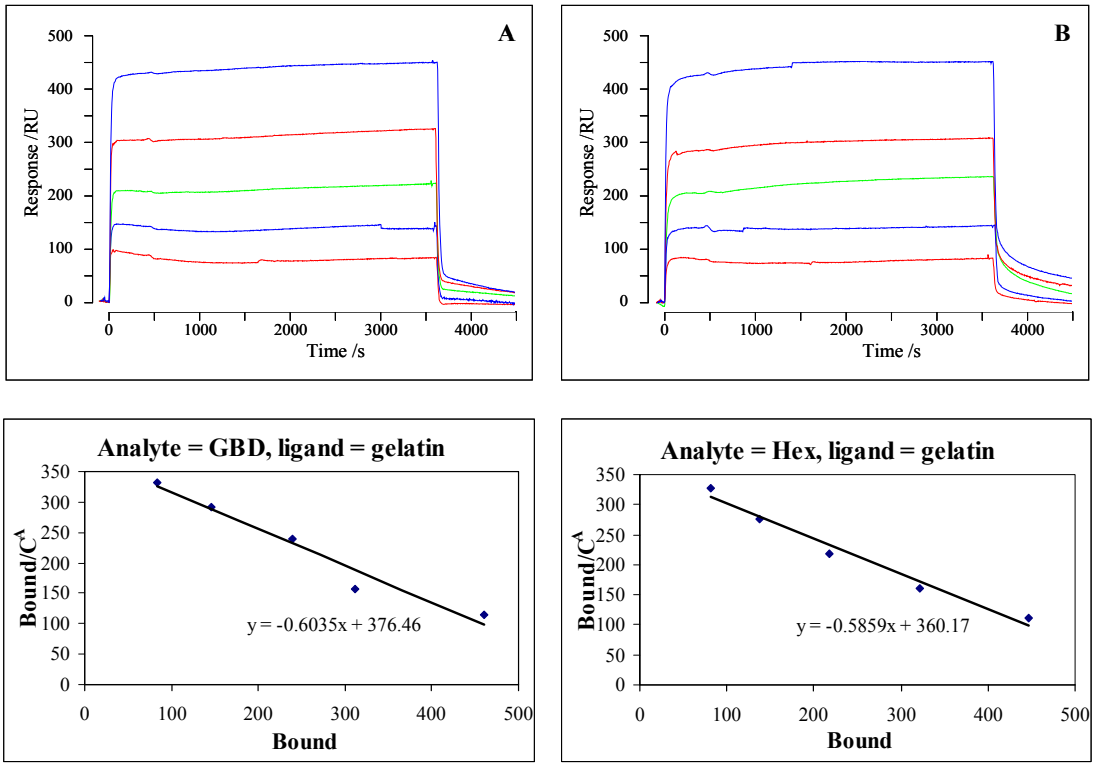



**Figure 4:5** SPR analysis of binding to gelatin. Binding of A, GBD (analyte) and B, Hex (analyte) to gelatin (ligand) which was covalently coupled to the chip. The Scatchard plot of each dataset is set below each sensorgram ($C^{A}$ = concentration of injected analyte).

Both GBD and Hex gave very similar sensorgram profiles confirming that both native and recombinant material behaved in a similar way (Figure 4:5). The $K_d$s for gelatin binding were 0.604 ± 0.056 μM and 0.586 ± 0.073 μM respectively (n = 5, error [rmsd] calculated in OriginPro7.5) suggesting a strong affinity between the ligand and analyte. From the sensorgram, the $k_{on}$ and $k_{off}$ rates were judged to be high, implying a transient interaction as the analyte concentration quickly built up, and was readily washed away from the surface of the chip.

The ability of the FN fragment to bind to collagen was assessed by laying down collagen (ligand) on the chip and leaving the system to equilibrate at 37°C. Due to low levels of Hex expression, the experiment was conducted using GBD, as the previous result suggested that the response from the Hex would be very similar.

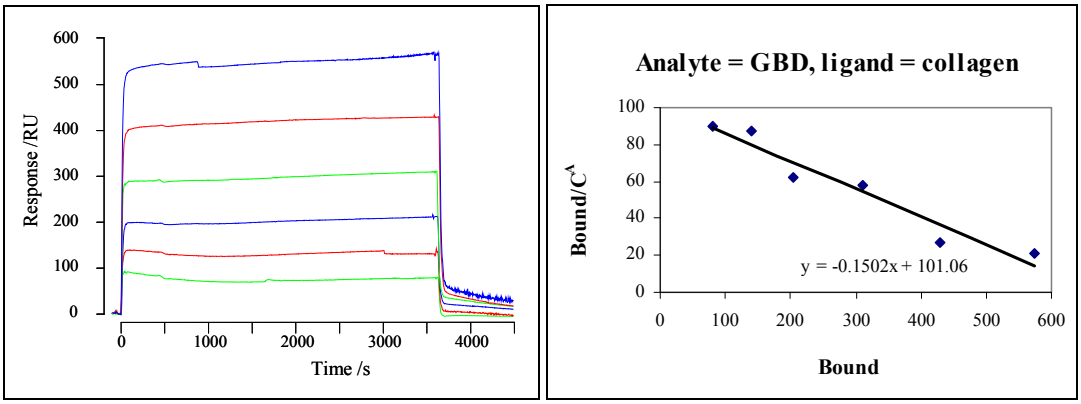



**Figure 4:6** SPR analysis of binding to collagen. Sensorgram of GBD (analyte) response to covalently coupled collagen (ligand). The Scatchard plot is alongside with a number of intermediate concentrations not shown in the sensorgram ($C^A$ = concentration of injected analyte).

The sensorgram profile was very similar to that of gelatin, displaying the same fast-on, fast-off characteristics, indicating that both ligands are similarly presented to the analyte (Figure 4:6). The $K_d$ for binding to collagen was 0.150 ± 0.049 μM which falls in the same order of magnitude as gelatin binding. The difference in $R_{max}$ (x-

intercept of the Scatchard plot) between gelatin and collagen can be explained by the ligand immobilisation levels. The $R_{max}$ for the GBD and Hex with gelatin (immobilisation level ~8000) were 624 and 615RU respectively, and the $R_{max}$ for the GBD with collagen (immobilisation level ~5500) was 467RU. The ratio of immobilisation of gelatin to collagen was ~1.45, implying that gelatin was more amenable to cross linking to the chip's surface at pH 4.0, and therefore more of the ligand present for the FN fragments to bind. This was reflected in the ratio of $R_{max}$ values for analyte binding to gelatin relative to collagen (~1.35). Therefore, collagen and gelatin both appear to display one (if not more) of the FN binding sites, and this site is equally accessible to ligand at 37°C.

### 4.3.2. Studying the property of glycosylation on fibronectin fragments

The role of the two N-linked sugars on the Hex (attached to N399 on $^{2}$Fn2 and N511 on $^{8}$Fn1) were investigated by removing high mannose sugars with the enzymes Endo $H_f$ and PNGase F. Although it was not possible to selectively remove one or other sugar from the Hex, smaller constructs of the Hex were expressed, and the properties of the remaining sugar chain investigated. The Triple ($^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2) and the Quad ($^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1) (which both only contain a sugar on N399) and $^{8}$Fn1**-$^{9}$Fn1 (which only contains a sugar on N511) were used for this purpose. The Hex, before and after trimming with Endo $H_f$ (to cleave the sugar chains to single GlcNAc sugar residues), was bound to a gelatin column and eluted with a urea gradient (Figure 4:7).

Both elution profiles were shown to be composed of several peaks with the profile of the protein after treatment having an extra high affinity peak. The first two peaks of

the protein after treatment were previously shown to be Hex fragments (although some gelatin binding character remained) and only the third corresponded to correctly folded undigested protein (chapter 2). The elution profile before treatment was missing the highest affinity peak and this only appeared after Endo $H_f$ treatment. Correctly folded undigested protein had to be present in both protein samples, but the absence of the third peak in the non-treated profile suggested that this species must coelute with weaker binding protein at an earlier point, and therefore the removal of the sugar chain increased the binding affinity.

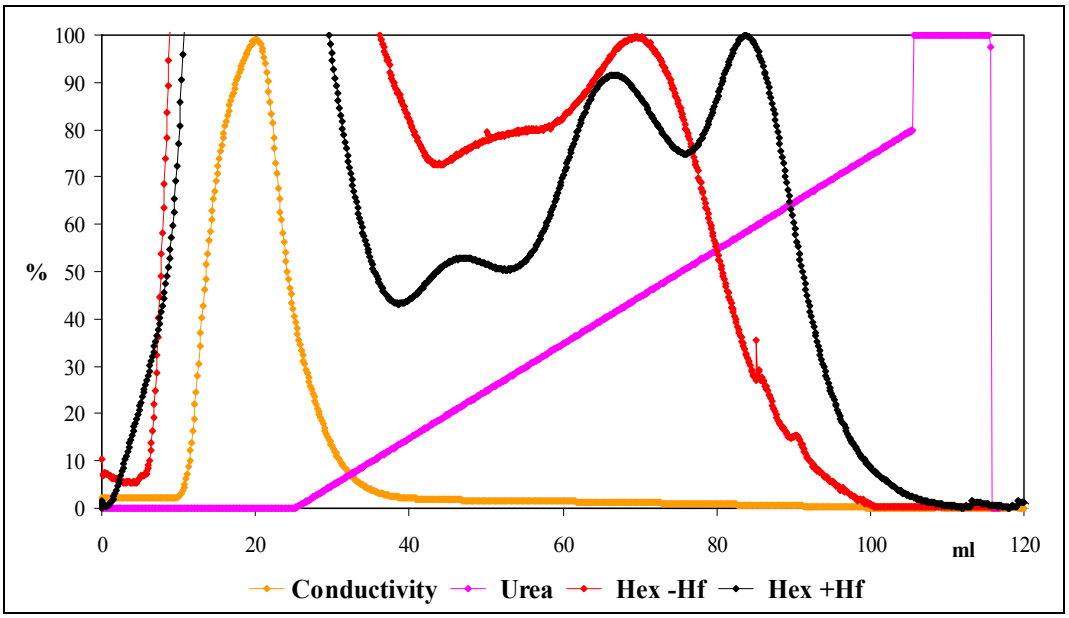



**Figure 4:7** The $A_{280}$ profile of the Hex eluting from the gelatin column in a gradient of 0-4 M urea before (Hex -Hf) and after (Hex +Hf) treatment with Endo $H_f$. The profile is normalised to the largest binding peak, and the void volume is indicated by the rise in conductivity.

In order to assess the individual contribution of the sugars to binding, smaller recombinantly expressed fragments of the Hex were investigated before and after treatment with sugar cleavage enzymes. The $^{8}$Fn1**-$^{9}$Fn1 module pair (R503K, N497Q; single sugar chain on N511) was treated with Endo $H_f$ (to leave a single GlcNAc) and PNGase F (to remove the entire high mannose sugar chain) and the

binding affinity was assessed (Figure 4:8). The Triple ($^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2) and the Quad ($^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1) (both contain a sugar chain on N399) were treated with Endo $H_f$ to assess the contribution of the sugar on binding (Figure 4:9 and Figure 4:10).

The $^{8}$Fn1**-$^{9}$Fn1 module pair with full length sugar chain (89 -Hf) and $^{8}$Fn1**-$^{9}$Fn1 with a single GlcNAc (89 +Hf) bound to the gelatin column with a similar affinity. This showed that binding is not inhibited by the presence of the full length sugar chain on N511. In contrast, the removal of the whole sugar chain caused loss of all the gelatin binding affinity, and the majority of the protein (89 +PNG) eluted with the salt peak. A small minority eluted even earlier at the void volume (at the point of BSA elution), suggesting that the digested protein formed a large aggregate which was unable to enter the sepharose beads, and was simply washed through the column. The major peak elutes before the start of the 0-4 M urea.

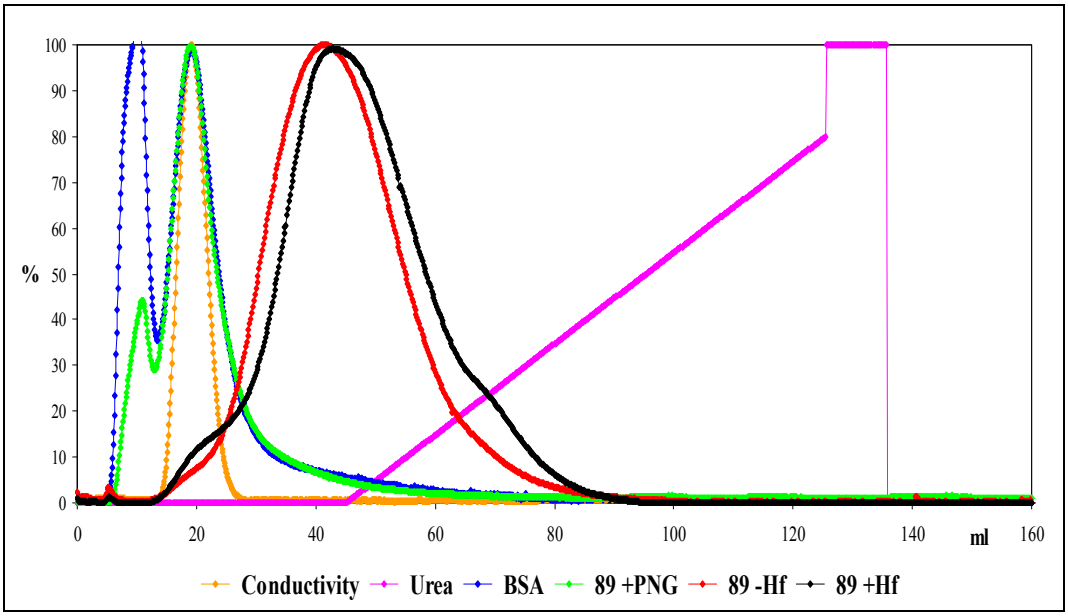



**Figure 4:8** The elution profile of $^{8}$Fn1**-$^{9}$Fn1 from the gelatin column in a 0-4 M urea gradient before (89 -Hf) and after treatment with Endo $H_f$ (89 +Hf) and PNGase F (89 +PNG). The profile is normalised to the largest peak. The void volume of the column is indicated by peak BSA and the peak in conductivity indicates the elution of salts from the column.

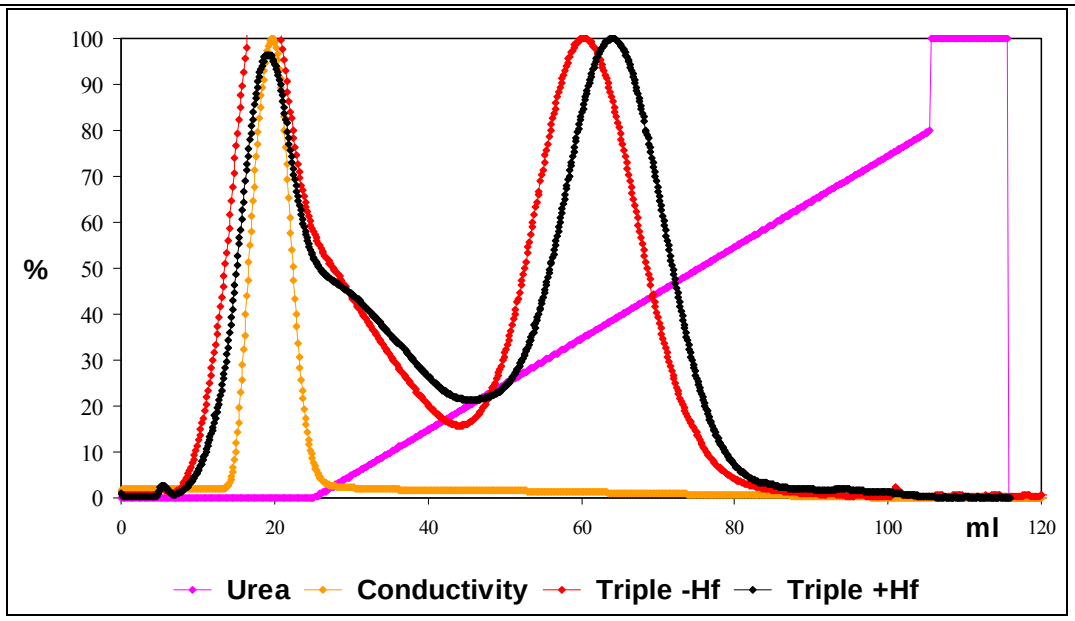



**Figure 4:9** The elution profile of the Triple ($^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2) before (Triple -Hf) and after (Triple +Hf) treatment with Endo $H_f$. The protein is eluted from the column with a 0-4 M urea gradient and the profile is normalised to the major protein peak. The salt peak is shown.

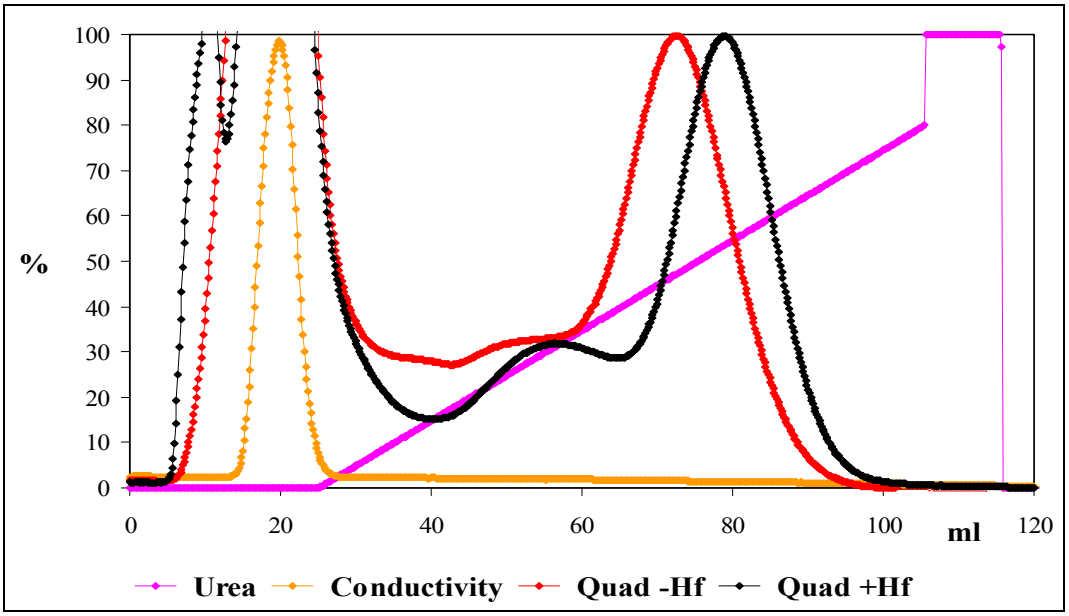



**Figure 4:10** The elution profile of the Quad ($^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1) before (Quad -Hf) and after (Quad +Hf) treatment with Endo $H_f$.

The effect of removing the sugar chain (linked to N399) from $^{2}$Fn2 was more distinct in the Quad than in the Triple, but far less so than removing the sugar chain (linked to

N511) from $^{8}$Fn1. The absence of the full length sugar chain in both the Triple (Figure 4:9) and the Quad (Figure 4:10) appeared to slightly increase the binding affinity to gelatin. The main binding peak eluted later in both cases after treatment with Endo $H_f$ with a more pronounced effect in the larger fragment. The smaller and weaker binding peak in both profiles was presumably misfolded or partially degraded material, which still retains some binding activity, but was eluted from the column at an earlier point than the tight binding peak.

As well as studying the effect of the high mannose sugar chain on binding, the increase in binding affinity of the Hex fragments was analysed, and in general, the larger the fragment, the tighter the binding to the gelatin column. The $^{8}$Fn1**-$^{9}$Fn1 module pair was retarded by the column, but eluted isocratically, whereas the longer fragments required urea to disrupt the binding interaction (Figure 4:11).

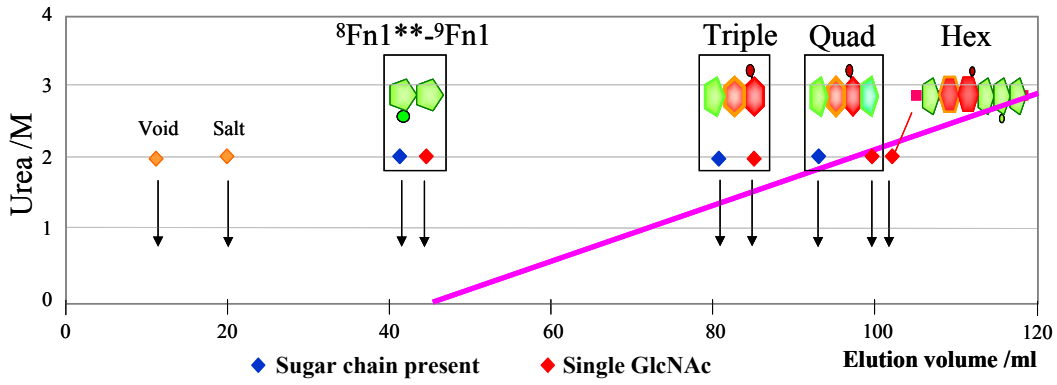



**Figure 4:11** The major peak elution volumes of four of the gelatin binding fragments. The fragments were applied to the column, washed with 45 ml column buffer, and a urea gradient was applied (purple line). The peak elution volumes of the fragments from the gelatin column were recorded in triplicate before and after Endo $H_f$ treatment. The Hex with sugar chain present is not included in this analysis as it was not possible to isolate the binding peak from the proteolytically degraded material. The void volume and salt peak are indicated.

### 4.3.3. Further fibronectin fragment binding properties

Further FN fragments were expressed and their binding properties characterised by gelatin affinity chromatography. The resolution of the experiment was increased by decreasing the diameter of the column and packing the gelatin sepharose to twice its original length, to allow for a better binding comparison. These fragments were shown to elute isocratically, avoiding the use of urea, and allowing the proteins to elute by dissociation rather than unfolding.

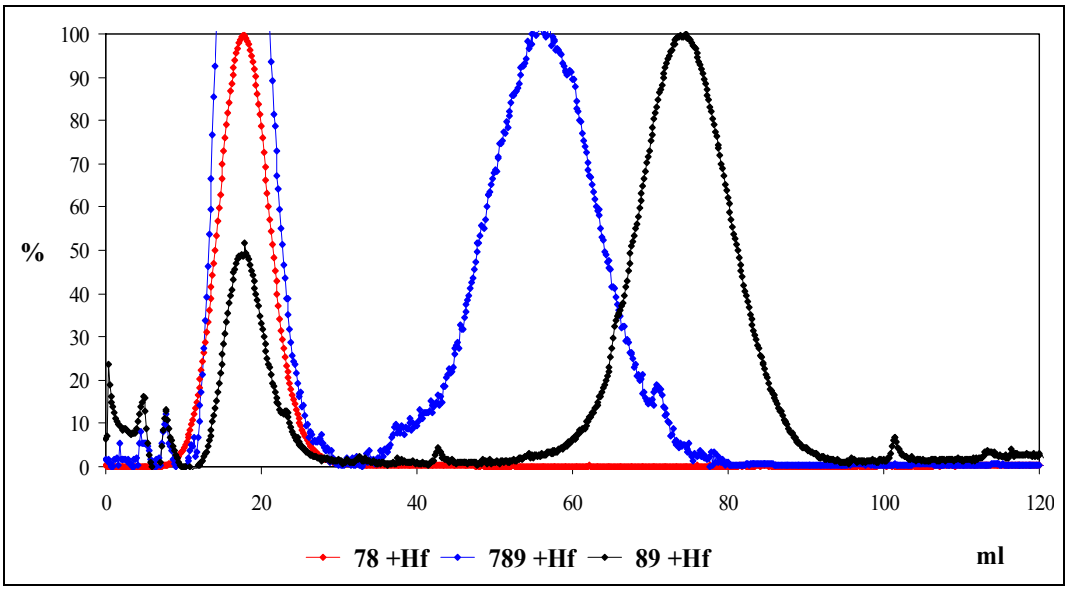



**Figure 4:12** Isocratic elution of FN fragments from a reduced diameter, double length gelatin column (compared to the original). All constructs were treated with Endo $H_f$ to trim the high mannose sugar chain before being passed down the column.

In contrast to the previous conclusion with FN fragments from the N-terminus of the Hex, the longer $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 from the C-terminus did not bind more tightly to gelatin than $^{8}$Fn1**-$^{9}$Fn1. The addition of $^{7}$Fn1 to $^{8}$Fn1**-$^{9}$Fn1 ($^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1) resulted in a reduction of the binding affinity. Also unexpectedly, the removal of $^{9}$Fn1 from the $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 fragment ($^{7}$Fn1-$^{8}$Fn1**) removed all binding affinity

for gelatin (Figure 4:12). The $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 had both been shown to be folded by one-dimensional NMR (chapter 3).

The influence of the C-terminal $His_6$ tag on binding was investigated as previous results suggested that the $^{7}$Fn1-$^{8}$Fn1** FN fragment should bind to gelatin (Katagiri et al., 2003). The $^{8}$Fn1**-$^{9}$Fn1 module pair, expressed with and without $His_6$ tag, was passed down the gelatin column, and the tagged protein eluted nearly 10ml earlier than the untagged protein. Although this shift in peak elution was significant, the tagged $^{8}$Fn1**-$^{9}$Fn1 remained bound to the column for nearly 10ml longer than $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1-$His_6$ (Table 3). The complete loss of binding by $^{7}$Fn1-$^{8}$Fn1** is unlikely to be caused by the $His_6$ tag and is more likely to be caused by the loss of $^{9}$Fn1 module from the end of the construct. This implies that the $^{9}$Fn1 module is required at the C-terminus to allow smaller FN fragments to bind.

| | $^{7}$Fn1-$^{8}$Fn1** (+ $His_6$) | $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 (+ $His_6$) | **$^{8}$Fn1**-$^{9}$Fn1 (- $His_6$)** | $^{8}$Fn1**-$^{9}$Fn1 (+ $His_6$) |
|---|---|---|---|---|
| Elution peak /ml | 18.1 | 56.8 | **75.8** | 66.3 |

**Table 3** The peak elution volumes for C-terminal Hex FN fragments. All constructs have a C-terminal $His_6$ tag with the exception from the $^{8}$Fn1**-$^{9}$Fn1 (- $His_6$) (Bold).

### 4.3.4. NMR collagen peptide titration

The nature of $^{8}$Fn1*-$^{9}$Fn1 (R503K) binding to collagen was investigated by NMR titration with four different synthetic peptides taken from the collagen sequence. There was no change to the [$^{1}$H-$^{15}$N]-HSQC spectra of $^{8}$Fn1*-$^{9}$Fn1 in response to peptide 1 (N-terminal to the collagenase cut site) or peptide 3 (hydrophobic N-telopeptide). There were some small changes in crosspeak positions in response to

peptide 4 (possible FN binding region) but the most dramatic changes were with peptide 2 (immediately C-terminal to the collagenase cut site) (Figure 4:13). (Peptide 2 will be referred to as "peptide" from this point unless otherwise stated in the text).

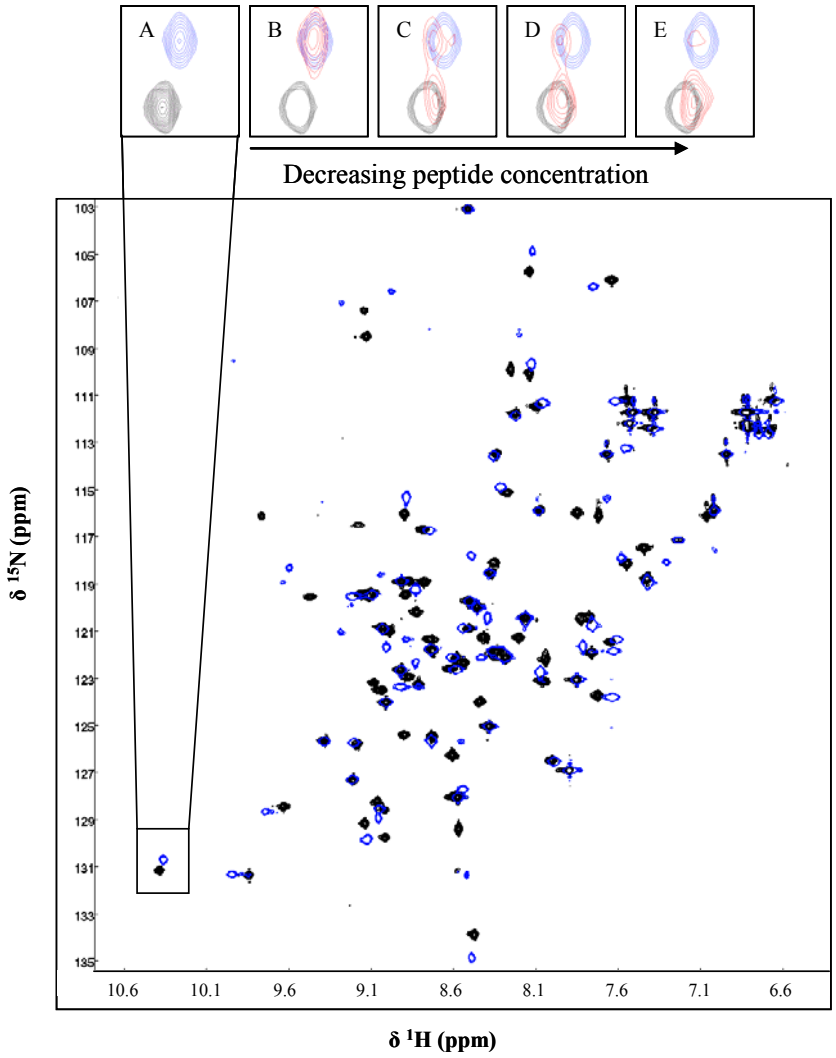



**Figure 4:13** [$^{1}$H-$^{15}$N]-HSQC spectra of $^{8}$Fn1*-$^{9}$Fn1 before (black filled crosspeaks) and after (blue open crosspeaks) NMR titration with 0.4mM peptide 2 at 25°C. Above: peak movement as the concentration of peptide 2 was reduced from 0.4mM to 0.05mM. A, before (black) and after 0.4mM peptide 2 (blue). B, 0.3mM peptide 2 (red). C, 0.2mM peptide 2 (red). D, 0.1mM peptide 2(red). E, 0.05mM peptide 2 (red).

The four peptides were titrated with $^{4}$Fn1-$^{5}$Fn1 and the Quad as controls for the binding (conducted by Dr. I. Vakonakis). As expected, there was no binding to $^{4}$Fn1-$^{5}$Fn1 which eliminated the probability of any fortuitous interaction with Fn1 modules.

Peptides 2 and 4 both bound to the Quad, although with far smaller crosspeak movement, which implied that the Quad and $^{8}$Fn1*-$^{9}$Fn1 can share the same binding sites on collagen.

Due to the slow exchange regime, the crosspeak shifts in the peptide titrated [$^{1}$H-$^{15}$N]-HSQC spectrum could not be mapped in the same way as for a fast exchange regime by following titration points back to the original crosspeaks; time restraints precluded sequential assignment of the bound state. The crosspeaks in the peptide titrated [$^{1}$H-$^{15}$N]-HSQC were mapped to the closest unambiguous peak in the original [$^{1}$H-$^{15}$N]-HSQC.

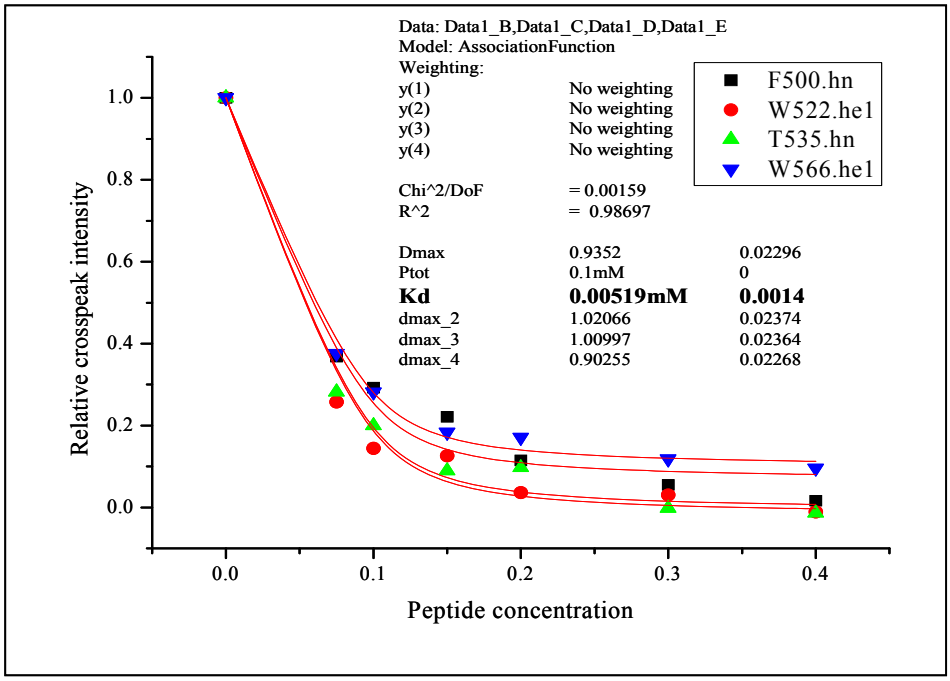



**Figure 4:14** Relative "free" crosspeak intensity plotted against peptide concentration. As the concentration of peptide 2 was decreased, the relative crosspeak intensity at the 0 mM peptide crosspeak position was recorded. The fitted curves (red lines) and iteration results are shown.

In order to calculate the affinity constant, discrete crosspeaks at 0 mM and 0.4 mM peptide concentration were required, ideally with crosspeak movement in both $^{1}$H and

$^{15}$N dimensions (in order to capture effects which may otherwise be neglected if only a single dimension was used). The bound protein appeared to be in slow exchange as the bound and unbound crosspeaks did not average out but remained as discrete peaks as the concentration of peptide was decreased. Two crosspeaks from both modules were selected (G507 and W522ε, T535 and W566ε), and as the peptide concentration was lowered, the "returning" crosspeak intensity was recorded at each titration point. The data were fitted to an exponential function and a $K_d$ value of 5.18 ± 1.4 μM was calculated for the binding (Figure 4:14).

### 4.3.5. Collagen peptide interaction site on $^{8}$Fn1*-$^{9}$Fn1

The $^{8}$Fn1*-$^{9}$Fn1 [$^{1}$H-$^{15}$N]-HSQC crosspeaks before and after the addition of 0.4 mM peptide were compared and the majority of those crosspeaks in the peptide bound spectrum could be assigned. The crosspeak shifts were measured and weighted (Figure 4:15). The largest shifts were seen in the 'D' and 'E' strands of $^{8}$Fn1, in the 'A' strand of $^{9}$Fn1, and in the linking residues (Figure 4:16).

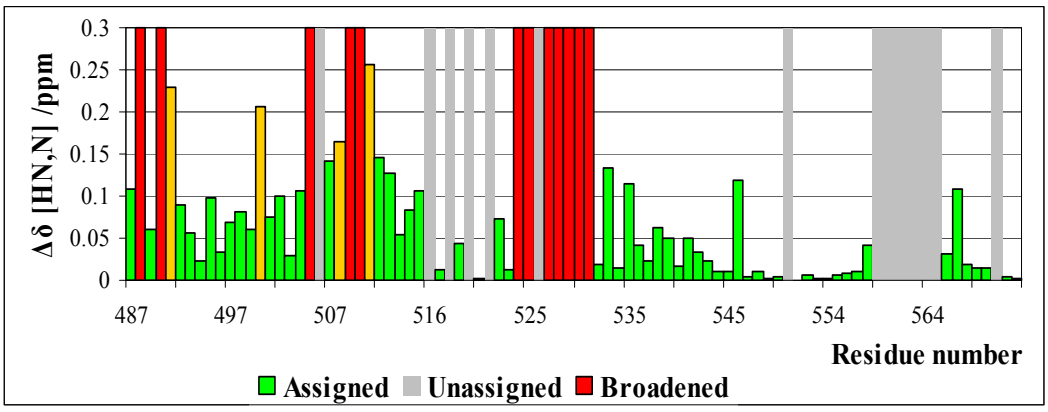



**Figure 4:15** Weighted shift map of the $^{8}$Fn1*-$^{9}$Fn1 [$^{1}$H-$^{15}$N]-HSQC crosspeaks before and after the addition of 0.4mM peptide. Green bars indicate residues with weighted shift under 0.15ppm, orange bars represent residues shifted between 0.15ppm and 0.3ppm, and red bars indicate residues whose peaks experienced severe line broadening in the bound state and could not be assigned. Grey bars indicate residues that were not assigned in either spectrum.



**Figure 4:16** Schematic representation of the $^{8}$Fn1*-$^{9}$Fn1 with the weighted shifts mapped onto the backbone using the same colour scheme as Figure 4:15 (coordinates provided by Dr. A. R. Pickford based on the $^{2}$Fn1-$^{3}$Fn1 crystal structure (Rudino-Pinera et al., 2007)). The $^{9}$Fn1 is set into the background of the figure since only the "A" strand is perturbed by peptide binding. Side chains of loop residues affected in binding are shown (diagram drawn in Pymol).

## 4.4. Discussion

### 4.4.1. Recombinantly expressed protein interacts with collagen

SPR provided a convenient method for comparing the function of recombinantly expressed Hex and proteolytically derived GBD without introducing tags or chemically modifying the protein which may have affected binding. Both proteins produced similar sensorgram profiles with almost identical equilibrium constants, confirming that both native and recombinant material interact with collagen in a similar way.

The Ingham group have produced similar binding profiles using short 30-residue collagen-like peptides as the ligand and proteolytically derived GBD as the analyte (Ingham et al., 2002). Their measured $K_d$ by SPR is very weak (1.8mM) but the synthetic 30-residue ligand $(POG)_{10}$ only vaguely mimics native collagen. In the same study they report a much tighter interaction between CNBr fragments of calf skin type I collagen and proteolytically derived GBD, measured by fluorescence anisotropy, and report a much tighter interaction with a $K_d$ of 1 to 4μM. The affinity of the interaction is within the same order of magnitude as determined here by SPR.

The $k_{on}$ and $k_{off}$ rates observed by SPR were high and could not be reliably fitted. These rapid on and off rates imply a transient interaction between analyte and ligand, but the abundance of binding sites per collagen chain (Ingham et al., 2002) would allow the GBD to associate and dissociate multiple times before being washed away from the sensorchip surface.

The SPR experiments were conducted at 37°C to be physiologically relevant and at a temperature where the collagen helix was predicted to be slightly unwound. The $K_d$ of GBD binding to collagen and gelatin were similar which suggested that the FN binding sites on gelatin were also exposed in triple helical collagen.

### 4.4.2. Role of glycosylation on the gelatin binding domain

Gelatin affinity chromatography served as both a method for purifying FN fragments and an analytical tool for comparing the affinity of binding. By expressing shortened fragments of the Hex, the influence of sugar chains linked to N399 and Asn511 on gelatin binding were compared. The removal of the bulk sugar chain from $^{8}$Fn1 (with Endo $H_f$) from the $^{8}$Fn1**-$^{9}$Fn1 module pair (linked to N511) had little effect on gelatin binding. In contrast, the complete removal of the sugar chain (with PNGase) resulted in the complete loss of binding, consistent with previous experiments that showed that the sugar residue was critical in stabilising the protein fold (Millard et al., 2005). Therefore, N511-linked high mannose sugar chain does not appear to inhibit or enhance binding of $^{8}$Fn1 to gelatin, but the presence of at least one N-linked carbohydrate residue at N511 is essential for this function.

The effect of removing the bulk carbohydrate chain from $^{2}$Fn2 (linked to N399) increased the gelatin binding affinity and this was more pronounced in the Quad over the Triple. The Ingham group investigated the gelatin binding properties of the 30kDa proteolytic fragment from the GBD (equivalent to the Quad), and found very little difference before and after treatment with Endo F to cleave the glucosamine linkage between the GlcNAc residues and the asparagine (Ingham et al., 1995).

Whether the high mannose sugar chain was properly removed was unclear, as they report that cleavage of with Endo F may have been incomplete. Endo F has been shown to become less effective as the size of a protein increases (Elder and Alexander, 1982) and the enzyme may not have been able to remove the glycosylation from the 30kDa protein. Endo $H_f$ cleavage is considered an efficient system for removing high mannose, N-linked carbohydrates (NEB), and may have been more successful than Endo F. The result of removing the carbohydrate from the Triple and the Quad suggests that the chain could partly obscure the binding interface between the $^{2}$Fn2 module and gelatin and this has a more prominent effect in the larger construct. The sugar chain, attached to a long flexible loop, could easily move to block the proposed $^{2}$Fn2 gelatin binding site (protein data bank ID code 1e8b (Pickford et al., 2001)).

The GBD is well documented to bind more tightly to gelatin than any smaller fragments of this domain (Ingham et al., 1989), (Hynes, 1990). In general, smaller constructs have weaker binding, as demonstrated with the Quad, Triple, and $^{8}$Fn1**-$^{9}$Fn1. Unexpectedly, the larger $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 construct bound more weakly than the smaller $^{8}$Fn1**-$^{9}$Fn1, and the $^{7}$Fn1-$^{8}$Fn1 module pair did not show any affinity for gelatin under these conditions. Urea was needed to remove the "N-terminal" constructs from the gelatin column containing Fn2 modules, whereas the "C-terminal" constructs without Fn2 modules eluted isocratically.

These results agree in part with those from the Ingham group who measured binding affinity of N-terminal $His_6$- and GFP-tagged GBD constructs by gelatin affinity

chromatography (Katagiri et al., 2003). Their $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1 (Hex) bound more tightly than $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1 (Quad) but their $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2 (Triple) showed no binding. Their $^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1 bound more tightly than $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2-$^{7}$Fn1 (Quad) and $^{7}$Fn1-$^{8}$Fn1 bound more tightly still. The discrepancies highlight the difficulties in working with gelatin and unknowns in eluting with urea. Enough urea will unfold any protein, and it is unclear whether the Ingham group were measuring elution by dissociation from the gelatin column, or elution by denaturation. The 30kDa GFP tag may have stabilised the smaller FN fragments and prolonged the interaction with gelatin, whereas in other cases, the tag may have sterically hindered binding and promoted dissociation from the column. The narrower and longer column used here was designed to eliminate the need for elution by urea and give a better indication of binding affinity. The C-terminal $His_6$ tag was shown to slightly affect binding but its size should not have been sufficient to cause loss of binding in the $^{7}$Fn1-$^{8}$Fn1** construct.

### 4.4.3. Fibronectin fragments bind to a synthetic collagen peptide

NMR titration of $^{8}$Fn1*-$^{9}$Fn1 with peptides from different regions of the collagen sequence allowed a comprehensive analysis of the binding interaction at the molecular level. The titration with peptide 2 allowed some residues of $^{8}$Fn1*-$^{9}$Fn1 involved in binding to be mapped onto a predicted model and allowed an estimate of the binding constant. The sequence of peptide 2 was taken from type I collagen (α1(I) chain), initiating two residues downstream of the collagenase cut site (Chung et al., 2004). This stretch is thought to be less stable due to its lowered Pro and Hyp (4-Hydroxyproline) content, and had previously been suggested as a FN binding region (Pagett et al., 2005).

Peptide 1 was taken from the same region of collagen, designed to include the collagenase cleavage cut site, and the 13 preceding residues. Any binding to $^{8}$Fn1*-$^{9}$Fn1 was too weak to be observed (a similar peptide to peptide 1 had previously been shown to bind to the Triple but only with a $K_d$ of ~6 mM (Pagett et al., 2005)). Peptide 4 was designed to explore the possibility of binding at an alternative site on collagen (Ingham et al., 2002). Peptide 3 was designed to eliminate the possibility that $^{8}$Fn1-$^{9}$Fn1 binds to any hydrophobic region of the collagen sequence and the absence of binding confirmed that there is some specificity in this NMR titration.

The affinity of $^{8}$Fn1-$^{9}$Fn1 for peptide 2 was in the same order of magnitude as that determined by the Ingham group. They suggested that the CNBr derived collagen fragment with N-terminus "GPRGFOGSO" (which is most similar to peptide 2 "GQRGVVGLOGQR**GERGFOGLO**GY") bound with a 1.0 ± 0.8 μM $K_d$ as judged by fluorescence anisotropy. By NMR, it was not possible to calculate a $K_d$ for binding to peptide 4. Attempts to measure the $K_d$ values by fluorescence were unsuccessful as the intrinsic fluorescence of $^{8}$Fn1*-$^{9}$Fn1 was too low, indicating that the three Trp side chains on the module pair are buried within the protein.

The Ingham group suggest another of their CNBr derived collagen fragments with N-terminus "OGHRGFSGL" (which is most similar to peptide 4 "ARGLOGTAGLOGM**KGHRGFSGL**DGAKY") bound with a $K_d$ of 3.4 ± 4.0 μM. All the binding collagen fragments from the Ingham group interacted with the GBD within this range of affinity, which agrees with the affinity determined for $^{8}$Fn1*-$^{9}$Fn1 binding to peptide 2.

The shift of individual crosspeaks in response to the peptide titration allowed tentative mapping of the peptide binding onto the $^{8}$Fn1*-$^{9}$Fn1 model structure. The majority of affected residues were within $^{8}$Fn1, although direct and indirect interactions could not be distinguished from these data. The contribution of $^{9}$Fn1 in binding was unclear as the majority of these residues were unperturbed. In this preliminary map, the direction of the peptide was unknown, and the peptide residues involved in binding were unconfirmed, so this schematic representation was appropriate (Figure 4:17). Computation and molecular dynamics/docking could be used at a later date to determine more information about this interaction.

### 4.4.4. The role of the each module in the gelatin binding domain

The difference in affinity of the C-terminal GBD fragments for gelatin may be explained by the proposed peptide binding schematic representation (Figure 4:17). PDB coordinates, provided by Dr. A. R. Pickford based on the $^{2}$Fn1-$^{3}$Fn1 crystal structure (Rudino-Pinera et al., 2007), were used as a basis to visualise the perturbations on $^{8}$Fn1**-$^{9}$Fn1. Gelatin affinity chromatography showed that $^{8}$Fn1**-$^{9}$Fn1 binds most tightly to gelatin (of these C-terminal FN fragments), but binding was reduced by the addition of $^{7}$Fn1, and binding was lost by the subsequent subtraction of the $^{9}$Fn1 module. The addition of $^{7}$Fn1 to $^{8}$Fn1**-$^{9}$Fn1 may have sterically hindered the binding of the peptide to the "A" strand of $^{9}$Fn1 and loop of $^{8}$Fn1**. The $^{7}$Fn1 could possibly partially inhibit the "A" strand of $^{8}$Fn1** from making contact with one of the ends of the peptide, causing a reduction in binding affinity. In addition, the model shows that the removal of $^{9}$Fn1 from $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 would reduce the size of the peptide binding interface by the loss of the linker

and "A" strand of $^{9}$Fn1. This would imply a very important role for the "E" strand of $^{8}$Fn1**, the linker region, and the "A" strand of $^{9}$Fn1 in gelatin binding.

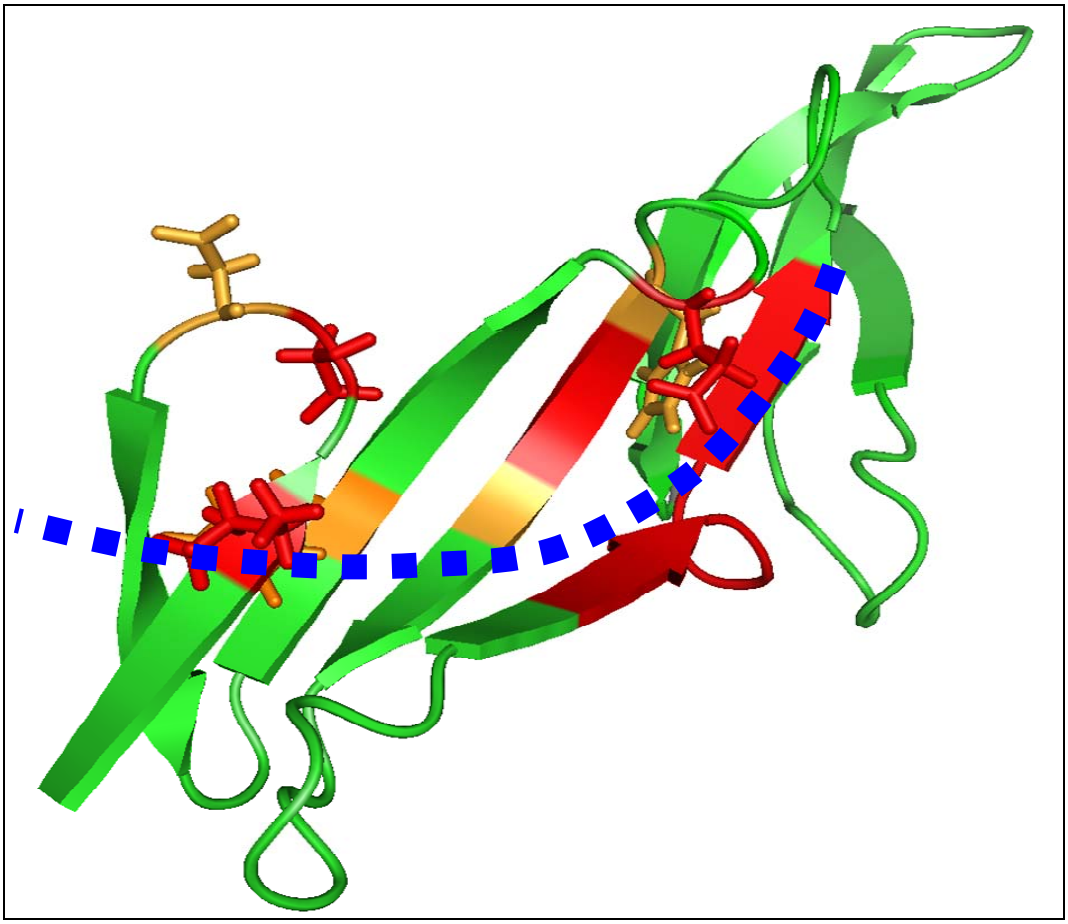


**Figure 4:17** A schematic representation to demonstrate how the collagen peptide may bind to the $^{8}$Fn1-$^{9}$Fn1 module pair. The diagram is orientated to give the reader a sense of looking "down" the $^{8}$Fn1 in the same way that a peptide would lie across the $^{8}$Fn1. The $^{9}$Fn1 is set into the background of the figure since only the "A" strand is perturbed by peptide binding. The speculative position of the bound peptide is shown (blue dotted line). The residues most affected in binding are shown in red, those less affected are in orange, and those unaffected in green (coordinates provided by Dr. A. R. Pickford based on the $^{2}$Fn1-$^{3}$Fn1 crystal structure (Rudino-Pinera et al., 2007)). Side chains of loop residues affected in binding are shown (diagram drawn in Pymol).

If the Quad and $^{8}$Fn1**-$^{9}$Fn1 are able to bind to gelatin, what is the role of $^{7}$Fn1? The Ingham group showed that the removal of $^{7}$Fn1 from the GBD to create a "gap construct" caused an almost negligible drop in the urea concentration at peak elution (3.0 M) from gelatin-sepharose (compared to 3.3 M for GBD) (Katagiri et al., 2003). In contrast, the added presence of $^{7}$Fn1 caused the Quad to bind much more tightly to

the gelatin column than the Triple, and even more confusingly, the presence of $^{7}$Fn1 decreased the affinity of $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 for gelatin. These results may be interpreted to imply that $^{7}$Fn1 does not actually contribute to gelatin binding directly, but instead provides a linker between $^{1}$Fn2-$^{2}$Fn2 and $^{8}$Fn1**-$^{9}$Fn1, which holds together the fully active gelatin binding domain. The result of the gap construct from the Ingham lab suggested that the $^{7}$Fn1 "linker" can be reduced or removed with little impact on the overall gelatin binding affinity of the GBD. Acting as a linker, $^{7}$Fn1 could stabilise the Triple by protecting from the denaturing effect of urea, and increase the apparent binding affinity of the Quad. Therefore, $^{7}$Fn1 may fulfil a similar role to $^{6}$Fn1, which has been suggested to have little or no role in direct gelatin binding, but provides a structural nucleus for the other modules to associate (Litvinovich et al., 1991), (Pagett et al., 2005).

## 4.5. References


Cavanagh, J. (2007) *Protein NMR Spectroscopy*. Elsevier Academic Press London.

Chung, L., Dinakarpandian, D., Yoshida, N., Lauer-Fields, J.L., Fields, G.B., Visse, R. and Nagase, H. (2004) Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis. *Embo J*, **23**, 3020-3030.

Elder, J.H. and Alexander, S. (1982) endo-beta-N-acetylglucosaminidase F: endoglycosidase from Flavobacterium meningosepticum that cleaves both high-mannose and complex glycoproteins. *Proc Natl Acad Sci U S A*, **79**, 4540-4544.

Engvall, E., Bell, M.L. and Ruoslahti, E. (1981) Affinity chromatography of collagen on collagen-binding fragments of fibronectin. *Coll Relat Res*, **1**, 505-516.

Engvall, E. and Ruoslahti, E. (1977) Binding of soluble form of fibroblast surface protein, fibronectin, to collagen. *Int J Cancer*, **20**, 1-5.

Engvall, E., Ruoslahti, E. and Miller, E.J. (1978) Affinity of fibronectin to collagens of different genetic types and to fibrinogen. *J Exp Med*, **147**, 1584-1595.

Forastieri, H. and Ingham, K.C. (1985) Interaction of gelatin with a fluorescein-labeled 42-kDa chymotryptic fragment of fibronectin. *J Biol Chem*, **260**, 10546-10550.

Guidry, C., Miller, E.J. and Hook, M. (1990) A second fibronectin-binding region is present in collagen alpha chains. *J Biol Chem*, **265**, 19230-19236.

Hynes, R. (1990) *Fibronectins*. Springer-Verlag, Berlin, Germany.

Ingham, K.C., Brew, S.A. and Migliorini, M. (2002) Type I collagen contains at least 14 cryptic fibronectin binding sites of similar affinity. *Arch Biochem Biophys*, **407**, 217-223.

Ingham, K.C., Brew, S.A. and Migliorini, M.M. (1989) Further localization of the gelatin-binding determinants within fibronectin. Active fragments devoid of type II homologous repeat modules. *J Biol Chem*, **264**, 16977-16980.

Ingham, K.C., Brew, S.A. and Novokhatny, V.V. (1995) Influence of carbohydrate on structure, stability, and function of gelatin-binding fragments of fibronectin. *Arch Biochem Biophys*, **316**, 235-240.

Isaacs, B.S., Brew, S.A. and Ingham, K.C. (1989) Reversible unfolding of the gelatin-binding domain of fibronectin: structural stability in relation to function. *Biochemistry*, **28**, 842-850.

Kadler, K.E., Holmes, D.F., Trotter, J.A. and Chapman, J.A. (1996) Collagen fibril formation. *Biochem J*, **316 ( Pt 1)**, 1-11.

Katagiri, Y., Brew, S.A. and Ingham, K.C. (2003) All six modules of the gelatin-binding domain of fibronectin are required for full affinity. *J Biol Chem*, **278**, 11897-11902.

Kleinman, H.K., McGoodwin, E.B., Martin, G.R., Klebe, R.J., Fietzek, P.P. and Woolley, D.E. (1978) Localization of the binding site for cell attachment in the alpha1(I) chain of collagen. *J Biol Chem*, **253**, 5642-5646.

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

Lindmark, H. and Guss, B. (1999) SFS, a novel fibronectin-binding protein from Streptococcus equi, inhibits the binding between fibronectin and collagen. *Infect Immun*, **67**, 2383-2388.

Litvinovich, S.V., Strickland, D.K., Medved, L.V. and Ingham, K.C. (1991) Domain structure and interactions of the type I and type II modules in the gelatin-binding region of fibronectin. All six modules are independently folded. *J Mol Biol*, **217**, 563-575.

Miekka, S.I., Ingham, K.C. and Menache, D. (1982) Rapid methods for isolation of human plasma fibronectin. *Thromb Res*, **27**, 1-14.

Millard, C.J., Campbell, I.D. and Pickford, A.R. (2005) Gelatin binding to the $^{8}$Fn1-$^{9}$Fn1 module pair of human fibronectin requires site-specific N-glycosylation. *FEBS Lett*, **579**, 4529-4534.

Mulder FA, Schipper D, Bott R, Boelens R. (1999) Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J Mol Biol*. **292**, 111-23.

Pagett, A., Campbell, I.D. and Pickford, A.R. (2005) Gelatin binding to the $^{6}$Fn1-$^{1}$Fn2-$^{2}$Fn2 fragment of fibronectin is independent of module-module interactions. *Biochemistry*, **44**, 14682-14687.

Pickford, A.R., Smith, S.P., Staunton, D., Boyd, J. and Campbell, I.D. (2001) The hairpin structure of the (6)F1(1)F2(2)F2 fragment from human fibronectin enhances gelatin binding. *Embo J*, **20**, 1519-1529.

Rudino-Pinera, E., Ravelli, R.B., Sheldrick, G.M., Nanao, M.H., Korostelev, V.V., Werner, J.M., Schwarz-Linek, U., Potts, J.R. and Garman, E.F. (2007) The solution and crystal structures of a module pair from the Staphylococcus aureus-binding site of human fibronectin--a tale with a twist. *J Mol Biol*, **368**, 833-844.

Sottile, J., Shi, F., Rublyevska, I., Chiang, H.Y., Lust, J. and Chandler, J. (2007) Fibronectin-dependent collagen I deposition modulates the cell response to fibronectin. *Am J Physiol Cell Physiol*.

Van Der Merwe, P. (2001) *Protein-ligand interactions: hydrodynamics and calorimetry*. Oxford University Press, Oxford.

Vuento, M., Salonen, E., Osterlund, K. and Stenman, U.H. (1982) Essential charged amino acids in the binding of fibronectin to gelatin. *Biochem J*, **201**, 1-8.

# 5. Stimulating fibroblast migration with FN

## 5.1. Fibroblast migration

Cell migration in multicellular organisms must be tightly controlled in order that cells are placed in the correct positions to form coherent and organised structures. The proteins which make up the ECM form a network of intimate contacts which can enhance or inhibit cell migration within the matrix. Cells have to respond to both diffusible and non-diffusible chemical and physical cues to adapt to their environment. Several of the chemical cues have been attributed to extracellular matrix proteins. In particular, a novel truncated form of FN (70kDa) was shown to stimulate cell migration (Schor et al., 2003). This potent migration-stimulating factor (MSF) has been suggested to be an effector in promoting cancer development (Schor et al., 1988) (See Figure 5:1).

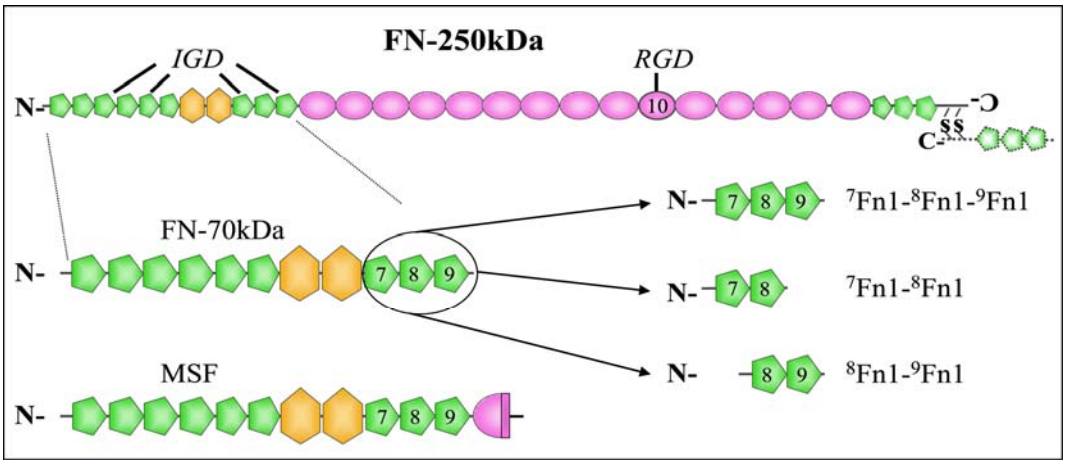



**Figure 5:1** The modular structure of FN, migration-stimulating factor (MSF), and Fn1 module fragments.

MSF was originally identified in fetal and cancer patient fibroblasts (Schor et al., 1988). *In vitro*, exposure of human dermal fibroblasts to MSF provokes a change in

phenotype, causing cells to migrate into 3-dimensional gels of native type-I collagen (Schor et al., 2003) (Figure 5:2). Full length FN is, however, devoid of MSF-like activity, suggesting that cryptic sites are exposed in MSF (Schor et al., 2003) as well as in certain N-terminal FN fragments (Schor et al., 1996).

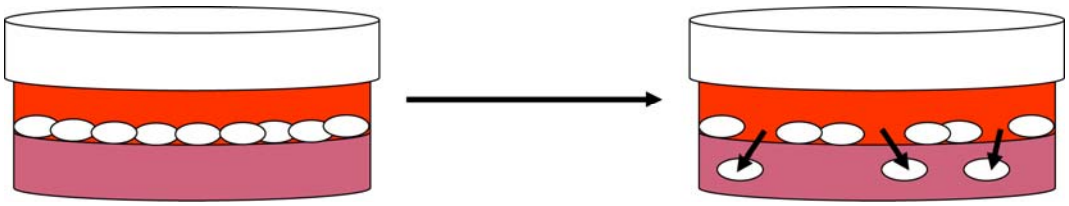


**Figure 5:2** An illustration of the cell migration experiment (Schor et al., 2003). Cells (white) are plated out onto a 3D- collagen matrix (dark red - lower layer) in the presence serum-free minimal essential medium (MEM) containing effector (light red - upper layer). The control plate was without effector. After 4 days incubation with MSF or other IGD containing peptides, the cell phenotype will change and cause the cells to migrate into the collagen matrix.

### 5.1.1. The IGD motif

The 70kDa N-terminal fragment of FN is composed of nine Fn1 modules and two Fn2 modules. Four of the nine Fn1 modules contain a highly conserved Ile-Gly-Asp (IGD) sequence found within the short loop between the two anti-parallel β-sheets of the module (Figure 5:3). The motogenic activity of MSF on fibroblasts has been attributed to this loop since mutation of IGD to DGI in modules 7 and 9 abolishes MSF bioactivity (Schor et al., 2003). Furthermore, soluble synthetic tri- and tetra-peptides containing the IGD amino acid motif can mimic MSF properties (Schor et al., 1999). Like MSF, these peptides stimulate the migration of human dermal fibroblasts into 3-dimensional type I collagen gels, although with much reduced potency; half maximal activity is observed with femtomolar and micromolar concentrations for MSF and IGD peptides respectively (Schor et al., 1999). The

addition of Ser or Gln C-terminal to the IGD tri-peptide was shown to enhance MSF-like activity by a factor of 10 (Schor et al., 1999).

The demonstration of bioactivity in a tri-peptide is reminiscent of several other activation motifs. The Arg-Gly-Asp (RGD) cell adhesion motif in the $^{10}$Fn3 module was the first tri-peptide to be isolated and has been shown to interact with integrin receptors (Ruoslahti and Pierschbacher, 1987), (Xiong et al., 2002). Synthetic peptides based on Leu-Asp-Val (LDV) and Arg-Glu-Asp-Val (REDV), found in the IIICS alternatively spliced site in FN, can mimic the activity of intact FN at least partially (Humphries et al., 1986; Mould and Humphries, 1991). While the RGD sequence has been shown to bind integrins and promote cell adhesion and fibril formation, the exact binding partner for the IGD motif is unknown, but its ability to promote cell migration, which is abolished by antibody to αvβ3 (Schor et al., 1999), suggests a possible interaction with integrins.

Full length fibronectin can form fibrils, and although much is known about promotion of fibrillogenesis by cells, little is known about the precise mechanism. Most models implicate integrin binding via the RGD recognition sequence. A recent paper surprisingly showed that the absence of a functional RGD motif in FN did not compromise the assembly of FN fibrils in mutant embryos or cells. A possible explanation suggested was that FN contains a novel integrin binding motif in its N-terminal region (Takahashi et al., 2007). The importance of the N-terminal 70kDa fragment was also emphasized recently when it was shown to initiate FN fibril formation in fibronectin-null mouse fibroblasts (Tomasini-Johansson et al., 2006).

These papers support the idea that fibrillogenesis might be initiated through integrin cell attachment at the N-terminus via a binding site other than the RGD region.

The structure-function relationships of the IGD sequence in the context of intact Fn1 domains are explored here. Of particular interest are the third and fourth IGD motifs in MSF, one in the seventh Fn1, and one in $^{9}$Fn1. Here the role of the IGD motif is examined by studying the properties of a series of fragments, namely $^{8}$Fn1-$^{9}$Fn1 and $^{7}$Fn1-$^{8}$Fn1 module pairs, each containing one IGD motif, and the $^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1 fragment that contains two IGD motifs. The effect of site specific mutations on migration and structure are examined using fibroblast migration assays and solution state NMR. Both IGD motifs are shown to have significant migration stimulating activity while contained within fully folded fibronectin modules.

## 5.2. Methods

### 5.2.1. Protein production

The human fibronectin modules $^{8}$Fn1**-$^{9}$Fn1 (residues 485-577), $^{7}$Fn1-$^{8}$Fn1** (residues 434-528), and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 (residues 434-577) containing two conservative point mutations (N497Q, R503K) were cloned into the *Pichia pastoris* expression vector pPICZλ (as previously described in chapter 2). Further point mutations were then introduced in these constructs to disrupt the IGD motifs. The following single mutations were introduced into $^{8}$Fn1**-$^{9}$Fn1 to disrupt the IGD motif in $^{9}$Fn1: I541R, I541V, I541A, D543E, D543A, and S544Q. The DGI mutant was introduced through two point mutations: I541R, D543E. A single mutation, I449R, was introduced in $^{7}$Fn1-$^{8}$Fn1** to disrupt the IGD motif in $^{7}$Fn1. The $^{9}$Fn1 IGD motif in $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 was disrupted with mutation I541R, and both IGD motifs were disrupted using the I449R, I541R double mutant. The nomenclature $^{8}$Fn1**-$^{9}$Fn1 (His) indicates a construct which contained a C-terminal $His_6$ tag. A C-terminal $His_6$ tag was present on all $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 constructs. The protein was deglycosylated and purified as previously described in chapter 2.

### 5.2.2. Cells

Collagen gel migration experiments were performed with human skin fibroblasts FSF44 (conducted by Dr. I.R. Ellis). Stock cultures were maintained in Eagle's Minimal Essential Medium containing 15% donor calf serum, as previously described (Schor, 1980).

### 5.2.3. Migration assays

The collagen gel assay (conducted by Dr. I.R. Ellis) was performed in 30mm plastic culture dishes containing preformed 2ml type I collagen gel (Figure 5:2), (Schor et al., 2003). 1ml serum-free MEM containing 4x the desired final concentration of effector molecule, and 1ml trypsinised fibroblasts (density 2x105 cells/ml) suspended in MEM containing 4% donor calf serum, were plated in duplicate onto the collagen gel to give a final volume of 4ml in 1% donor calf serum. After a 4 day incubation period at 37°C, cell migration into the 3-dimensional gel, in response to the effector, was determined by using an inverted-phase microscope to count the cells on the surface of the gel, and then focusing down through the gel to count the cells within the gel (Schor, 1980). This operation was repeated with 15 randomly selected areas on the gel surface and the results compared to a control plate without effector.

### 5.2.4. NMR spectroscopy

All NMR spectra were recorded at 25°C using spectrometers built in house, on samples of 1-2mM protein in 150mM NaCl, 20mM sodium phosphate, 1mM dioxane, 95% $H_2O$, 5% $D_2O$ at pH 6.6. NMR processing and analysis were performed as detailed in chapter 3. Backbone NH and Hα assignments have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number 15447 for $^{8}$Fn1*-$^{9}$Fn1 ($^{8}$Fn1*-$^{9}$Fn1 (R503K)) (see Appendix C).

## 5.3. Results

### 5.3.1. The location of the IGD motif

The solution structures of several Fn1 module pairs have already been solved by NMR (Potts et al., 1999), (Rudino-Pinera et al., 2007). The secondary structure map of $^{8}$Fn1**-$^{9}$Fn1 (as mapped using inter-strand NOEs (chapter 3; Figure 3.11) indicated that the IGD motif in $^{9}$Fn1 is found within a tight turn between the B strand and the C strand, and a model based on the structure of the $^{2}$Fn1-$^{3}$Fn1 module pair confirmed that the loop was accessible to the solvent (Figure 5:3).

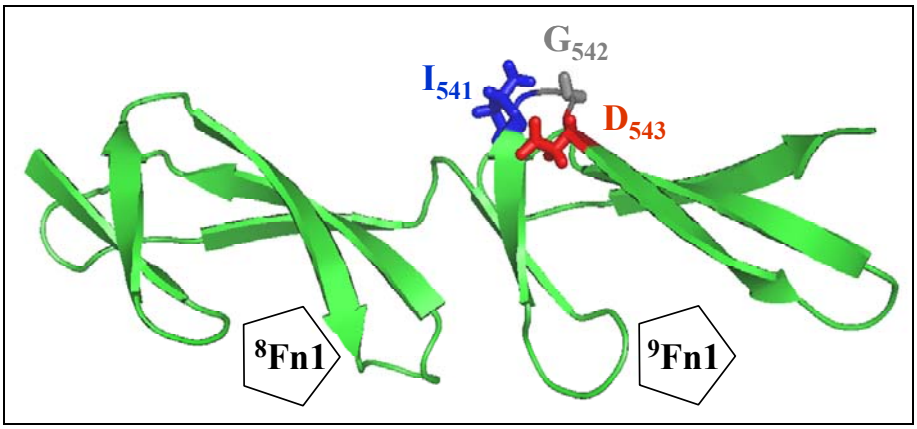



**Figure 5:3** A 3-dimensional model of the $^{8}$Fn1**-$^{9}$Fn1 module pair based on the secondary structure data in chapter 3; Figure 3:11, and the X-ray structure of the $^{2}$Fn1-$^{3}$Fn1 module pair (Rudino-Pinera et al., 2007). A short linker joins the two modules and the IGD motif falls with a tightly constrained loop in the $^{9}$F1. The side chains of the tightly constrained IGD loop are highlighted (isoleucine = blue, glycine = grey, aspartic acid = red) (http://www.pymol.org).

The heteronuclear $\{^{1}H\text{-}\}^{15}N$ NOE was used to map the internal motion of residues on a subnanosecond timescale, showing the module pair to be relatively rigid, especially in the short linker between modules and in the IGD loop of $^{9}$Fn1 (Figure 5:4). Those values lower than 0.65 indicate regions with fast internal mobility, the majority falling within the loop regions of the protein.

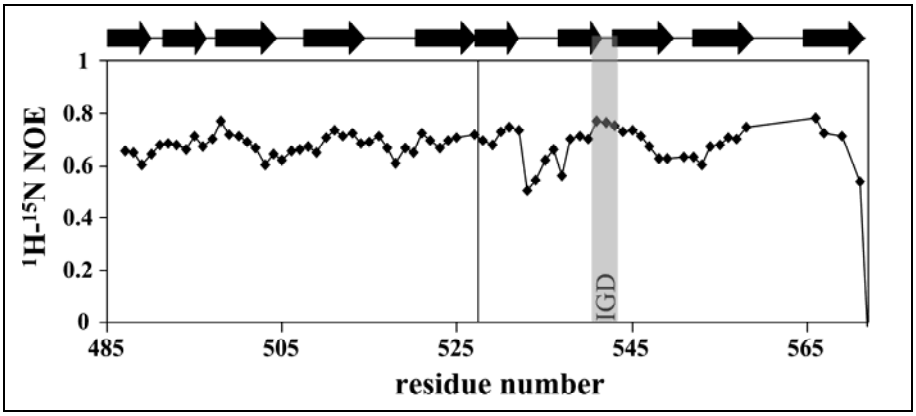



**Figure 5:4** The heteronuclear {$^{1}$H-}$^{15}$N NOE of $^{8}$Fn1*-$^{9}$Fn1. The horizontal line indicates the module boundary and the boxed area indicates the position of the IGD motif. The secondary structure is indicated in schematic form. The DE loop of the $^{9}$Fn1 remains unassigned due to intermediate exchange of the residues on the NMR timescale.

### 5.3.2. Expression, purification, and characterisation of IGD mutants

Conservative and non-conservative point mutations were introduced into $^{8}$Fn1**-$^{9}$Fn1 to interrupt the IGD site in the $^{9}$Fn1 module. These constructs were labelled, expressed, and purified; their fold was examined by comparing each [$^{1}$H-$^{15}$N]-HSQC spectrum with that of the unmodified $^{8}$Fn1**-$^{9}$Fn1 module pair. All seven mutants were equally soluble and were folded in the same way as shown by the limited $^{1}H_{N}$ and $^{15}$N chemical shift differences from the "wild-type" $^{8}$Fn1**-$^{9}$Fn1 (Figure 5:5), (Figure 5:6). As expected, there were minor shifts of peaks from residues in the IGD and of residue S544 which immediately follows. Other observed localised shifts were in the cross-strand residue C530 (in strand A of $^{9}$F1) and C558 (in strand D of $^{9}$F1); these sidechains are predicted by structural homology models to point towards the IGD loop (data not shown). Apart from these minor local perturbations around the mutated residue, there was no evidence for a change in the overall fold.

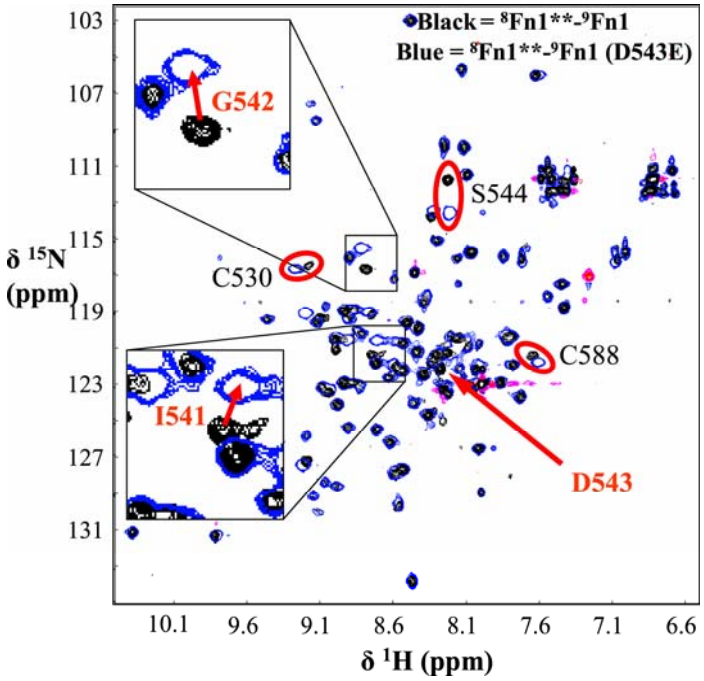



**Figure 5:5** Overlay of the "wild type" and the D543E $^{8}$Fn1**-$^{9}$Fn1 [$^{1}$H-$^{15}$N]-HSQC spectra. The mutated residue D543 disappears from the $^{8}$Fn1**-$^{9}$Fn1 spectrum and there are significant crosspeak shifts surrounding the IGD site. Chemical environments are largely unchanged for the other residues and the module fold is unperturbed. Spectra were taken at pH 6.6 and 25°C on a 750MHz spectrometer.

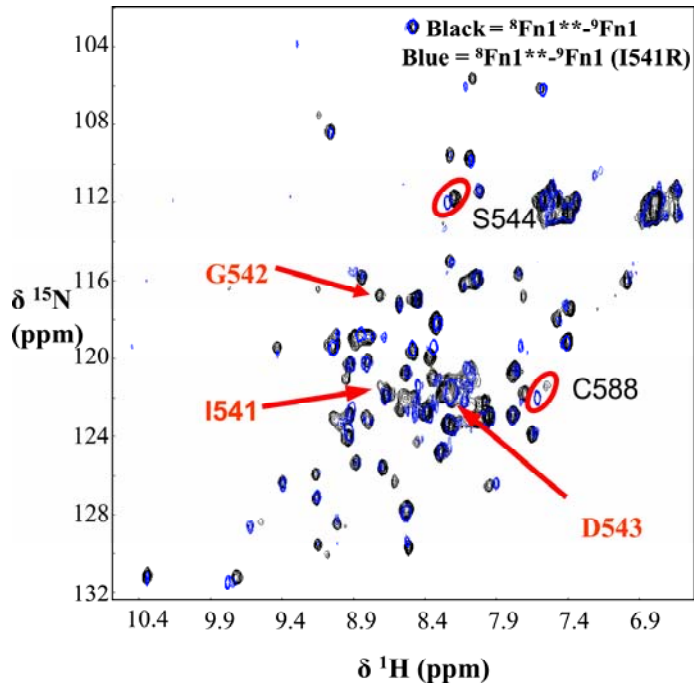



**Figure 5:6** Overlay of the "wild type" and the I541R $^{8}$Fn1**-$^{9}$Fn1 [$^{1}$H-$^{15}$N]-HSQC spectra. The majority of the I541R spectrum can be assigned from the $^{8}$Fn1**-$^{9}$Fn1 spectrum and those residues close to the IGD loop are indicated. The majority of residues are unaffected by the mutation and the module fold is unperturbed. Spectra were taken at pH 6.6 and 15°C on a 750MHz spectrometer.

### 5.3.3. An intact IGD motif is required to stimulate fibroblast migration

The effect of each module pair on the behaviour of fibroblast cells was assessed in a collagen gel migration assay (Figure 5:7). Increasing concentrations of the module pairs were added to cells plated on the surface of 3-dimensional gels of native type I collagen fibres. In general, eight cells were seen to have migrated in each of the randomly selected areas on the plate with effector after 4 days, compared to two cells per randomly selected area on the control plate without effector (Schor et al., 2006).

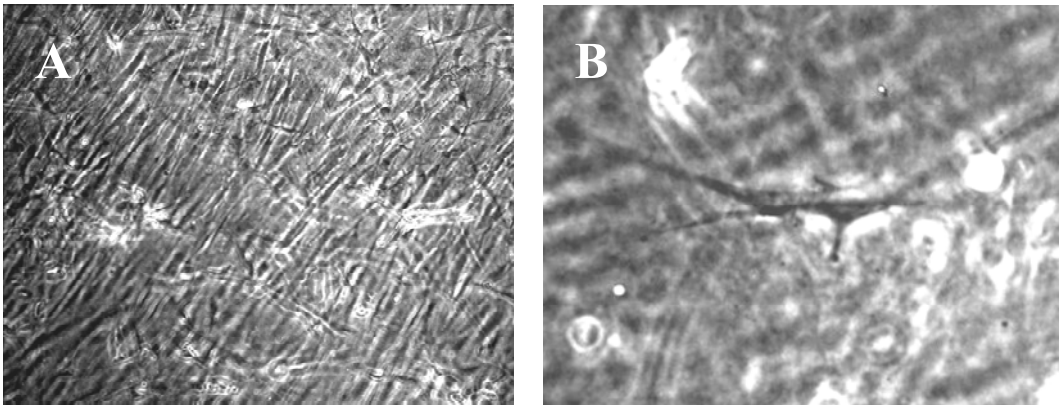



**Figure 5:7** A light microscope image of the collagen gel migration assay after a 4 day incubation period with effector at 37°C. A, confluent fibroblast cells can be seen on the collagen gel surface. B, an enlarged image of a fibroblast cell migrating into the collagen matrix.

In a similar manner to that of MSF and IGD peptides (Schor et al., 1999), $^{8}$Fn1**-$^{9}$Fn1 with an intact IGD motif stimulated fibroblast migration into this matrix, in a bell-shaped dose-response fashion. Maximal bioactivity was expressed at concentrations of 100 pmolar (Figure 5:8A). As predicted from peptide studies (Schor et al., 1999), the module pairs with a mutation at I541 (VGD, AGD, RGD) did not affect cell migration (Figure 5:8A). The DGI mutant was also unable to stimulate migration (Figure 5:8B). In contrast, those module pairs carrying a mutation at D543 (IGE and IGA) showed some bioactivity but this was significantly reduced with respect to the intact IGD at the same protein concentrations (Figure 5:8B). It has also

been suggested that the fourth amino acid in the IGD loop could affect migration (Schor et al., 1999). The amino acid immediately following the IGD was also mutated from Ser to Gln in $^{8}$Fn1**-$^{9}$Fn1 but did not change the extent of cell migration (Figure 5:8B).

The $^{8}$Fn1**-$^{9}$Fn1 module pair (R503K, N497Q) was shown to be as active as $^{8}$Fn1*-$^{9}$Fn1 (R503K) indicating that the removal of the glycosylation site does not affect the bioactivity of the protein (Figure 5:8B). $^{8}$Fn1**-$^{9}$Fn1 (His) was also shown to exhibit the same bioactivity as the module pair without the His tag (Figure 5:8B).

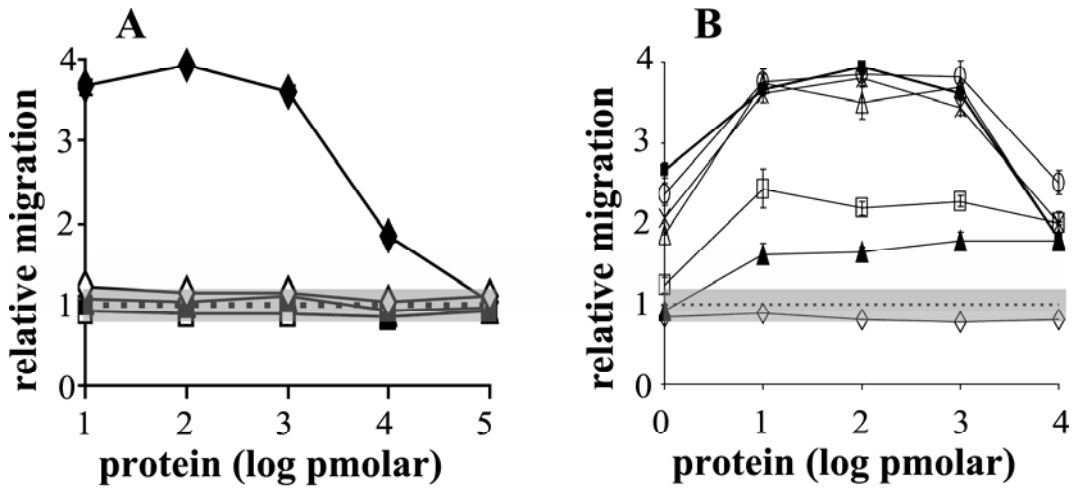



**Figure 5:8** The effects of $^{8}$Fn1**-$^{9}$Fn1 module pairs on cell migration. Fibroblasts were plated on native collagen gels in the presence of intact IGD and mutant test proteins at the concentrations indicated. After a 4-day incubation period, the cells present on the surface and within the 3D-collagen matrix were counted. The percentage of cells that migrated into the gel is expressed relative to the control cultures (cells plated onto native collagen without effector). For clarity, the error bars are omitted for those results that fall within the error of the control, but are indicated by the shaded area. A, $^{8}$Fn1**-$^{9}$Fn1 (R503K, N497) module pair; IGD (—♦—), VGD (-◊-), AGD (-▲-), RGD (-□-). B, $^{8}$Fn1**-$^{9}$Fn1 module pair; IGD (—■—), IGE (-□-), IGA (-▲-), DGI (-◊-), IGD (Q) {= $^{8}$Fn1**-$^{9}$Fn1 (S544Q)} (-X-), IGD' {= $^{8}$Fn1**-$^{9}$Fn1 with C-terminal His tag} (-Δ-), IGD (*) {= $^{8}$Fn1*-$^{9}$Fn1 (R503K)} (-○-).

### 5.3.4. The relative roles of IGD in $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1

The $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 fragments were judged by SDS-PAGE to express with yields of approximately 80mg/l and 40mg/l, respectively. The lower

yield of the module triple is presumably due to some misfolding, and therefore degradation, of this larger recombinant fragment. The identity of the proteins was confirmed by N-terminal sequencing and ESI-MS (measured molecular weights of 12061Da and 18393Da were within 1Da of theoretical). The recombinant material was isotopically labelled with $^{15}N$ and examined by NMR. The dispersion of crosspeaks in [$^{1}H$-$^{15}N$]-HSQC spectra was indicative of a fully folded protein (data not shown).

Following the results from the $^{8}$Fn1**-$^{9}$Fn1 mutations the relative roles of the IGD motifs in $^{7}$F1 and $^{9}$F1 were explored by replacing IGD with RGD in $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1. The proteins were successfully expressed, purified to homogeneity and characterised by western blotting and ESI-MS. The two IGD motifs in $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 were mutated in turn by introducing single point mutations I449R and I541R, before introducing both mutations to create a double mutant. The structure of the mutants, before and after the Ile to Arg mutation, was judged not to have changed significantly by comparing resolved up-field methyl peaks in 1D NMR spectra (data not shown).

### 5.3.5. The IGD motifs in $^{7}$F1 and $^{9}$F1 can act independently

The $^{7}$Fn1-$^{8}$Fn1** module pair with the intact IGD stimulated fibroblast migration at the same concentration as $^{8}$Fn1**-$^{9}$Fn1 (Figure 5:9). The $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 fragment with both IGD motifs stimulated fibroblast migration with a factor of 100 higher potency with maximal activity at 1 pmolar/IGD. Mutation of the $^{7}$Fn1 IGD motif in $^{7}$Fn1-$^{8}$Fn1** caused significant loss of bioactivity. Independent mutation of the

seventh and the ninth IGD motifs in $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 reduced the maximal bioactivity to the level of $^{7}$Fn1-$^{8}$Fn1** but the activity was absent in the double mutant (Figure 5:8B).

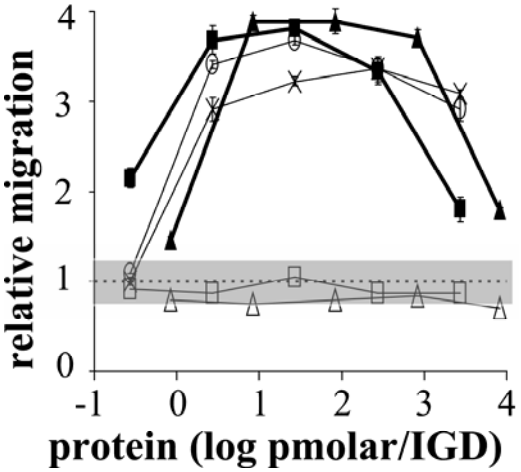



**Figure 5:9** Fibroblast migration into collagen gels. The response to different constructs is shown; $^{7}$Fn1-$^{8}$Fn1** and $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1. The percentage of cells that migrated into the 3D-collagen matrix after 4 days is expressed relative to the control cultures (no effector present). $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 (R503K, N497); 2 x IGD (—■—), RGD $^{7}$Fn1 (-○-), RGD $^{9}$Fn1 (-x-), 2 x RGD (-□-). $^{7}$Fn1-$^{8}$Fn1** module pair; IGD (—▲—), RGD (-Δ-).

## 5.4. Discussion

### 5.4.1. Fibronectin fragments containing the IGD motif are bioactive

Work in collaboration with Dr. I.R. Ellis (Dundee) has allowed an extension of earlier work on the bioactivity of soluble synthetic IGD peptides (Schor et al., 1999) to that of IGD motifs in the context of folded Fn1 modules. Various Fn1 module fragments, including several IGD mutants, were produced in *P. pastoris* and their ability to stimulate fibroblast migration was assessed in a collagen gel migration assay. All of the proteins produced had a similar fold, as judged by NMR. A 10,000 fold increase in potency was observed in moving from peptides to recombinant module pairs.

The conservative I541V mutation was designed to mimic the VGD of the first Fn1 module; it was unable to promote cell migration in the context of $^{8}$Fn1**-$^{9}$Fn1. The I541R mutation was designed to introduce an RGD sequence, a motif which promotes cell adhesion when in other regions of FN e.g. $^{10}$Fn3 (Ruoslahti and Pierschbacher, 1987). This mutation did not, however, promote cell migration here; neither did I541A. These results suggest that the Ile is critical for the observed bioactivity.

The bioactivity of $^{7}$Fn1-$^{8}$Fn1** is very similar to that of $^{8}$Fn1**-$^{9}$Fn1 (Figure 5:8A and Figure 5:9). Substitution of I449R in the $^{7}$Fn1-$^{8}$Fn1** module pair also resulted in a protein devoid of migration stimulating activity.

About half the bioactivity remained in the protein with the conservative D543E mutation. Some activity also remained for the D543A mutation but it was

significantly lower than the "wild type" $^{8}$Fn1**-$^{9}$Fn1. These results suggest that the Asp residue in the IGD is required for optimal bioactivity but is less critical than the Ile. The "reverse" mutant DGI, was also devoid of motogenic activity.

It has been suggested that the IGDS of $^{9}$Fn1 is more active than the IGDQ in $^{7}$Fn1 (Schor et al., 1999), but the mutation S544Q did not change the bioactivity of the protein. This finding suggests that the fourth amino acid is less important in the context of the $^{8}$Fn1**-$^{9}$Fn1 module pair than in the unconstrained peptide form.

### 5.4.2. Increased potency in the context of a folded protein

The increased potency of the protein in comparison to the peptide is remarkable. Constraining the ends of the IGD motif in the $^{9}$Fn1 module may favour the orientation of ligand binding which accounts for migration stimulation at much lower concentrations. A cyclic IGD peptide containing 20 amino acids from around the IGD site is also more potent than a similar length unconstrained IGD peptide (data not shown). Cyclic RGD peptides can inhibit cell attachment at a 20-fold lower concentration than linear peptides (Kumagai et al., 1991). The insertion of RGD motif into mutant lysozyme significantly increased the cell adhesion activity (Yamada et al., 1994), again implying that the introduction of conformational constraints can increase affinity to the integrin receptor, although a constrained RGD sequence in the IGD loop of Fn1 modules does not enhance cell migration. In the current study, the IGD motif as presented by the $^{9}$Fn module structure has markedly increased the bioactivity compared to free peptides.

The heteronuclear $\{^{1}H-\}^{15}N$ NOE data on $^{8}$Fn1*-$^{9}$Fn1 here, as well as previous dynamics studies of $^{2}$Fn1-$^{3}$Fn1 and $^{4}$Fn1-$^{5}$Fn1 (Potts et al., 1999), show that the IGD motif is in a relatively immobile part of the module. This inflexibility in the IGD loop is quite different from the considerable flexibility of the RGD loop observed in $^{10}$F3 (Main et al., 1992).

The possibility of cooperativity was also explored by studying the effect of two IGD motifs in one protein using a triple module construct ($^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1) containing an IGD motif in both $^{7}$Fn1 and $^{9}$Fn1. The maximal activity of two IGD motifs occurred at significantly lower concentration than a single IGD (the concentration dependence of the MSF activity of different peptides and proteins is summarised in Table 4). It is interesting to note that the gelatin binding domain, which contains the same two IGD motifs, is some 1000 fold more potent than the triple module (Schor et al., 1996). Migration stimulating activity was only lost from the triple module protein after mutation of both of the IGD sites, confirming that they are both able to stimulate migration independently as well as acting together.

| IGD peptides | $^{8}$Fn1**-$^{9}$Fn1 | $^{7}$Fn1-$^{8}$Fn1**-$^{9}$Fn1 | GBD | MSF |
|---|---|---|---|---|
| $10^{6}$ | $10^{2}$ | 1 | $10^{-4}$ | $10^{-6}$ |

**Table 4** The relative potency of IGD motifs in different contexts. The peptide/protein concentration (pmolar) showing maximal motogenic activity in the fibroblast migration assay was divided by the number of IGD motifs in that construct. GBD = gelatin-binding domain, MSF = migration stimulating factor.

### 5.4.3. A potential IGD signalling pathway

The profile of cell migration at different concentrations takes a bell shape suggesting that a threshold concentration is required to initiate cell migration and that a negative

feedback mechanism occurs at higher concentrations. This is somewhat similar to a previous observation (Fukai et al., 1991), where a 21kDa fragment of FN stimulated the migration of fibroblasts in a bell-shape dose-response pattern. It was previously shown (Schor et al., 1999) that fibroblast migration stimulated by IGD peptides is dependent on both αvβ3 functionality and tyrosine phosphorylation of focal adhesion kinase. The reduced migration at high concentration may thus arise from effects on integrin clustering and/or subsequent downstream signalling.

While integrin αvβ3 cannot be confirmed as the prime candidate receptor for the IGD motif, it is clear that the presence of αvβ3 and its downstream signalling components (phosphatidylinositol-3-kinase and focal adhesion kinase) are required for IGD motogenic activity (Schor et al., 1999). αvβ3 integrin also appears to cooperate with activated growth factor receptors, such as epidermal growth factor receptors, in mediating their respective motogenic signalling in response to ligand binding (Ellis et al., 2007). The RGD loop has been established as a key recognition sequence for αvβ3 (Xiong et al., 2002), but changing IGD to RGD in Fn1 domains results in loss, not gain, of motogenic activity. This apparent discrepancy may be due to the inherent inflexibility of the RGD in this context.

A recent report suggested that Asn can spontaneously deamidate to isoAsp to create an isoDGR sequence (Curnis et al., 2006) that can interact with αvβ3. There is a potential isoDGR in $^{7}$Fn1 that could constitute a binding ligand for the αvβ3. But since the $^{8}$Fn1**-$^{9}$Fn1 lacks the NGR sequence that could deamidate to isoDGA, and $^{7}$Fn1-$^{8}$Fn1** and $^{8}$Fn1**-$^{9}$Fn1 both exhibit the same potency in the collagen gel

migration assay, it is unlikely that this receptor-activation mechanism is part of the IGD response. Of course integrins could play a critical mechanistic role in mediating motogenic signalling by IGD ligation to some unknown cell surface receptor.

### 5.4.4. The IGD motif appears multiple times in fibronectin

The four IGD motifs at the N-terminus of FN are invariant in human, rat, and mouse (see Figure 5:1). Three are also invariant in frogs, with a conservative mutation in the fourth. From the current work and knowledge of Fn1 module structure it seems likely that all these IGD motifs will display bioactivity as long as the Fn1 modules are exposed. Only two out of the four IGD motifs of MSF ($^{7}$Fn1 and $^{9}$Fn1) appear to be active regarding stimulation of fibroblast migration (5). Consistent with this conclusion, a proteolytically-derived gelatin-binding domain of FN (GBD, Sigma), which contains both $^{7}$Fn1 and $^{9}$Fn1, displays motogenic activity, whereas the proteolytically-derived $^{1}$Fn1-$^{5}$Fn1 (Sigma) was inactive (Schor et al., 2003). This suggests that $^{3}$Fn1 and $^{5}$Fn1 may be cryptically hidden in MSF or $^{1}$Fn1-$^{5}$Fn1. Even if all four IGD motifs were exposed and bioactive in MSF, a simple association between exposed IGD motifs and motogenic activity would not explain the huge differences in potency between MSF and IGD-containing peptides (Table 4). MSF (with four IGDs in four Fn1 modules) has been shown to stimulate fibroblast migration at 100 fold lower concentration than the gelatin-binding domain (GBD) (with two IGDs) (Schor et al., 1996), (Schor et al., 2006). This could suggest that IGD motifs in larger fragments are able to act cooperatively to give a much increased response.

The enormous enhancements of IGD potency in different contexts is, however, not easy to explain. One possibility is that the increased bioactivity arises from a precise geometric presentation of the different binding motifs. Another more likely possibility is that several steps are involved; for example an initial binding to matrix, followed by receptor encounter and intracellular signalling. It is additionally possible that the IGD-containing Fn1 modules might partially denature the collagen triple helix, exposing RGD triplets, to which the αvβ3 integrin could then bind. This would explain the high potency of those IGD-containing Fn1 modules which have high collagen/gelatin binding activity, and distinguish them from the N-terminal domain, which has no motogenic activity. Of further interest is the observation that MSF displays proteolytic activity (Houard et al., 2005).

### 5.4.5. The bioactive IGD motif may be cryptically hidden

Another intriguing observation is that full length FN does not show bioactivity (Schor et al., 2003). The usual explanation is that IGD motifs are hidden in assembled, intact FN but bioactive when exposed; for example, as shown here and previously (Schor et al., 1996), bioactivity appears in smaller fragments. FN fragments created by proteolysis exhibit many properties not seen in full length FN (Homandberg, 1999; Ingham et al., 2004), and cryptic sites can be exposed simply by selective sequence deletion (Sechler et al., 2001). Mechanical tension has also been shown to unravel FN (Oberhauser et al., 2002) and potential new sites can be revealed through conformational change and molecular rearrangement. The number of cryptic sites revealed at any one time could define the size and nature of the response.

Previous studies have suggested that biological activities of FN such as fibrillogenesis can be controlled by concealing active motifs within the tertiary structure (Vakonakis et al., 2007). The 70kDa FN fragment has been shown to be critical in fibril assembly (Schwarzbauer, 1991) and much work has gone into establishing the matrix assembly site. A possible explanation is that the IGD motifs studied here could also contribute to fibrillogenesis.

## 5.5. References


Bright, J.R., Pickford, A.R., Potts, J.R. and Campbell, I.D. (2000) Preparation of isotopically labeled recombinant fragments of fibronectin for functional and structural study by heteronuclear nuclear magnetic resonance spectroscopy. *Methods Mol Biol*, **139**, 59-69.

Curnis, F., Longhi, R., Crippa, L., Cattaneo, A., Dondossola, E., Bachi, A. and Corti, A. (2006) Spontaneous formation of L-isoaspartate and gain of function in fibronectin. *J Biol Chem*, **281**, 36466-36476.

Ellis, I.R., Schor, A.M. and Schor, S.L. (2007) EGF AND TGF-alpha motogenic activities are mediated by the EGF receptor via distinct matrix-dependent mechanisms. *Exp Cell Res*, **313**, 732-741.

Fukai, F., Suzuki, H., Suzuki, K., Tsugita, A. and Katayama, T. (1991) Rat plasma fibronectin contains two distinct chemotactic domains for fibroblastic cells. *J Biol Chem*, **266**, 8807-8813.

Homandberg, G.A. (1999) Potential regulation of cartilage metabolism in osteoarthritis by fibronectin fragments. *Front Biosci*, **4**, D713-730.

Houard, X., Germain, S., Gervais, M., Michaud, A., van den Brule, F., Foidart, J.M., Noel, A., Monnot, C. and Corvol, P. (2005) Migration-stimulating factor displays HEXXH-dependent catalytic activity important for promoting tumor cell migration. *Int J Cancer*, **116**, 378-384.

Humphries, M.J., Akiyama, S.K., Komoriya, A., Olden, K. and Yamada, K.M. (1986) Identification of an alternatively spliced site in human plasma fibronectin that mediates cell type-specific adhesion. *J Cell Biol*, **103**, 2637-2647.

Ingham, K.C., Brew, S.A. and Erickson, H.P. (2004) Localization of a cryptic binding site for tenascin on fibronectin. *J Biol Chem*, **279**, 28132-28135.

Kumagai, H., Tajima, M., Ueno, Y., Giga-Hama, Y. and Ohba, M. (1991) Effect of cyclic RGD peptide on cell adhesion and tumor metastasis. *Biochem Biophys Res Commun*, **177**, 74-82.

Main, A.L., Harvey, T.S., Baron, M., Boyd, J. and Campbell, I.D. (1992) The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell*, **71**, 671-678.

Mould, A.P. and Humphries, M.J. (1991) Identification of a novel recognition sequence for the integrin alpha 4 beta 1 in the COOH-terminal heparin-binding domain of fibronectin. *Embo J*, **10**, 4089-4095.

Oberhauser, A.F., Badilla-Fernandez, C., Carrion-Vazquez, M. and Fernandez, J.M. (2002) The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J Mol Biol*, **319**, 433-447.

Potts, J.R., Bright, J.R., Bolton, D., Pickford, A.R. and Campbell, I.D. (1999) Solution structure of the N-terminal F1 module pair from human fibronectin. *Biochemistry*, **38**, 8304-8312.

Rudino-Pinera, E., Ravelli, R.B., Sheldrick, G.M., Nanao, M.H., Korostelev, V.V., Werner, J.M., Schwarz-Linek, U., Potts, J.R. and Garman, E.F. (2007) The solution and crystal structures of a module pair from the Staphylococcus aureus-binding site of human fibronectin--a tale with a twist. *J Mol Biol*, **368**, 833-844.

Ruoslahti, E. and Pierschbacher, M.D. (1987) New perspectives in cell adhesion: RGD and integrins. *Science*, **238**, 491-497.

Schor, S.L. (1980) Cell proliferation and migration on collagen substrata in vitro. *J Cell Sci*, **41**, 159-175.

Schor, S.L., Ellis, I., Banyard, J. and Schor, A.M. (1999) Motogenic activity of IGD-containing synthetic peptides. *J Cell Sci*, **112 ( Pt 22)**, 3879-3888.

Schor, S.L., Ellis, I., Dolman, C., Banyard, J., Humphries, M.J., Mosher, D.F., Grey, A.M., Mould, A.P., Sottile, J. and Schor, A.M. (1996) Substratum-dependent stimulation of fibroblast migration by the gelatin-binding domain of fibronectin. *J Cell Sci*, **109 ( Pt 10)**, 2581-2590.

Schor, S.L., Ellis, I.R., Harada, K., Motegi, K., Anderson, A.R., Chaplain, M.A., Keatch, R.P. and Schor, A.M. (2006) A novel 'sandwich' assay for quantifying chemo-regulated cell migration within 3-dimensional matrices: wound healing cytokines exhibit distinct motogenic activities compared to the transmembrane assay. *Cell Motil Cytoskeleton*, **63**, 287-300.

Schor, S.L., Ellis, I.R., Jones, S.J., Baillie, R., Seneviratne, K., Clausen, J., Motegi, K., Vojtesek, B., Kankova, K., Furrie, E., Sales, M.J., Schor, A.M. and Kay, R.A. (2003) Migration-stimulating factor: a genetically truncated onco-fetal fibronectin isoform expressed by carcinoma and tumor-associated stromal cells. *Cancer Res*, **63**, 8827-8836.

Schor, S.L., Schor, A.M., Grey, A.M. and Rushton, G. (1988) Foetal and cancer patient fibroblasts produce an autocrine migration-stimulating factor not made by normal adult cells. *J Cell Sci*, **90 ( Pt 3)**, 391-399.

Schwarzbauer, J.E. (1991) Identification of the fibronectin sequences required for assembly of a fibrillar matrix. *J Cell Biol*, **113**, 1463-1473.

Sechler, J.L., Rao, H., Cumiskey, A.M., Vega-Colon, I., Smith, M.S., Murata, T. and Schwarzbauer, J.E. (2001) A novel fibronectin binding site required for fibronectin fibril growth during matrix assembly. *J Cell Biol*, **154**, 1081-1088.

Takahashi, S., Leiss, M., Moser, M., Ohashi, T., Kitao, T., Heckmann, D., Pfeifer, A., Kessler, H., Takagi, J., Erickson, H.P. and Fassler, R. (2007) The RGD motif in fibronectin is essential for development but dispensable for fibril assembly. *J Cell Biol*, **178**, 167-178.

Tomasini-Johansson, B.R., Annis, D.S. and Mosher, D.F. (2006) The N-terminal 70-kDa fragment of fibronectin binds to cell surface fibronectin assembly sites in the absence of intact fibronectin. *Matrix Biol*, **25**, 282-293.

Vakonakis, I., Staunton, D., Rooney, L.M. and Campbell, I.D. (2007) Interdomain association in fibronectin: insight into cryptic sites and fibrillogenesis. *Embo J*, **26**, 2575-2583.

Xiong, J.P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L. and Arnaout, M.A. (2002) Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science*, **296**, 151-155.

Yamada, T., Uyeda, A., Kidera, A. and Kikuchi, M. (1994) Functional analysis and modeling of a conformationally constrained Arg-Gly-Asp sequence inserted into human lysozyme. *Biochemistry*, **33**, 11678-11683.

# 6. The gelatin binding domain

## 6.1. A multifaceted domain

Over the past 50 years, knowledge and understanding of the extracellular matrix has undergone a revolution, from being regarded as a simple inert scaffold to one of an interactive medium. The matrix is now known to be constantly remodelled through two-way signalling between cells and matrix proteins. Knowledge of the structure of FN has also gone through a similar revolution from an extended beads-on-a-string model to one of a more compact globular arrangement. The modules of FN have been shown to be carefully assembled to promote contacts within their domain and hence enhance the active properties of each domain. The gelatin binding domain of FN, which is composed of six modules, remained to be comprehensively characterised both in terms of interaction between the individual modules, and as one entire functional unit. The aim of this study was to further our understanding of the protein, both in terms of structure and function, building on the work from this lab and from others round the world.

### 6.1.1. Successful expression of gelatin binding domain fragments

Heterologous overexpression of FN fragments in *P. pastoris* was a convenient way of producing milligram quantities of labelled and unlabelled protein for study. *E. coli* may not always have been able to deal with the complex disulphide bonding pattern and glycosylation of the Hex, with two bonds per module (12 bonds in all) and three glycosylation sites, but *P. pastoris* was able to cope with the post-translational modification and express fully folded functional protein. The low cost of labelled

fermentation reagents meant that protein could be expressed in *P. pastoris* in large enough amounts for NMR.

NMR was a convenient method for examining the structure of the fragments. One-dimensional NMR was used to verify the folding of unlabelled material. Multi-dimensional NMR was especially useful in characterising the fold of the modules, in particular the $^{8}$Fn1-$^{9}$Fn1 module pair. The technique was invaluable for following the collagen peptide binding to $^{8}$Fn1-$^{9}$Fn1 at atomic resolution and for examining the fold of the various IGD mutants.

New insight into the properties of the gelatin binding domain was gained through two techniques. Analytical gelatin affinity chromatography enabled the gelatin binding properties of different length fragments to be explored, the role of the sugar chains on $^{2}$Fn2 and $^{8}$Fn1 to be characterised, and the contribution of each of the modules and module components in binding to be analysed. The cell migration assay was used to explore the bioactive IGD motif in the context of the folded protein and the relative potency of each of the separate fragments assessed.

## 6.2. Collagen binding *in vivo*

Collagen is notoriously difficult to work with as under physiological conditions triple helical molecules spontaneously self associate to form insoluble fibrils. Gelatin is more widely used as an experimental tool to further our understanding of the gelatin binding domain of FN. The binding of the gelatin binding domain to collagen has

been reasonably well characterised *in vitro* but the physiological relevance of the interaction *in vivo* has often been questioned. Could the binding simply be fortuitous, providing a useful tool for purification and isolation of FN from other cell components, but serve no active purpose in the context of a cell? Although the studies here have not primarily addressed this question, the NMR based study with the collagen peptides and $^{8}$Fn1-$^{9}$Fn1 showed a direct and reasonably tight binding interaction between the two. The recent study indicated that lethal mutations in major ligand binding regions of collagen I may impede interactions with FN, causing osteogenesis imperfecta (OI), which shows support for an interaction (Marini et al., 2007). The tightness of binding to only two peptides (of the four used in this study) showed specificity and presented good evidence that an interaction would occur if the two came together in the extracellular matrix.

### 6.2.1. Is collagen the ligand for fibronectin?

Could the ligand for the gelatin binding domain be something other than collagen/gelatin? The gelatin coupled to affinity columns was almost certainly not pure, as the source of the gelatin was biological (calf skin), and may have contained material other than denatured collagen. The gelatin could have been contaminated with another yet undetermined ligand for FN. The Ingham group non-specifically digested calf skin type I collagen (Sigma) (Ingham et al., 2002), and although the N-terminal sequencing suggested the material binding to FN was collagen, sequence alignments of these N-termini flagged up several potential FN binding partners other than collagen. Reassuringly, the BLASTp search of one of their suggested FN binding partners (GPRGFOGSO selected at random from the 14 published sequences) matched type I collagen, but this peptide also had a high score match to other human

proteins: adiponectin precursor protein (adipocyte, C1q and collagen domain-containing protein), scavenger receptor class A member 3 (cellular stress response gene protein), and collagen alpha-3(VI) chain precursor (a different member of the collagen family), to name but a few. Being able to confirm a specific binding interaction by NMR was a key step in reinforcing that the $^{8}$Fn1-$^{9}$Fn1 module pair and collagen peptide are in fact interacting ligands although this does not rule out other binding partners.

The result of the BLASTp search suggested that the peptide is not exclusively found in collagen and there may be an element of competition in binding to FN. The sequence alignment of "peptide 2" (used in the NMR titration) matched similar proteins to the Ingham peptide, indicating that other collagen-like proteins also contain a similar sequence. These proteins could be present as impurities in the gelatin column, implying that FN binding to the column cannot solely be attributed to collagen. Therefore the gelatin affinity column remains a good tool for investigating a binding interaction, and NMR confirmed that at least some of the contribution is from gelatin, but due to the lack of homogeneity in the column there remains uncertainty over the number of different ligands competing for the gelatin binding domain of FN.

### 6.2.2. Orientation of binding

Both the Quad and $^{8}$Fn1-$^{9}$Fn1 showed affinity for the same two peptides in the NMR titration (Quad experiments conducted by Dr. I. Vakonakis). The NMR titration was not configured to assess which residues of the peptide were bound, which allows for

the possibility that they could bind simultaneously, with one showing a higher affinity to an upstream sequence and the other to a downstream sequence. There is ample space on the 20mer synthetic peptide for the Quad to bind to one end and the $^{8}$Fn1-$^{9}$Fn1 to the other (Figure 6:1). Any steric restraints imposed through simultaneous binding could be further investigated using a labelled peptide and unlabelled protein to confirm which residues of the peptide are bound by the Quad and $^{8}$Fn1-$^{9}$Fn1.

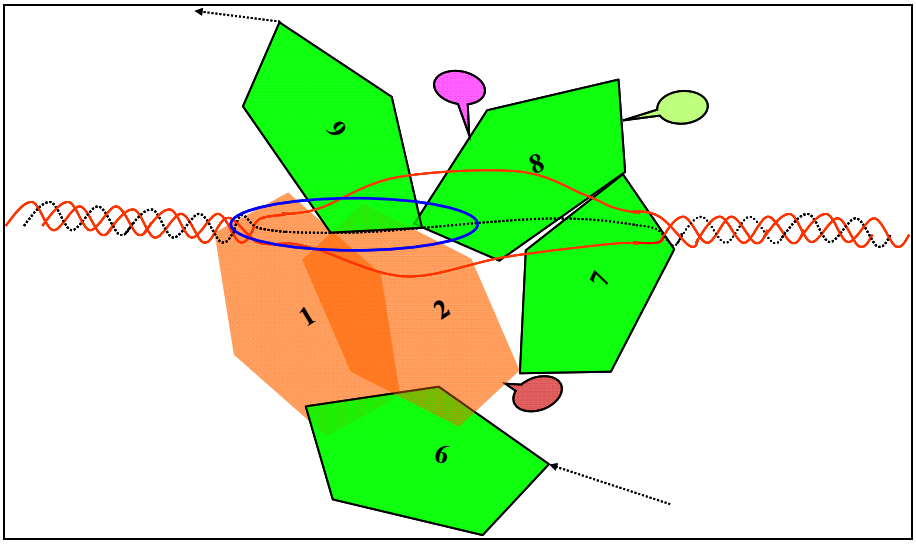



**Figure 6:1** Possible binding orientation of the FN fragment to the same strand of the same collagen triple helix (black dotted line). The Quad could bind to one part of the collagen strand while the $^{8}$Fn1-$^{9}$Fn1 fragment binds to the adjacent part of the same collagen strand (circled in blue). Arrows indicate the position of the FN modules preceding and following the gelatin binding domain. Fn1 modules = green, Fn2 modules = orange, bubbles = high mannose sugar chains, collagen helix; non-binding strands = red, binding strand = dotted black.

An alternative binding mode would be to bind to the same sequence in *different* strands of collagen. The triple helical nature of collagen would be ideally suited to this as FN could bind across the molecule, perpendicular to the amino acid backbone, where the slight stagger in the helix would facilitate and promote FN binding to the two identical α1(I) chains. A third model would suggest that the FN fragment binds to *different* peptide 2 (or additional?) sequences in *different* collagen molecules within the collagen fibril.

## 6.3. The role of the fibronectin IGD motif

Sequence alignments of modules from FN showed that the IGD motif is conserved in four out of nine N-terminal Fn1 modules and also in Fn1 modules from other species. The high level of conservation had already prompted investigation with synthetic IGD peptides (Schor et al., 1999), and the aim here was to build on the existing work, using the same cell migration assay. The scale of increased potency of the IGD motif in the context of the folded module pair was completely unexpected. The additional $10^6$ times enhancement of the full domain over the module pair suggested more than a simple doubling in the number of IGD motifs. The downstream signalling response was only fully activated in the presence of an exposed IGD motif and occurred with most potency in the context of the fully folded molecule. This dramatic rise in potency strongly suggests that an amplification process must be activated once a threshold signalling level is reached. This effect is seen with integrin clustering (Fernandez et al., 1998) which could imply that integrins are involved in direct binding to the IGD ligand or indirectly as a downstream effector.

How this motif causes cells to migrate into a collagen matrix and the identity of the IGD "receptor" remains to be investigated. The IGD motif may bind directly to a membrane bound receptor (in the same way as the RGD motif of FN binds to the integrin α5β1), or equally possible, may bind to an unidentified extracellular adapter protein. The IGD pathway has been suggested to involve integrin α5β1, integrin αVβ3, and downstream effectors focal adhesion kinase and PI3 (Ellis et al., 2007). Somehow the "receptor" must feed into this pathway. The dramatic enhancement seen in the cell migration assays could involve a two step mechanism where the FN

molecule must initially bind another molecule (collagen?) before binding the "activator" ligand.

## 6.4. Cooperativity within the gelatin binding domain

In a dissection approach, investigating individual modules in minute detail can often take priority over understanding the function of the complete domain. However, by building up from a single pair of modules, information about the whole domain can be pieced together. The gelatin binding domain appears to function cooperatively both in collagen binding and in the presentation of the IGD motif. Learning more about how each of the modules contributes to collagen binding can indirectly lead to information about how the modules cooperate and their spatial arrangement with respect to one another. This information can then aid the interpretation of structural data, and with a structural model, the biology of the domain can be better understood. For example, the NMR data from the bound peptide suggested a model to describe how the $^{8}$Fn1-$^{9}$Fn1 modules cooperate to bind the collagen peptide. The gelatin affinity chromatography was then able to expand the model by suggesting a spatial arrangement of $^{7}$Fn1 with respect to $^{8}$Fn1-$^{9}$Fn1. The addition of $^{7}$Fn1 caused partial reduction (but not loss) of gelatin binding affinity in the $^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1 construct, suggesting that the binding face of $^{8}$Fn1 was inhibited by the presence of $^{7}$Fn1. This suggested that the IGD motif of $^{9}$Fn1 is unlikely to be obscured by $^{7}$Fn1, and suggested that $^{7}$Fn1's own IGD motif is accessible in $^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1. The obtained secondary structure information for $^{8}$Fn1-$^{9}$Fn1 confirmed the accessibility of the IGD loop and the similarity of this fold to other Fn1 module pairs; it also allowed the structure of $^{8}$Fn1-$^{9}$Fn1 to be modelled. Knowledge of the position of the IGD motif in

conjunction with the cell migration data allows a more thorough interpretation of the data and therefore the biology of the domain is better understood.

### 6.4.1. Future work

The two threads of this study could both be taken on to better understand collagen binding and IGD signalling. The NMR structure of the $^{8}$Fn1-$^{9}$Fn1 module pair in complex with the collagen "peptide 2" would be an initial step in defining the orientation and exact binding interface on the FN module. The collagen peptide could be made longer (or even triple helical) and binding to $^{7}$Fn1-$^{8}$Fn1-$^{9}$Fn1, the Quad, or even the Hex could be observed by NMR. NMR may allow an estimate of the binding constant, and with enough material, this binding constant could be confirmed with isothermal titration calorimetry. The ultimate goal would be to crystallise the Hex with bound ligand, although this has so far proved intractable.

With respect to the bioactive IGD motif, the natural ligand on the cell surface remains undefined. The $^{8}$Fn1-$^{9}$Fn1 module pair could be used to "fish" for the ligand from the membrane fraction of cell lysate. The identification of a ligand (cell surface receptor or soluble extracellular matrix protein) would lead to a whole new catalogue of experiments.

Numerous scientists have spent hundreds of hours investigating FN and its gelatin binding domain, and it is only in the past few years that all the data has really started to make sense. The dissection approach and structural study of small fragments of FN has developed the big picture of how this domain is assembled and how the individual

components contribute to function. The addition of the IGD motif data has allowed another function to be attributed to this domain in addition to the collagen binding role pioneered by many earlier scientists.

## 6.5. References


Ellis, I.R., Schor, A.M. and Schor, S.L. (2007) EGF AND TGF-alpha motogenic activities are mediated by the EGF receptor via distinct matrix-dependent mechanisms. *Exp Cell Res*, **313**, 732-741.

Fernandez, C., Clark, K., Burrows, L., Schofield, N.R. and Humphries, M.J. (1998) Regulation of the extracellular ligand binding activity of integrins. *Front Biosci*, **3**, d684-700.

Ingham, K.C., Brew, S.A. and Migliorini, M. (2002) Type I collagen contains at least 14 cryptic fibronectin binding sites of similar affinity. *Arch Biochem Biophys*, **407**, 217-223.

Marini, J.C., Forlino, A., Cabral, W.A., Barnes, A.M., San Antonio, J.D., Milgrom, S., Hyland, J.C., Körkkö, J., Prockop, D.J., De Paepe, A., Coucke, P., Symoens, S., Glorieux, F.H., Roughley, P.J., Lund, A.M., Kuurila-Svahn, K., Hartikka, H., Cohn, D.H., Krakow, D., Mottes, M., Schwarze, U., Chen, D., Yang, K., Kuslich, C., Troendle, J., Dalgleish, R. and Byers, P.H. (2007) Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum Mutat*, **28,** 209-21.

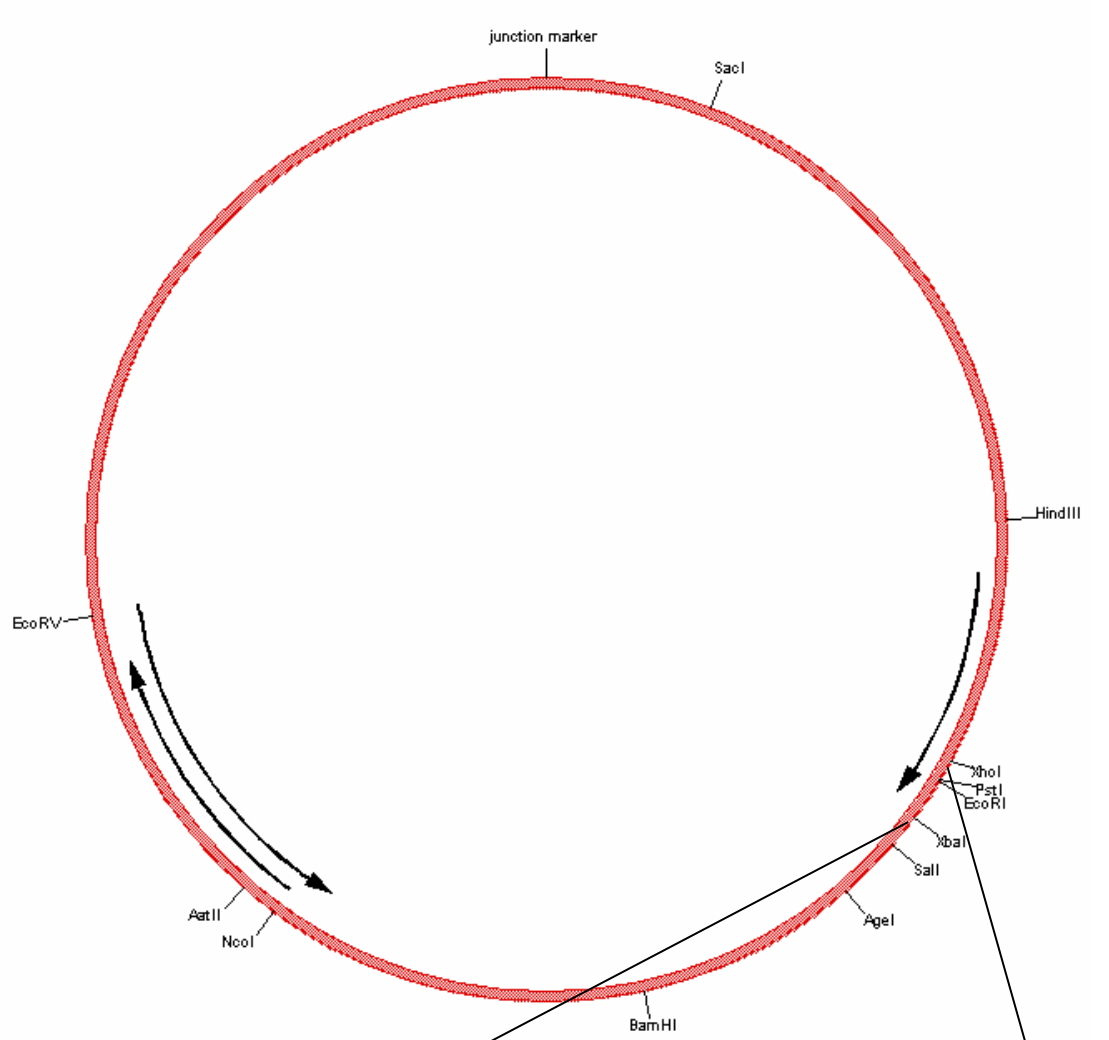
Schor, S.L., Ellis, I., Banyard, J. and Schor, A.M. (1999) Motogenic activity of IGD-containing synthetic peptides. *J Cell Sci*, **112 ( Pt 22)**, 3879-3888.

# pPICZλ vector sequence

Original pPICZα B vector with pLITMUS insert (bold) flanked by the *EcoRI* and *XbaI* insertion sites (underlined)

AGATCTAACATCCAAAGACGAAAGGTTGAATGAAACCTTTTTGCCATCCGACATCCACAGGTCCATTCT
CACACATAAGTGCCAAACGCAACAGGAGGGGATACACTAGCAGCAGACCGTTGCAAACGCAGGACCTCC
ACTCCTCTTCTCCTCAACACCCACTTTTGCCATCGAAAAACCAGCCCAGTTATTGGGCTTGATTGGAGC
TCGCTCATTCCAATTCCTTCTATTAGGCTACTAACACCATGACTTTATTAGCCTGTCTATCCTGGCCCC
CCTGGCGAGGTTCATGTTTGTTTATTTCCGAATGCAACAAGCTCCGCATTACACCCGAACATCACTCCA
GATGAGGGCTTTCTGAGTGTGGGGTCAAATAGTTTCATGTTCCCCAAATGGCCCAAAACTGACAGTTTA
AACGCTGTCTTGGAACCTAATATGACAAAAGCGTGATCTCATCCAAGATGAACTAAGTTTGGTTCGTTG
AAATGCTAACGGCCAGTTGGTCAAAAAGAAACTTCCAAAAGTCGGCATACCGTTTGTCTTGTTTGGTAT
TGATTGACGAATGCTCAAAAATAATCTCATTAATGCTTAGCGCAGTCTCTCTATCGCTTCTGAACCCCG
GTGCACCTGTGCCGAAACGCAAATGGGGAAACACCCGCTTTTTGGATGATTATGCATTGTCTCCACATT
GTATGCTTCCAAGATTCTGGTGGGAATACTGCTGATAGCCTAACGTTCATGATCAAAATTTAACTGTTC
TAACCCCTACTTGACAGCAATATATAAACAGAAGGAAGCTGCCCTGTCTTAAACCTTTTTTTTTATCAT
CATTATTAGCTTACTTTCATAATTGCGACTGGTTCCAATTGACAAGCTTTTGATTTTAACGACTTTTAA
CGACAACTTGAGAAGATCAAAAAACAACTAATTATTCGAAACGATGAGATTTCCTTCAATTTTTACTGC
TGTTTTATTCGCAGCATCCTCCGCATTAGCTGCTCCAGTCAACACTACAACAGAAGATGAAACGGCACA
AATTCCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATT
TTCCAACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGA
AGGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCTGCAGG**AATTCCTGCAGGATATCTGGATCCACGAAG**
**CTTCCCATGGTGACGTCACCGGTT**CTAGAACAAAAACTCATCTCAGAAGAGGATCTGAATAGCGCCGTC
GACCATCATCATCATCATTGAGTTTGTAGCCTTAGACATGACTGTTCCTCAGTTCAAGTTGGGCAC
TTACGAGAAGACCGGTCTTGCTAGATTCTAATCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGC
AGGCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTTTGTTTC
TTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCTGATGAATATCTTGTGGTAGGGGTTTGGG
AAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACTCCTCTTCAGAGTACAGAAGATTAAGTGA
GACCTTCGTTTGTGCGGATCCCCCACACACCATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCC
AGATTTTCTCGGACTCCGCGCATCGCCGTACCACTTCAAAACACCCAAGCACAGCATACTAAATTTTCC
CTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCGTACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCT
CGTTTCTTTTTCTTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTGAAATTTTTTT
TTTTAGTTTTTTTCTCTTTCAGTGACCTCCATTGATATTTAAGTTAATAAACGGTCTTCAATTTCTCAA
GTTTCAGTTTCATTTTTCTTGTTCTATTACAACTTTTTTTACTTCTTGTTCATTAGAAAGAAAGCATAG
CAATCTAATCTAAGGGGCGGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGAC
AAGGTGAGGAACTAAACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCC
GGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGT
GTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTG
GCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTC
CGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGC
GACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTCCGACGGCGGCCCACG
GGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTCGATATCATGTAATTAGTTATGTCACGCTT
ACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGT
CCCTATTTATTTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTT
TCTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTC
GAAGGCTTTAATTTGCAAGCTGGAGACCAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAA
AAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGT
GCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGT
GCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGT
AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGC
TCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG
TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTT
GATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATC

**pPICZλ:** Vector map




**pPICZλ:** MCS



# Gelatin binding domain of fibronectin

Protein sequence; residue numbers are of mature FN (upper, bold)
DNA sequence (lower)

**[6]Fn1;** 274 - 323

```
     Y   G   H   C   V   T   D   S   G   V   V   Y   S   V   G   M
|TAT|GGC|CAC|TGT|GTC|ACA|GAC|AGT|GGT|GTG|GTC|TAC|TCT|GTG|GGG|ATG

  Q   W   L   K   T   Q   G   N   K   Q   M   L   C   T   C   L   G
|CAG|TGG|CTG|AAG|ACA|CAA|GGA|AAT|AAG|CAA|ATG|CTT|TGC|ACG|TGC|CTG|GGC

  N   G   V   S   C   Q   E   T   A   V   T   Q   T   Y   G   G   N
|AAC|GGA|GTC|AGC|TGC|CAA|GAG|ACA|GCT|GTA|ACC|CAG|ACT|TAC|GGT|GGC|AAC
```

**[1]Fn2;** 324 - 383

```
     S   N   G   E   P   C   V   L   P   F   T   Y   N   G   R   T
|TCA|AAT|GGA|GAG|CCA|TGT|GTC|TTA|CCA|TTC|ACC|TAC|AAT|GGC|AGG|ACG

  F   Y   S   C   T   T   E   G   R   Q   D   G   H   L   W   C   S
|TTC|TAC|TCC|TGC|ACC|ACA|GAA|GGG|CGA|CAG|GAC|GGA|CAT|CTT|TGG|TGC|AGC

  T   T   S   N   Y   E   Q   D   Q   K   Y   S   F   C   T   D   H
|ACA|ACT|TCG|AAT|TAT|GAG|CAG|GAC|CAG|AAA|TAC|TCT|TTC|TGC|ACA|GAC|CAC

  T   V   L   V   Q   T   Q   G   G   N
|ACT|GTT|TTG|GTT|CAG|ACA|CGA|GGA|GGA|AAT
```

**[2]Fn2;** 384 - 433

```
     S   N   G   A   L   C   H   F   P   F   L   Y   N   N   H   N
|TCC|AAT|GGT|GCC|TTG|TGC|CAC|TTC|CCC|TTC|CTA|TAC|AAC|AAC|CAC|AAT

  Y   T   D   C   T   S   E   G   R   R   D   N   M   K   W   C   G
|TAC|ACT|GAT|TGC|ACT|TCT|GAG|GGC|AGA|AGA|GAC|AAC|ATG|AAG|TGG|TGT|GGG

  T   T   Q   N   Y   D   A   D   Q   K   F   G   F   C   P   M   A
|ACC|ACA|CAG|AAC|TAT|GAT|GCC|GAC|CAG|AAG|TTT|GGG|TTC|TGC|CCC|ATG|GCT
```

**[7]Fn;** 434 - 484

**A H E E I C T T N E G V M Y R I**

|GCC|CAC|GAG|GAA|ATC|TGC|ACA|ACC|AAT|GAA|GGG|GTC|ATG|TAC|CGC|ATT

**G D Q W D K Q H D M G H M M R C T**

|GGA|GAT|CAG|TGG|GAT|AAG|CAG|CAT|GAC|ATG|GGT|CAC|ATG|ATG|AGG|TGC|ACG

**C V G N G R G E W T C I A Y S Q L**

|TGT|GTT|GGG|AAT|GGT|CGT|GGG|GAA|TGG|ACA|TGC|ATT|GCC|TAC|TCG|CAG|CTT

**R**

|CGA

**[8]Fn1*;** 485 - 527

**D Q C I V D D I T Y N V N D T F**

|GAT|CAG|TGC|ATT|GTT|GAT|GAC|ATC|ACT|TAC|AAT|GTG|AAC|GAC|ACA|TTC

**H K K* H E E G H M L N C T C F G Q**

|CAC|AAG|AAG|CAT|GAA|GAG|GGG|CAC|ATG|CTG|AAC|TGT|ACA|TGC|TTC|GGT|CAG

**G R G R W K C D P V**

|GGT|CGG|GGC|AGG|TGG|AAG|TGT|GAT|CCC|GTC

**[9]Fn1;** 528 - 577

**D Q C Q D S E T G T F Y Q I G D**

|GAC|CAA|TGC|CAG|GAT|TCA|GAG|ACT|GGG|ACG|TTT|TAT|CAA|ATT|GGA|GAT

**S W E K Y V H G V R Y Q C Y C Y G**

|TCA|TGG|GAG|AAG|TAT|GTG|CAT|GGT|GTC|AGA|TAC|CAG|TGC|TAC|TGC|TAT|GGC

**R G I G E W H C Q P L Q T Y P S S**

|CGT|GGC|ATT|GGG|GAG|TGG|CAT|TGC|CAA|CCT|TTA|CAG|ACC|TAT|CCA|AGC|TCA

# Backbone NH and Hα assignments for $^{8}$Fn1*-$^{9}$Fn1

(Deposited in the BioMagResBank, accession number 15447)

- The original sequence entered:
  DQCIVDDITYNVNDTFHKRHEEGHMLNCTCFGQGRGRWKCDPVDQC
  QDSETGTFYQIGDSWEKYVHGVRYQCYCYGRGIGEWHCQPLQ

- Chemical shift ambiguity code definitions:
  1: Unique (geminal atoms and geminal methyl groups with identical chemical shifts are assumed to be assigned to stereospecific atoms)
  2: Ambiguity of geminal atoms or geminal methyl proton groups

| ID | Fibro index ID | αα ID | Atom ID | Atom type | Chem shift value | C.S. value error | C.S. ambig code |
|---|---|---|---|---|---|---|---|
| 1 | 486 | GLN | HE21 | H | 6.835 | 0.01 | 2 |
| 2 | 486 | GLN | HE22 | H | 7.43 | 0.01 | 2 |
| 3 | 486 | GLN | NE2 | N | 112.352 | 0.08 | 1 |
| 4 | 487 | CYS | H | H | 8.895 | 0.01 | 1 |
| 5 | 487 | CYS | N | N | 116.012 | 0.08 | 1 |
| 6 | 488 | ILE | H | H | 8.913 | 0.01 | 1 |
| 7 | 488 | ILE | N | N | 125.371 | 0.08 | 1 |
| 8 | 489 | VAL | H | H | 8.587 | 0.01 | 1 |
| 9 | 489 | VAL | N | N | 127.966 | 0.08 | 1 |
| 10 | 490 | ASP | H | H | 9.14 | 0.01 | 1 |
| 11 | 490 | ASP | N | N | 129.131 | 0.08 | 1 |
| 12 | 491 | ASP | H | H | 8.249 | 0.01 | 1 |
| 13 | 491 | ASP | N | N | 109.863 | 0.08 | 1 |
| 14 | 492 | ILE | H | H | 7.727 | 0.01 | 1 |
| 15 | 492 | ILE | N | N | 123.703 | 0.08 | 1 |
| 16 | 493 | THR | H | H | 7.79 | 0.01 | 1 |
| 17 | 493 | THR | N | N | 120.316 | 0.08 | 1 |
| 18 | 494 | TYR | H | H | 9.197 | 0.01 | 1 |
| 19 | 494 | TYR | N | N | 125.774 | 0.08 | 1 |
| 20 | 495 | ASN | H | H | 9.1 | 0.01 | 1 |
| 21 | 495 | ASN | HD21 | H | 7.029 | 0.01 | 2 |
| 22 | 495 | ASN | HD22 | H | 8.09 | 0.01 | 2 |
| 23 | 495 | ASN | N | N | 119.357 | 0.08 | 1 |
| 24 | 495 | ASN | ND2 | N | 115.793 | 0.08 | 1 |
| 25 | 496 | VAL | H | H | 8.531 | 0.01 | 1 |
| 26 | 496 | VAL | N | N | 120.709 | 0.08 | 1 |
| 27 | 497 | ASN | H | H | 9.178 | 0.01 | 1 |
| 28 | 497 | ASN | HD21 | H | 7.679 | 0.01 | 2 |
| 29 | 497 | ASN | HD22 | H | 8.966 | 0.01 | 2 |
| 30 | 497 | ASN | N | N | 119.455 | 0.08 | 1 |
| 31 | 497 | ASN | ND2 | N | 113.506 | 0.08 | 1 |
| 32 | 498 | ASP | H | H | 8.305 | 0.01 | 1 |
| 33 | 498 | ASP | N | N | 121.377 | 0.08 | 1 |
| 34 | 499 | THR | H | H | 8.112 | 0.01 | 1 |

| ID | Fibro index ID | αα ID | Atom ID | Atom type | Chem shift value | C.S. value error | C.S. ambig code |
|---|---|---|---|---|---|---|---|
| 35 | 499 | THR | N | N | 109.947 | 0.08 | 1 |
| 36 | 500 | PHE | H | H | 7.839 | 0.01 | 1 |
| 37 | 500 | PHE | N | N | 115.987 | 0.08 | 1 |
| 38 | 501 | HIS | H | H | 8.747 | 0.01 | 1 |
| 39 | 501 | HIS | N | N | 119.168 | 0.08 | 1 |
| 40 | 502 | LYS | H | H | 8.237 | 0.01 | 1 |
| 41 | 502 | LYS | N | N | 121.242 | 0.08 | 1 |
| 42 | 503 | ARG | H | H | 8.713 | 0.01 | 1 |
| 43 | 503 | ARG | N | N | 125.43 | 0.08 | 1 |
| 44 | 504 | HIS | H | H | 9.065 | 0.01 | 1 |
| 45 | 504 | HIS | N | N | 128.416 | 0.08 | 1 |
| 46 | 505 | GLU | H | H | 8.561 | 0.01 | 1 |
| 47 | 505 | GLU | N | N | 129.745 | 0.08 | 1 |
| 48 | 507 | GLY | H | H | 8.127 | 0.01 | 1 |
| 49 | 507 | GLY | N | N | 105.628 | 0.08 | 1 |
| 50 | 508 | HIS | H | H | 7.438 | 0.01 | 1 |
| 51 | 508 | HIS | N | N | 117.461 | 0.08 | 1 |
| 52 | 509 | MET | H | H | 8.887 | 0.01 | 1 |
| 53 | 509 | MET | N | N | 119.157 | 0.08 | 1 |
| 54 | 510 | LEU | H | H | 8.877 | 0.01 | 1 |
| 55 | 510 | LEU | N | N | 122.899 | 0.08 | 1 |
| 56 | 511 | ASN | H | H | 9.461 | 0.01 | 1 |
| 57 | 511 | ASN | HD22 | H | 8.464 | 0.01 | 1 |
| 58 | 511 | ASN | N | N | 119.359 | 0.08 | 1 |
| 59 | 511 | ASN | ND2 | N | 133.86 | 0.08 | 1 |
| 60 | 512 | CYS | H | H | 8.438 | 0.01 | 1 |
| 61 | 512 | CYS | N | N | 121.184 | 0.08 | 1 |
| 62 | 513 | THR | H | H | 8.297 | 0.01 | 1 |
| 63 | 513 | THR | N | N | 115.065 | 0.08 | 1 |
| 64 | 514 | CYS | H | H | 8.532 | 0.01 | 1 |
| 65 | 514 | CYS | N | N | 122.223 | 0.08 | 1 |
| 66 | 515 | PHE | H | H | 9.116 | 0.01 | 1 |
| 67 | 515 | PHE | N | N | 130.045 | 0.08 | 1 |
| 68 | 516 | GLY | H | H | 9.162 | 0.01 | 1 |
| 69 | 516 | GLY | N | N | 107.566 | 0.08 | 1 |
| 70 | 517 | GLN | H | H | 8.821 | 0.01 | 1 |
| 71 | 517 | GLN | HE21 | H | 6.693 | 0.01 | 2 |
| 72 | 517 | GLN | HE22 | H | 7.571 | 0.01 | 2 |
| 73 | 517 | GLN | N | N | 123.228 | 0.08 | 1 |
| 74 | 517 | GLN | NE2 | N | 111.157 | 0.08 | 1 |
| 75 | 518 | GLY | H | H | 9.13 | 0.01 | 1 |
| 76 | 518 | GLY | N | N | 108.548 | 0.08 | 1 |
| 77 | 519 | ARG | H | H | 7.549 | 0.01 | 1 |
| 78 | 519 | ARG | N | N | 118.099 | 0.08 | 1 |
| 79 | 520 | GLY | H | H | 9.779 | 0.01 | 1 |
| 80 | 520 | GLY | N | N | 116.078 | 0.08 | 1 |
| 81 | 521 | ARG | H | H | 8.853 | 0.01 | 1 |
| 82 | 521 | ARG | N | N | 118.809 | 0.08 | 1 |

| ID | Fibro index ID | αα ID | Atom ID | Atom type | Chem shift value | C.S. value error | C.S. ambig code |
|---|---|---|---|---|---|---|---|
| 83 | 522 | TRP | H | H | 8.456 | 0.01 | 1 |
| 84 | 522 | TRP | HE1 | H | 10.379 | 0.01 | 1 |
| 85 | 522 | TRP | N | N | 124.041 | 0.08 | 1 |
| 86 | 522 | TRP | NE1 | N | 131.071 | 0.08 | 1 |
| 87 | 523 | LYS | H | H | 8.384 | 0.01 | 1 |
| 88 | 523 | LYS | N | N | 118.505 | 0.08 | 1 |
| 89 | 524 | CYS | H | H | 8.816 | 0.01 | 1 |
| 90 | 524 | CYS | N | N | 120.128 | 0.08 | 1 |
| 91 | 525 | ASP | H | H | 9.079 | 0.01 | 1 |
| 92 | 525 | ASP | N | N | 123.103 | 0.08 | 1 |
| 93 | 527 | VAL | H | H | 8.033 | 0.01 | 1 |
| 94 | 527 | VAL | N | N | 122.241 | 0.08 | 1 |
| 95 | 528 | ASP | H | H | 8.622 | 0.01 | 1 |
| 96 | 528 | ASP | N | N | 126.197 | 0.08 | 1 |
| 97 | 529 | GLN | H | H | 7.828 | 0.01 | 1 |
| 98 | 529 | GLN | N | N | 120.516 | 0.08 | 1 |
| 99 | 530 | CYS | H | H | 9.181 | 0.01 | 1 |
| 100 | 530 | CYS | N | N | 116.499 | 0.08 | 1 |
| 101 | 531 | GLN | H | H | 9.037 | 0.01 | 1 |
| 102 | 531 | GLN | HE21 | H | 6.699 | 0.01 | 2 |
| 103 | 531 | GLN | HE22 | H | 6.772 | 0.01 | 2 |
| 104 | 531 | GLN | N | N | 123.459 | 0.08 | 1 |
| 105 | 531 | GLN | NE2 | N | 112.416 | 0.08 | 1 |
| 106 | 532 | ASP | H | H | 8.009 | 0.01 | 1 |
| 107 | 532 | ASP | N | N | 126.506 | 0.08 | 1 |
| 108 | 533 | SER | H | H | 7.775 | 0.01 | 1 |
| 109 | 533 | SER | N | N | 121.895 | 0.08 | 1 |
| 110 | 534 | GLU | H | H | 8.719 | 0.01 | 1 |
| 111 | 534 | GLU | N | N | 121.825 | 0.08 | 1 |
| 112 | 535 | THR | H | H | 7.626 | 0.01 | 1 |
| 113 | 535 | THR | N | N | 106.008 | 0.08 | 1 |
| 114 | 536 | GLY | H | H | 8.087 | 0.01 | 1 |
| 115 | 536 | GLY | N | N | 111.473 | 0.08 | 1 |
| 116 | 537 | THR | H | H | 7.428 | 0.01 | 1 |
| 117 | 537 | THR | N | N | 118.798 | 0.08 | 1 |
| 118 | 538 | PHE | H | H | 8.066 | 0.01 | 1 |
| 119 | 538 | PHE | N | N | 123.132 | 0.08 | 1 |
| 120 | 539 | TYR | H | H | 8.987 | 0.01 | 1 |
| 121 | 539 | TYR | N | N | 121.052 | 0.08 | 1 |
| 122 | 540 | GLN | H | H | 9.099 | 0.01 | 1 |
| 123 | 540 | GLN | N | N | 119.37 | 0.08 | 1 |
| 124 | 541 | ILE | H | H | 8.749 | 0.01 | 1 |
| 125 | 541 | ILE | N | N | 121.34 | 0.08 | 1 |
| 126 | 542 | GLY | H | H | 8.783 | 0.01 | 1 |
| 127 | 542 | GLY | N | N | 116.657 | 0.08 | 1 |
| 128 | 543 | ASP | H | H | 8.335 | 0.01 | 1 |
| 129 | 543 | ASP | N | N | 121.822 | 0.08 | 1 |
| 130 | 544 | SER | H | H | 8.227 | 0.01 | 1 |

| ID | Fibro index ID | αα ID | Atom ID | Atom type | Chem shift value | C.S. value error | C.S. ambig code |
|---|---|---|---|---|---|---|---|
| 131 | 544 | SER | N | N | 111.734 | 0.08 | 1 |
| 132 | 545 | TRP | H | H | 8.278 | 0.01 | 1 |
| 133 | 545 | TRP | HE1 | H | 9.599 | 0.01 | 1 |
| 134 | 545 | TRP | N | N | 122.135 | 0.08 | 1 |
| 135 | 545 | TRP | NE1 | N | 128.367 | 0.08 | 1 |
| 136 | 546 | GLU | H | H | 8.502 | 0.01 | 1 |
| 137 | 546 | GLU | N | N | 119.607 | 0.08 | 1 |
| 138 | 547 | LYS | H | H | 8.419 | 0.01 | 1 |
| 139 | 547 | LYS | N | N | 119.877 | 0.08 | 1 |
| 140 | 548 | TYR | H | H | 8.933 | 0.01 | 1 |
| 141 | 548 | TYR | N | N | 122.932 | 0.08 | 1 |
| 142 | 549 | VAL | H | H | 8.973 | 0.01 | 1 |
| 143 | 549 | VAL | N | N | 124.125 | 0.08 | 1 |
| 144 | 551 | GLY | H | H | 8.476 | 0.01 | 1 |
| 145 | 551 | GLY | N | N | 103.068 | 0.08 | 1 |
| 146 | 552 | VAL | H | H | 7.863 | 0.01 | 1 |
| 147 | 552 | VAL | N | N | 122.948 | 0.08 | 1 |
| 148 | 553 | ARG | H | H | 8.529 | 0.01 | 1 |
| 149 | 553 | ARG | N | N | 127.679 | 0.08 | 1 |
| 150 | 554 | TYR | H | H | 9.439 | 0.01 | 1 |
| 151 | 554 | TYR | N | N | 126.212 | 0.08 | 1 |
| 152 | 555 | GLN | H | H | 8.928 | 0.01 | 1 |
| 153 | 555 | GLN | HE21 | H | 6.841 | 0.01 | 1 |
| 154 | 555 | GLN | HE22 | H | 7.393 | 0.01 | 1 |
| 155 | 555 | GLN | N | N | 118.973 | 0.08 | 1 |
| 156 | 555 | GLN | NE2 | N | 111.665 | 0.08 | 1 |
| 157 | 556 | CYS | H | H | 9.194 | 0.01 | 1 |
| 158 | 556 | CYS | N | N | 127.163 | 0.08 | 1 |
| 159 | 557 | TYR | H | H | 8.161 | 0.01 | 1 |
| 160 | 557 | TYR | N | N | 120.429 | 0.08 | 1 |
| 161 | 558 | CYS | H | H | 7.643 | 0.01 | 1 |
| 162 | 558 | CYS | N | N | 121.424 | 0.08 | 1 |
| 163 | 566 | TRP | H | H | 8.997 | 0.01 | 1 |
| 164 | 566 | TRP | HE1 | H | 9.818 | 0.01 | 1 |
| 165 | 566 | TRP | N | N | 128.561 | 0.08 | 1 |
| 166 | 566 | TRP | NE1 | N | 131.267 | 0.08 | 1 |
| 167 | 567 | HIS | H | H | 8.336 | 0.01 | 1 |
| 168 | 567 | HIS | N | N | 113.795 | 0.08 | 1 |
| 169 | 569 | GLN | H | H | 8.996 | 0.01 | 1 |
| 170 | 569 | GLN | N | N | 120.514 | 0.08 | 1 |
| 171 | 571 | LEU | H | H | 8.411 | 0.01 | 1 |
| 172 | 571 | LEU | N | N | 124.999 | 0.08 | 1 |
| 173 | 572 | GLN | H | H | 7.915 | 0.01 | 1 |
| 174 | 572 | GLN | N | N | 126.913 | 0.08 | 1 |