MOLECULAR AND STRUCTURAL CHARACTERISATION OF THE HUMAN FIBRILLIN-1 N-C TERMINAL INTERACTION



David Yadin

Department of Biochemistry

and

Lincoln College

A thesis submitted for the degree of Doctor of Philosophy
at the University of Oxford, Trinity Term 2012

MOLECULAR AND STRUCTURAL CHARACTERISATION OF THE HUMAN FIBRILLIN-1 N-C TERMINAL INTERACTION

David Yadin, Lincoln College and Department of Biochemistry, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy, Trinity Term 2012

ABSTRACT

Fibrillins are modular, disulphide-rich glycoproteins that assemble into microfibrils in the extracellular matrix (ECM). These microfibrils are critical structural elements of many non-elastic and elastic connective tissues. They also regulate the availability of transforming growth factor-β signalling molecules in the ECM. Defects in microfibrils are associated with acquired and inherited connective tissue disorders. In particular, mutations in the human *FBN1* gene, which encodes fibrillin-1, are associated with a spectrum of diseases, including Marfan syndrome (MFS).

One of the proposed initial steps in microfibril assembly is the interaction between the N- and C-terminal regions of fibrillin monomers. The minimal regions of human fibrillin-1 required for an interaction *in vitro* were previously identified: the four N-terminal domains, from the fibrillin unique N-terminal (FUN) domain to the third epidermal growth factor-like (EGF) domain (FUN-EGF3), and the three C-terminal calcium-binding EGF-like (cbEGF) domains (cbEGF41-43). Here, fragments corresponding to these regions were produced and shown to interact in pull-down and surface plasmon resonance assays. In addition, the structure of the FUN-EGF3 fragment was determined using nuclear magnetic resonance spectroscopy. This showed the novel structure of the FUN domain and the interdomain interfaces in this region of fibrillin.

Combining structural and sequence conservation data may help to identify regions of FUN-EGF3 important for binding to cbEGF41-43. Here, the interaction was probed by site-directed mutagenesis. However, substituting individual residues in FUN-EGF3 with alanine did not abrogate binding to cbEGF41-43. Three MFS-associated residue substitutions were also introduced into the FUN-EGF3 fragment. While they did not abolish the interaction with cbEGF41-43, they did cause misfolding. Two of these substitutions, N57D and W71R, also resulted in the defective secretion of a larger N-terminal fragment by fibroblast cells, suggesting a potential mechanism of disease pathogenesis. Although specific residues involved in the N-C interaction were not identified here, the FUN-EGF3 structure will be vital for understanding the molecular surfaces involved in microfibril assembly and growth factor binding.

Acknowledgements

I am very grateful for the generous funding from the Wellcome Trust, which has supported me for the last four years.

Special thanks go to Professors Penny Handford and Christina Redfield for their supervision, guidance and patience. I am very grateful to Penny for allowing me to work on this project, and for being a constant source of new ideas and directions, especially when experiments were not going to plan. I am also very grateful to Christina for teaching me about NMR, training me to use the spectrometers, helping me with data analysis and generally being an unlimited source of information about NMR and protein structure. This work would not have been possible without either of them.

I would particularly like to thank Dr Sacha Jensen, who started the N-C interaction project in the lab and supervised me during my rotation project in my first year. I am extremely grateful for the discussions, the ideas he has given me, the large amount of time he spent proofreading parts of my thesis, as well as being good company in the lab and at conferences.

I am grateful to David Stoddart and Ian Robertson for the work they did on the N-C terminal interaction during their rotation projects, before I started in the lab. Ian has also been a great source of information about fibrillins and LTBPs and interesting conversations. I am also very grateful to Jo McNaught-Davis and Paul Evans for the work they did on the N-C terminal interaction during their Part II projects.

Thanks also go to Dr David Staunton for his assistance with mass spectrometry and SPR.

I am very grateful to everyone in the Handford lab. To Paul for helping me to keep my sanity, to Sarah and Mouli for being good 'neighbours' in the office and to Pat, Philip, Shaoyan, Matt (an honorary member of the lab), Sophie and Laurie for providing a fun working environment. Thanks also to Louise Walport from the Redfield lab for discussions about NMR structure calculations.

Finally, special thanks to my family who have supported me during my time in Oxford, through the ups and downs. And to my friends, who have kept me entertained and distracted from work outside the lab.

Declaration

I declare that this work is entirely my own except where stated in the text.

In addition to instances in the text, assistance with acquisition and processing of some NMR spectra was provided by Dr Christina Redfield.

Abbreviations

1D	One-dimensional
2 x YT	Two times yeast tryptone broth
2D	Two-dimensional
3D	Three-dimensional
AD	Acromicric dysplasia
ADAMTS	A disintegrin and metalloprotease with thrombospondin motif
ADAMTSL	A disintegrin and metalloprotease with thrombospondin motif-like
AMD	Age-related macular degeneration
APS	Ammonium persulphate
A_x	Absorbance of light with x nm wavelength
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
cbEGF	Calcium-binding epidermal growth factor-like (domain)
cbEGF32-34	Fragment comprising calcium-binding epidermal growth factor-like domains 32-34 of fibrillin
cbEGF32-34-BirA	cbEGF32-34 fragment with a C-terminal BirA tag
cbEGF41-43	Fragment comprising calcium-binding epidermal growth factor-like domains 41-43 of fibrillin
cbEGF41-43-BirA	cbEGF41-43 fragment with a C-terminal BirA tag
cbEGF41-43-BirA-bio	Biotinylated cbEGF41-43-BirA fragment
CL	Cutis laxa
COSY	Correlation spectroscopy
CT	Constant time
CTAB	Cetyl trimethyl ammonium bromide
D13OPC	1,2-O-ditridecyl-*sn*-glycero-3-phosphocholine
DHOPC	1,2-dihexyl-*sn*-glycero-3-phosphocholine
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulphoxide
dNTP	Deoxyribonucleotide
DQF	Double-quantum filter
DSL	Delta/Serrate/Lag2 (domain)
DTT	Dithiothreoitol

E. coli	*Escherichia coli*
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor-like (domain)
EGTA	Ethylene glycol tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Emilin	Elastin microfibril interface located protein
FPLC	Fast protein liquid chromatography
FUN	Fibrillin unique N-terminal (domain)
FUN-EGF3	Fragment comprising the fibrillin unique N-terminal and first three EGF-like domains of fibrillin
FUN-EGF3-BirA-bio	Biotinylated FUN-EGF3 fragment with a C-terminal BirA tag
GAG	Glycosaminoglycan
GD	Geleophysic dysplasia
GDF	Growth and differentiation factor
GFP	Green fluorescent protein
Gly-rich	Glycine-rich (region)
HBDB	Hydrogen bond database potential
HEPES	4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid
HHIP	Hedgehog interacting protein
His_6	Hexa-histidine (tag)
His-cbEGF32-34	cbEGF32-34 fragment with an N-terminal His_6 tag
His-cbEGF41-43	cbEGF41-43 fragment with an N-terminal His_6 tag
His-DSL-EGF3	Human Jagged-1 DSL-EGF3 fragment with an N-terminal His_6 tag
His-FUN-EGF3	FUN-EGF3 fragment with an N-terminal His_6 tag
HMQC	Heteronuclear multiple quantum coherence
HPLC	High performance liquid chromatography
HRP	Horse radish peroxidase
HSQC	Heteronuclear single quantum coherence
HEK	Human embryonic kidney (cells)
Hyb	Hybrid (domain)
INEPT	Insensitive nuclei enhanced by polarisation transfer
IPAP	In-phase anti-phase
IPTG	Isopropyl-β-D-1-thiogalactopyranoside

kb	Kilobase
K_d	Dissociation constant
LAP	Latency associated peptide
LLC	Large latent complex
LOX	Lysyl oxidase
LOXL	Lysyl oxidase-like
LTBP	Latent transforming growth factor-β binding protein
MAGP	Microfibril-associated glycoprotein
mAU	Milli absorbance units
MaxEnt	Maximum entropy
MEM	Minimal essential medium
MES	2-(N-Morpholino) ethanesulfonic acid
MFS	Marfan syndrome
MMP	Matrix metalloproteinase
Mw	Molecular weight
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
Nterm-Pro	Fragment comprising the fibrillin N-terminus to the Pro-rich region
PACE	Paired basic amino acid cleaving enzyme
PBS	Phosphate-buffered saline
PBSE	Phosphate-buffered saline with 0.02% (w/v) EDTA
PBS-T	PBS solution with 0.1% (v/v) Triton-X-100
PCR	Polymerase chain reaction
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
pK_a	$-\log_{10}(K_a)$, where K_a is the acid dissociation constant
ppm	Parts per million
Pro-rich	Proline-rich (region)
Rama	Ramachandran-based torsion angle database potential
RCF	Relative centrifugal force
RDC	Residual dipolar coupling
RGD	Asp-Gly-Asp motif
RMSD	Root mean square deviation

Rpm	Revolutions per minute
RU	Response units
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
SSS	Stiff skin syndrome
T25	25 cm^2 tissue culture flask
T75	75 cm^2 tissue culture flask
TEMED	*N*,*N*,*N*′,*N*′-Tetramethylethylenediamine
TB	Transforming growth factor-β binding-like (domain)
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline with 0.05% (v/v) Tween-20
TFA	Trifluoroacetic acid
TGFβ	Transforming growth factor-β
T_m	Melting temperature
TOCSY	Total correlation spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
WIF	Wnt-inhibitor factor
WMS	Weill-Marchesani syndrome
WT	Wildtype

TABLE OF CONTENTS

Abstract II

Acknowledgements III

Declaration IV

Abbreviations V

Chapter 1: Introduction

1.1 Extracellular matrix 2
1.2 Fibrillin microfibrils and elastic fibres 4
1.3 Fibrillin structure 6
1.4 Interdomain packing interactions 12
1.5 Fibrillin evolution 15
1.6 Fibrillin microfibril organisation 15
1.7 Fibrillin microfibril assembly 18
1.7.1 Furin-mediated processing of the N- and C-termini 20
1.7.2 Roles of the N- and C-terminal regions 21
1.8 Functions of ECM components that interact with fibrillin 24
1.8.1 Microfibril structure and assembly 24
1.8.2 Integrins: cell adhesion and role in microfibril assembly 26
1.8.3 Growth factor regulation 26
1.8.4 Elastic fibre assembly 27
1.9 Fibrillinopathies: inherited connective tissue disorders linked to fibrillin 30
1.9.1 Marfan syndrome 30
1.9.2 Other diseases linked to the FBN1 gene 32
1.9.3 Mechanisms of Marfan syndrome pathogenesis 32
1.9.4 Role of TGFβ in disease pathogenesis 35
1.9.5 Structural effects of MFS-causing mutations 36
1.9.6 Marfan syndrome and microfibril assembly 37
1.10 Thesis aim 40

Chapter 2: Materials and methods

2.1 DNA techniques 43
2.1.1 Polymerase chain reaction (PCR) 43
2.1.2 Ethanol precipitation 44

2.1.3 Agarose gel electrophoresis 44
2.1.4 Gel purification 45
2.1.5 Restriction enzyme digestion of DNA 45
2.1.6 DNA Ligation 45
2.1.7 PCR colony screen 46
2.1.8 DNA amplification 46
2.1.9 DNA quantitation 47
2.1.10 Site-directed mutagenesis 47
2.1.11 DNA sequencing 48
2.1.12 Phenol-chloroform extraction 48
2.1.13 Production of competent cells 48
2.1.14 Transformation of competent *E. coli* cells 49
2.2 Protein expression, purification, refolding and characterisation 50
2.2.1 Protein expression in *E. coli* 50
2.2.2 Isotopic labelling for NMR spectroscopy 50
2.2.3 Nickel affinity purification of reduced protein 51
2.2.4 Reverse phase HPLC 52
2.2.5 Prediction of protein properties and concentration measurement by spectrophotometry 53
2.2.7 His_6 tag removal using factor Xa 55
2.2.8 Ion exchange fast protein liquid chromatography (FPLC) 55
2.2.9 Site-specific biotinylation of BirA-tagged proteins 56
2.2.10 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) 56
2.2.11 Mass spectrometry 57
2.3 Limited proteolysis 58
2.4 Pull-down interaction assay 59
2.4.1 Pull-down 59
2.4.2 Western blotting 60
2.5 Surface plasmon resonance (SPR) 61
2.5.1 Surface immobilisation 61
2.5.2 Binding experiments 62
2.6 NMR spectroscopy 62
2.6.1 Sample preparation 63
2.6.2 Data acquisition and processing 63
2.6.2.1 One-dimensional (1D) ^{1}H experiments 63
2.6.2.2 Two-dimensional (2D) homonuclear ^{1}H experiments 64
2.6.2.3 2D heteronuclear experiments 64
2.6.2.4 3D double resonance experiments 66

2.6.2.5 3D triple resonance experiments 66
2.6.3 Interpretation and analysis of NMR spectra 68
2.6.4 Heteronuclear NOE experiment 68
2.6.5 Generation of distance restraints 68
2.6.6 Measurement of J-couplings from HMQC-J spectra 69
2.6.7 TALOS+ torsion angle prediction 69
2.6.8 Identification of slowly-exchanging amide protons 69
2.6.9 Measurement of residual dipolar couplings (RDCs) 70
2.6.9.1 Bicelle preparation 70
2.6.9.2 Monitoring bicelle alignment by 1D ^{2}H spectroscopy 70
2.6.9.3 Data acquisition and analysis 71
2.6.10 Titrations with $CaCl_2$ and $MgCl_2$ 71
2.6.11 Protein titrations 71
2.7 NMR structure calculation 72
2.7.1 Simulated annealing from an extended template 72
2.7.2 Analysis of prochiral swapping calculations 74
2.7.3 Refinement of initial structures 74
2.7.4 Fitting of RDCs to calculated structures 75
2.7.5 Average structure and ensemble RMSD calculations 76
2.7.6 Buried surface area and domain twist/tilt calculations 76
2.8 Fibroblast culture and transfection 77
2.8.1 Maintenance and culture of cells 77
2.8.2 Sub-culture of cells 78
2.8.3 Freezing, thawing and storage of cells 78
2.8.4 Stable transfection 79
2.8.5 Protein secretion assay 79

Chapter 3: Production, characterisation and interaction of human fibrillin-1 N- and C-terminal fragments

3.1 Introduction 82
3.2 Interaction of an immobilised FUN-EGF3 fragment with cbEGF41-43 in solution 84
3.3 Development of a pull-down assay using immobilised cbEGF41-43 86
3.4 Expression, purification and refolding of fibrillin fragments 87
3.4.1 Cloning of constructs encoding cbEGF41-43-BirA and cbEGF32-34-BirA 87
3.4.2 Protein expression, purification and refolding 89
3.4.2.1 Purification and refolding of FUN-EGF3 90
3.4.2.2 SDS-PAGE and mass spectrometry analysis of FUN-EGF3 and His-FUN-EGF3 93

3.4.2.3 Purification and refolding of cbEGF41-43-BirA 94
3.4.2.4 SDS-PAGE and mass spectrometry analysis of cbEGF41-43-BirA 96
3.4.2.5 Purification, refolding and characterisation of cbEGF32-34-BirA 97
3.5 Initial characterisation of FUN-EGF3 fragment by NMR spectroscopy 97
3.6 Characterisation of cbEGF41-43-BirA 100
3.6.1 Limited proteolysis 100
3.6.2 NMR spectroscopy 102
3.7 Interaction of immobilised cbEGF41-43-BirA with His-FUN-EGF3 105
3.7.1 Pull-down assay 105
3.7.2 Surface plasmon resonance-based interaction assay 107
3.8 Discussion 111

Chapter 4: Structure and dynamics of the FUN-EGF3 fragment

4.1 Introduction 115
4.2 Isotopic labelling of FUN-EGF3 for structure determination 117
4.3 Assignment of resonances in FUN-EGF3 117
4.4 Preliminary analysis of backbone dynamics 125
4.5 NOE assignment and restraint generation 127
4.6 Backbone torsion angles 130
4.6.1 Measurement of $^{3}J_{HNH\alpha}$ couplings 130
4.6.2 TALOS+ prediction 132
4.6.3 *Cis vs. trans* proline residues 132
4.7 Initial structure calculation by simulated annealing 136
4.7.1 Iterative NOE assignment 137
4.7.2 Stereospecific assignment using prochiral swapping 138
4.8 Refinement 140
4.8.1 Ramachandran torsion angle database potential 140
4.8.2 Hydrogen bonds 142
4.8.3 Residual dipolar couplings: concepts 144
4.8.4 RDC measurement 145
4.8.5 Use of RDCs in structure refinement 147
4.8.6 Refinement of FUN-EGF3 with RDCs 148
4.9 Final structure ensembles 150
4.9.1 Quality statistics 150
4.9.2 Structure precision 153
4.9.3 Structure validation 156
4.10 Domain structures in the FUN-EGF3 fragment 157

4.11 Interdomain packing 159
4.12 Comparisons with other interdomain interfaces in human fibrillin-1 164
4.13 Comparisons with EGF-EGF pairs in other extracellular proteins 165
4.14 Discussion 168

Chapter 5: Dissection of the fibrillin-1 N-C terminal interaction

5.1 Introduction 172
5.2 Production of FUN-EGF1 and EGF2-EGF3 fragments 174
5.2.1 Cloning of a construct encoding EGF2-EGF3 174
5.2.2 Expression, purification and refolding of FUN-EGF1 174
5.2.3 SDS-PAGE and mass spectrometry analysis of FUN-EGF1 and His-FUN-EGF1 176
5.2.4 Expression, purification and refolding of EGF2-EGF3: effect of a C-terminal BirA tag 177
5.2.5 SDS-PAGE and mass spectrometry analysis of EGF2-EGF3-BirA and His-EGF2-EGF3-BirA 179
5.3 Characterisation of FUN-EGF1 and EGF2-EGF3-BirA by NMR spectroscopy 180
5.4 Interaction of FUN-EGF1 and EGF2-EGF3-BirA fragments with cbEGF41-43-BirA 184
5.4.1 Pull-down assay 184
5.4.2 SPR assay 186
5.5 NMR titration: ^{15}N FUN-EGF3 with unlabelled cbEGF41-43-BirA 188
5.6 Probing the N-C interaction by site-directed mutagenesis of FUN-EGF3 190
5.6.1 Insights from sequence conservation of FUN-EGF3 190
5.6.2 Design of FUN-EGF3 fragments with alanine substitutions 192
5.6.3 Production of FUN-EGF3 fragments with single alanine substitutions 193
5.6.4 Use of 2D ^{1}H NMR spectroscopy in characterisation of FUN-EGF3 proteins 193
5.6.5 Characterisation of FUN-EGF3 fragments with single alanine substitutions 197
5.6.6 Interaction of modified FUN-EGF3 fragments with cbEGF41-43-BirA 200
5.6.6.1 Pull-down assay 200
5.6.6.2 SPR assay 201
5.7 Discussion 204

Chapter 6: Effects of MFS mutations on the N- and C-terminal domains of human fibrillin-1

6.1 Introduction 208
6.2 Insights into disease-linked missense mutations affecting the FUN-EGF3 region of fibrillin-1 209
6.4 Chapter aims 213
6.5 Production and characterisation of FUN-EGF3 fragments with MFS-associated substitutions 214
6.5.1 Site-directed mutagenesis 214
6.5.2 Protein production 214

6.5.3 SDS-PAGE and mass spectrometry analysis 219
6.5.4 Characterisation of N57D and P148S FUN-EGF3 by NMR spectroscopy 221
6.6 Effects of MFS-associated substitutions on the N-C terminal interaction 229
6.7 Effects of substitutions on protein secretion by MSU1.1 cells 230
6.7.1 pKG Nterm-Pro construct 230
6.7.2 Site-directed mutagenesis of pKG Nterm-Pro and transfection into MSU1.1 cells 230
6.7.3 Nterm-Pro secretion by pools of MSU1.1 clones 232
6.8 Production and characterisation of the Y2639N cbEGF41-43-BirA fragment 234
6.8.1 Site-directed mutagenesis 234
6.8.2 Protein production 236
6.8.3 Characterisation by SDS-PAGE and mass spectrometry 236
6.9 Interaction of Y2639N cbEGF41-43-BirA-bio with His-FUN-EGF3 237
6.10 Discussion 238

Chapter 7: Discussion

7.1 Summary 244
7.2 Production, characterisation of N- and C-terminal fibrillin-1 fragments 244
7.3 Interaction of N- and C-terminal fibrillin-1 fragments 246
7.4 Structure of the FUN-EGF3 fragment 247
7.5 Dissection of the N-C terminal interaction 250
7.6 Effects of MFS-associated mutations on the fibrillin-1 N- and C-termini 254
7.7 Implications for microfibril assembly 257
7.8 Implications for microfibril organisation 261
7.9 Implications for interactions of fibrillin-1 with microfibril-associated factors 263
7.10 Conclusions 263

References 265

Appendix

Appendix A: references for disease-linked mutations in *FBN1* affecting the FUN-EGF3 and cbEGF41-43 regions of fibrillin-1 286
Appendix B: oligonucleotide primer sequences 288
Appendix C: protein extinction coefficients and molecular weights 290
Appendix D: purification, refolding and characterisation of negative control fragment cbEGF32-34-BirA 291
Appendix E: ^{1}H, ^{15}N and ^{13}C resonance assignments for FUN-EGF3 294
Appendix F: fibrillin protein sequences used in alignments 299

Chapter 1

Introduction

1.1 Extracellular matrix

Extracellular matrix (ECM) is the medium that surrounds cells in complex multicellular organisms. In animals, ECM comprises a complex mixture of proteins and carbohydrates. Organs and tissues are defined and delimited by the composition of their ECM, as well their characteristic cell types. Specialised cells, predominantly fibroblasts, secrete ECM components and contribute to their assembly. The ECM is responsible for the structural integrity of tissues and for their mechanical properties. However, rather than being a static scaffold for cells, the ECM is dynamic − it is constantly remodelled during development and must be maintained throughout life, including responding to damage. This interplay between cells and their surrounding ECM requires mechanisms for regulation of cell adhesion, migration, signalling and differentiation (Engel *et al.*, 2011).

All types of animal ECM are derived from similar core components. The constituent proteins are usually modular, consisting of arrays of similar domain types (Hynes, 2009), and have extensive post-translational modifications. They can be glycoproteins, with N- and O-linked polysaccharide chains, or proteoglycans such as perlecan, aggrecan and versican, which are linked to glycosaminoglycans (GAGs). GAGs are linear polysaccharides of repeating disaccharides and are often highly negatively charged due to the presence of sulphate and carboxylate groups (Wight *et al.*, 2011). In some cases, such as hyaluronan, they are also present in an unconjugated form (Engel *et al.*, 2011). Collagens make up a significant proportion of proteins in the ECM and are distinguished by a characteristic G-X-Z repeat sequence (where X and Z are frequently proline and hydroxyproline respectively) (Birk *et al.*, 2011). This sequence results in a propensity to forming triple helix structures, with some collagens assembling further to produce long fibrils. Several other proteins assemble into oligomers and some, such as fibrillin and fibronectin, can also form

fibrils. The ability of these proteins to assemble into higher-order structures in the ECM is essential to their functions.

Connective tissues containing ECM can take many different forms, depending on their composition. Basement membranes are important for marking boundaries within and between tissues and can act as permeability barriers. For example, they are found between the epidermis and dermis layers of the skin, and in the kidney glomerulus. These membranes all comprise type IV collagen, glycoproteins laminin and nidogen, and sulphated proteoglycans such as perlecan, which assemble into a dense network (Miner and Mecham, 2011).

Strength and durability are conferred by fibres, the backbones of which are predominantly formed by fibrillar collagens or fibrillin. Collagens form inelastic rope-like structures in tissues such as tendons and ligaments (Engel *et al.*, 2011). Elastic tissues, such as the elastic laminae of blood vessels, heart valves or the skin, contain elastic fibres. A major component of these fibres is elastin, which is assembled from monomeric tropoelastin on a scaffold of fibrillin microfibrils (Yanagisawa and Davis, 2010). Some tissues, such as cartilage, have to resist a significant amount of compression, while allowing deformation. This is achieved by a network of collagen and fibronectin fibres surrounding proteoglycans, which are highly hydrated and so have a 'cushioning' effect (Engel *et al.*, 2011). In bones and teeth, additional strength and hardness results from mineralisation of the tissue, a process that depends on calcium (Engel *et al.*, 2011). Calcium is also important for maintaining the structure of calcium-binding proteins, most notably fibrillin.

1.2 Fibrillin microfibrils and elastic fibres

In 1986, fibrillin-1 was identified as the major component of the 10-12 nm diameter microfibrils in connective tissues using antibody labelling (Sakai *et al.*, 1986). These microfibrils are distinct from the 3-5 nm diameter collagen VI-containing fibrils and are present in a large number of different tissues. Specifically, fibrillin microfibrils occur in proximity to basement membranes like Descemet's membrane in the cornea, in elastic tissues, such as the skin, surrounding an amorphous core of elastin and also in elastin-free fibrils in tissues such as the ciliary zonules of the eye and kidney glomerulus (Figure 1.1A).

When extracted from ECM and viewed by electron or atomic force microscopy, microfibrils have a 'beads-on-a-string' appearance, with a characteristic average repeat distance of 50-60 nm between the beads (Baldock *et al.*, 2001; Cardy and Handford, 1998; Hanssen *et al.*, 1998; Keene *et al.*, 1991; Kielty *et al.*, 1991) (Figure 1.1B). By contrast, monomeric, extended fibrillin-1 has a length of ~150 nm (Sakai *et al.*, 1991). Furthermore, the structure of microfibrils is dependent on calcium. Microfibrils treated with $CaCl_2$, but not $MgCl_2$, adopt a linear conformation, which can be reversibly disrupted by treatment with the calcium-specific chelating agent ethylene glycol tetraacetic acid (EGTA) (Cardy and Handford, 1998; Kielty and Shuttleworth, 1993).

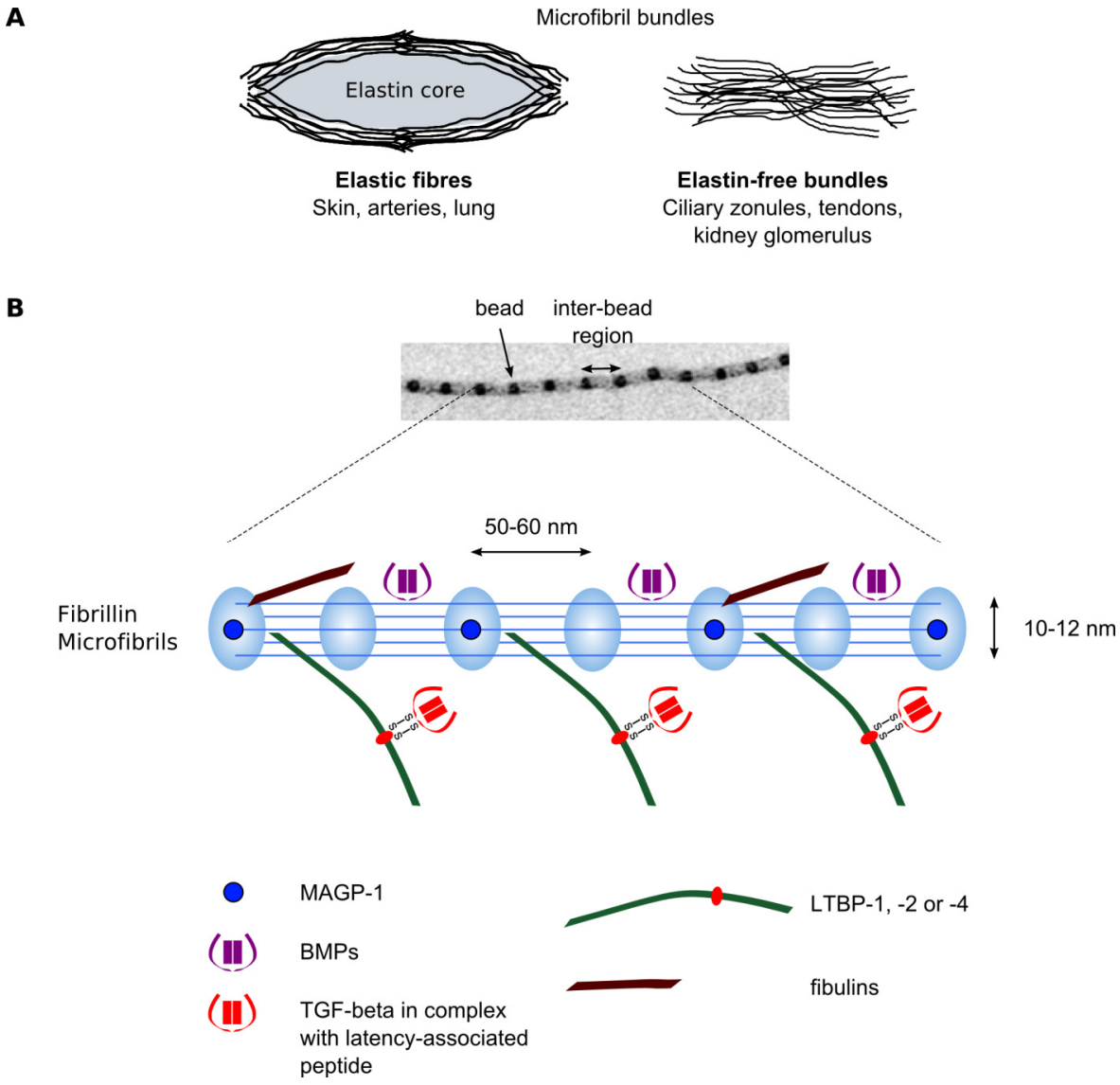


Figure 1.1 Fibrillin microfibrils in elastic and non-elastic tissues. A) Fibrillin microfibrils are found in elastic and non-elastic connective tissues. In elastic tissues, they surround a core of elastin, which is a polymer of tropoelastin. B) Microfibrils extracted from tissue, imaged by electron microscopy after rotary shadowing, have a characteristic beads-on-a-string appearance and a diameter of 10-12 nm. There is a repeat distance of 50-60 nm between the bead structures. Microfibrils are associated with a large number of accessory factors, particularly in elastic fibres. Some of these are depicted here. Adapted from Jensen *et al.*, 2012. Electron micrograph from Dorothy Halliday, D.Phil. thesis, 2000.

Elastic fibres comprise many different ECM components in addition to fibrillin and elastin, some of which associate directly with microfibrils. Microfibril associated glycoprotein-1 (MAGP-1) is an integral component of elastic fibres (Gibson *et al.*, 1986) and is localised to the bead structures of microfibrils (Henderson *et al.*, 1996). Like fibrillin-1, MAGP-1 is heavily cross-linked by disulphide bonds and transglutaminase-catalysed cross-links, and extraction from microfibrils requires strong denaturing and reducing agents (Gibson *et al.*, 1986; Qian and Glanville, 1997). MAGP-2, discovered later, is also localised to fibrillin microfibrils, but restricted to a smaller number of tissues than MAGP-1 (Gibson *et al.*, 1998).

Other major components of elastic fibres co-localise with microfibrils (Figure 1.1): the latent transforming growth factor-β binding proteins (LTBPs)-1, -2 and -4 (Gibson *et al.*, 1995; Isogai *et al.*, 2003; Ono *et al.*, 2009), bone morphogenetic proteins (BMPs) -4 and -7 (Sengle *et al.*, 2008a) and fibulins-2, -4, and -5 (Ono *et al.*, 2009). Lysyl oxidase (LOX) and LOX-like (LOXL) proteins are also found in elastic fibres, where they are thought to catalyse the cross-linking of tropoelastin into elastin (Wagenseil and Mecham 2007). Finally, there are also proteoglycans such as decorin, versican and vitronectin, and other glycoproteins, for example elastin microfibril interface located proteins (emilins) -1 and -2 (Kielty *et al.*, 2002).

1.3 Fibrillin structure

Fibrillin-1 is a ~350 kDa disulphide-rich glycoprotein with a modular structure (Corson *et al.*, 1993; Sakai *et al.*, 1986; Sakai *et al.*, 1991). The human *FBN1* gene, which encodes fibrillin-1, is located on chromosome 15 and is linked with the inherited connective tissue disorder Marfan syndrome (MFS) (Dietz *et al.*, 1991; Lee *et al.*, 1991; Maslen *et al.*, 1991). The structure of fibrillin-1 is dominated by 47 epidermal growth factor (EGF)-like domains, 43 of which contain calcium-binding consensus sequences – calcium-binding EGF-like (cbEGF) domains (Figure 1.2). cbEGF

domains are interspersed with TGFβ-binding-protein-like (TB) and hybrid (hyb) domains (Handford *et al.*, 2000; Pereira *et al.*, 1993). There are further human isoforms of fibrillin, fibrillin-2 (Lee *et al.*, 1991) and fibrillin-3 (Corson *et al.*, 2004), with similar domain organisation to fibrillin-1. However, their expression profiles are different from fibrillin-1: both are mainly expressed during development and fibrillin-3 is highly expressed in the brain.

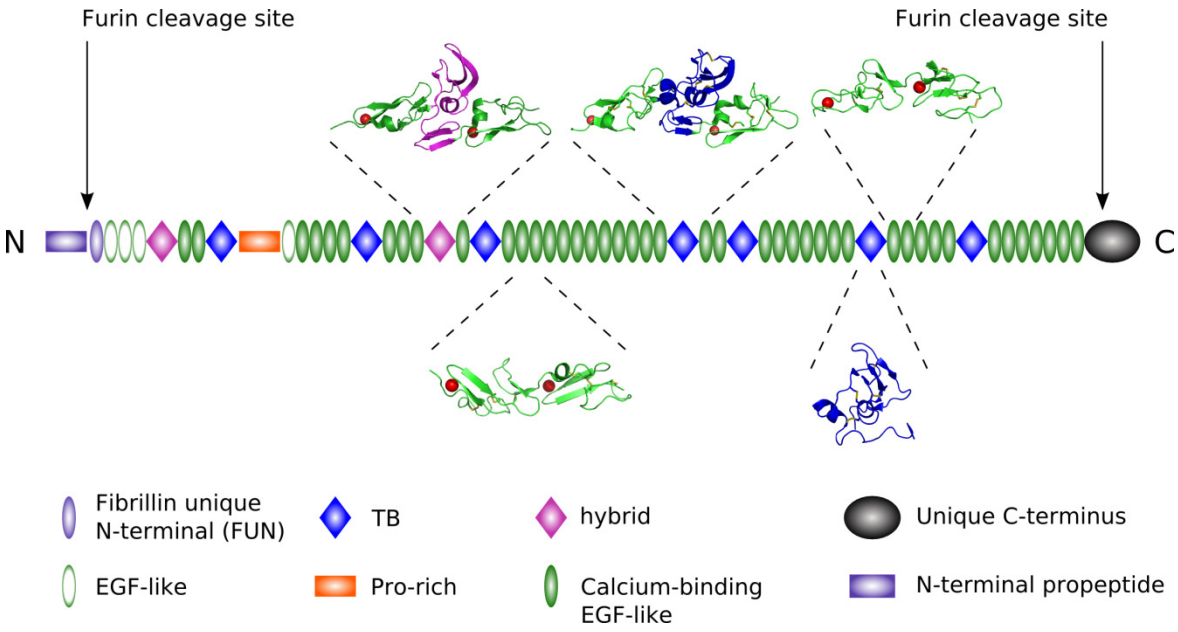


Figure 1.2 Structure and domain organisation of human fibrillin-1. Fibrillins have a modular structure dominated by EGF-like and cbEGF domains, with TB and hyb domains in between. The proline-rich region is replaced by a glycine-rich region in fibrillin-2 and a proline/glycine-rich region in fibrillin-3. There are unique regions with furin protease recognition sites at the N- and C-termini, including the fibrillin unique N-terminal (FUN) domain. Fragment domain structures are also shown: cbEGF9-hyb2-cbEGF10 (PDB ID 2W86), cbEGF12-cbEGF13 (PDB ID 1LMJ), cbEGF22-TB4-cbEGF23 (PDB ID 1UZQ), TB6 (1APJ), cbEGF32-cbEGF33 (1EMN). Structures were rendered in Pymol version 0.99.

There are high-resolution structures of several of the main fibrillin domain types. To date, it has been impossible to carry out detailed structural work on the full-length protein because of its large size, low solubility and heterogeneity due to glycosylation. Structure determination of small fragments comprising two to three domains using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy has proved to be a successful approach (Figure 1.2) (Downing *et al.*, 1996; Jensen *et al.*, 2009; Lee *et al.*, 2004; Smallridge *et al.*, 2003; Yuan *et al.*, 1997). Structural information about the various domain types is summarised below:

1.3.1 EGF-like domains

EGF-like domains are very common modules in extracellular proteins. They contain three disulphide bonds, with cysteines paired in a C1-C3, C2-C4, C5-C6 arrangement (Figure 1.3A-B); this is the same in every EGF domain (Bork *et al.*, 1996), with the exception of one domain in thrombomodulin-5 (Sampoli Benitez *et al.*, 1997). EGF domains typically have an antiparallel β-sheet structure with two major strands (forming the 'β-hairpin') and sometimes a third, minor strand; they do not have a significant hydrophobic core. cbEGFs contain an additional consensus sequence that is required for calcium binding: D/N-X-D/N-E/Q-X_m-D*/N*-X_n-Y/F, where X is any amino acid, m and n indicate a variable number of residues and * indicates potential β-hydroxylation (Handford *et al.*, 1991).

Structures of cbEGF domains show that backbone and side-chain carbonyl ligands coordinate the calcium ion with pentagonal bipyramidal geometry (Downing *et al.*, 1996; Rao *et al.*, 1995) (Figure 1.3B). In some cases, the seventh ligand is H_2O, while in others it is the backbone carbonyl of a residue in the β-hairpin (Lee *et al.*, 2004). In cases where a cbEGF immediately precedes another cbEGF, there is also a G-F/Y/W sequence between C5 and C6. The aromatic residue in this motif is involved in interdomain packing (Section 1.4) (Downing *et al.*, 1996; Smallridge *et al.*, 2003).

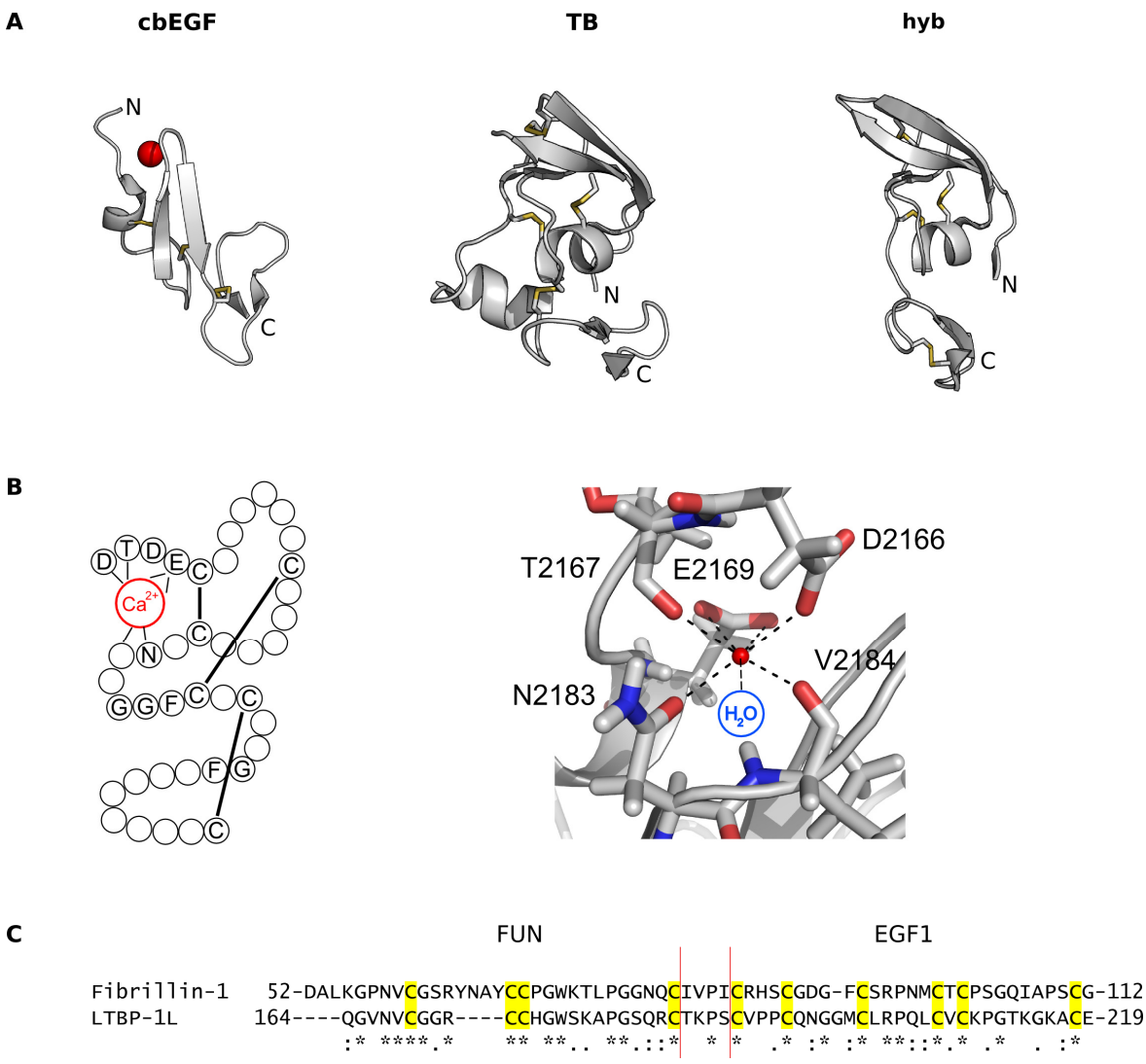


Figure 1.3 Structures of fibrillin domains. A) Cartoon representations of the major domain types from high-resolution fragment structures (Figure 1.2): cbEGF33, TB4, hyb2. The calcium ion in cbEGF33 is red and disulphide bonds are yellow. B) Calcium ion binding by domain cbEGF33. Left: two-dimensional representation, with key calcium-binding residues highlighted. Disulphide bonds are shown as thick, black lines. The contributions of the residues involved in chelation are shown with lines pointing towards the calcium ion (after Smallridge *et al.*, 2003). Right: structure of the calcium-binding site. Ligands are provided by backbone carbonyl groups (from T2167 and V2184) and side chains of the residues highlighted (indicated by dashed lines). The side chain of E2169 is a bidentate ligand and seventh ligand is H_2O. C) Sequence alignment of the fibrillin-1 and LTBP-1 FUN and EGF1 domains. Sequences were aligned with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Cysteines are highlighted in yellow and domain boundaries are demarcated with red lines. Key: '*' = conserved; ':' = very similar properties; '.' = slightly similar properties.

1.3.2 TB domains

TB domains occur exclusively in the fibrillins and LTBPs (Robertson *et al.*, 2011). They have four disulphide bonds, in a C1-C3, C2-C6, C4-C7, C5-C8 arrangement, and have a more globular structure than EGF-like domains (Figure 1.3A) (Lee *et al.*, 2004; Yuan *et al.*, 1997). The secondary structure consists of one four-strand β-sheet and one two-strand β-sheet, as well as two helices. The first helix lies at the centre of the domain and has a C-C-C motif at its C-terminal end, meaning that it is covalently linked to three portions of the surrounding polypeptide chain. A conserved aromatic residue (tryptophan in the majority of cases) contributes to the hydrophobic core, found as part of the G-X-X-W/F-G motif. The C-terminal β-hairpin is not linked to any other part of the domain by disulphide bonds, but does contain a G-F/Y consensus motif, as in cbEGF domains.

1.3.3 Hybrid domains

The hyb domain structure is a composite of TB and cbEGF domains (Figure 1.3A) (Corson *et al.*, 1993; Jensen *et al.*, 2009). Like the TB domain, the hyb domain has a globular fold, with a central helix, surrounded by three portions of β-sheet. It also has a G-X-X-W/F-G sequence and there is a C-C motif at the end of the helix, analogous to the cysteine triplet of the TB domain. However, the arrangement of its four disulphide bonds is different: C1-C3, C2-C5, C4-C6 and C7-C8. The fibrillin hyb1 domain also contains a ninth, unpaired cysteine. The C-terminal segment of the hyb domain is more like the C-terminal end of a cbEGF − the penultimate cysteine is disulphide-bonded to the last cysteine. The hyb domain also has a G-F/Y motif in the β-hairpin between C7 and C8.

1.3.4 Pro/Gly-rich region

Human fibrillin-1 has a unique proline-rich (Pro-rich) region between the TB1 and EGF4 domains. It has been suggested that it might act as a 'hinge' that would allow the fibrillin molecule to bend back on itself, by analogy with a similar region in dystrophin (Pereira *et al.*, 1993). However, there is no evidence in support of this hypothesis as there is no structural information about the region. In place of the Pro-rich region, fibrillin-2 has a glycine-rich (Gly-rich) sequence and fibrillin-3 has a proline- and glycine-rich region (Corson *et al.*, 2004). The functional significance of these differences is not clear, but they allow fibrillin isoforms to be distinguished as their domain organisation is otherwise very similar.

1.3.5 N- and C-terminal regions

The N-terminal region of human fibrillin-1 consists of a signal sequence, propeptide and the fibrillin unique N-terminal (FUN) domain. The propeptide sequence and the FUN domain are separated by a furin/paired basic amino cleaving enzyme (PACE) recognition site − R-A-K-R (residues 42-45) in human fibrillin-1. Little is known about the structure or function of the FUN domain. It has four glycine residues at its N-terminus, which are expected to be unstructured. Moreover, the FUN domain has four cysteine residues that are disulphide-bonded in recombinant fibrillin fragments (Reinhardt *et al.*, 2000a). Two of the cysteines occur as a C-C motif, but the arrangement of disulphide bonds is unknown. There is an homologous domain in human LTBP-1 and -2 (Robertson *et al.*, 2011), with the same arrangement of cysteines and followed by an EGF-like domain (Figure 1.3C).

There are no known homologues of the C-terminal region following the last cbEGF domain (cbEGF43 in human fibrillin-1) in other proteins. Like the fibrillin N-terminal region, the C-terminal

region is divided into two segments by a furin/PACE site: a sequence containing a C-X-X-C motif and the C-terminal propeptide. There is no structural information about either region. It was previously suggested that this region is similar to the C-terminal domain of fibulins (Giltay *et al.*, 1999), but there is in fact very low sequence homology. The fibulin C-terminal domain also contains two cysteines, but there are 5 or 6 intervening residues instead of two and there is no furin/PACE recognition sequence.

1.4 Interdomain packing interactions

High-resolution structures of fibrillin fragments provide insights into interdomain packing interactions. Many of these interactions in fibrillin involve calcium binding. cbEGF, TB and hyb domains have a G-F/Y/W motif in their C-terminal segment that packs with the next domain (Section 1.3; Figure 1.4). This packing interaction has implications for calcium binding – the calcium affinity of a domain may be significantly enhanced by the presence of an N-terminal domain. For example, in the cbEGF32-cbEGF33 domain pair, the dissociation constant (K_d) for calcium for cbEGF33 is ~25-fold lower than for cbEGF32 (Knott *et al.*, 1996), as measured by NMR. There is an even greater enhancement (~70-fold) for the cbEGF12-cbEGF13 pair (Smallridge *et al.*, 1999). Importantly, calcium binding stabilises these interdomain interfaces. An NMR study showed that the presence of calcium increased the rigidity of the backbone of the cbEGF32-cbEGF33 pair and that calcium affinity was inversely correlated with flexibility (Werner *et al.*, 2000). Thus, calcium is important for maintaining a rod-like arrangement of cbEGF arrays.

Similar observations have been made about TB-cbEGF and hyb-cbEGF domain pairs. With the exception of TB1, all TB or hyb domains in human fibrillin-1 are followed by a cbEGF. The K_d values for calcium binding by cbEGFs with an N-terminal TB domain vary from nM to mM, with the highest affinity pair being TB4-cbEGF23 and the lowest affinity TB6-cbEGF32 (Jensen *et al.*, 2005).

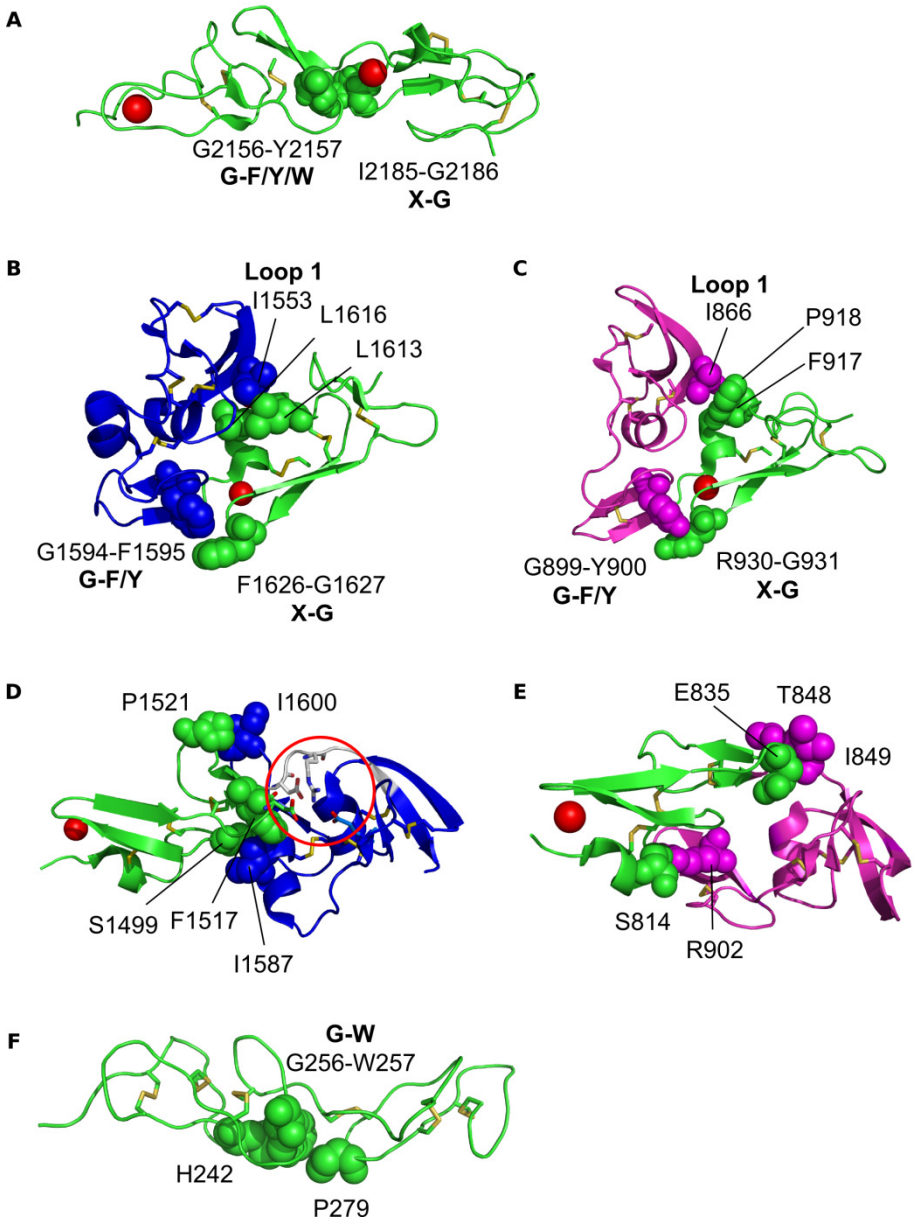


Figure 1.4 Interdomain packing interactions. A) cbEGF32-cbEGF33. Calcium ions are shown in red and atoms of side-chains are shown as spheres. Y2157 in cbEGF32 (G-F/Y/W motif) packs with I2185 in cbEGF33 (X-G motif). B) TB4-cbEGF23. G-F/Y and X-G motifs are highlighted, as in cbEGF32-cbEGF33. 'Loop 1' of the TB domain also packs with two leucines in a loop in cbEGF23. C) hyb2-cbEGF10. This interface has similar packing motifs to the TB-cbEGF interface – the two leucines are replaced by proline and phenylalanine. D) cbEGF22-TB4. Residues involved in non-polar and hydrophobic packing are shown as spheres. The circled area highlights charged residues in the cbEGF, interdomain linker (white) and TB domain (shown as sticks) that form salt bridges. E) cbEGF9-hyb2. The interface involves non-polar interactions. F) EGF1-EGF2 pair from human Jagged-1 (PDB ID 2VJ2). W257 in the G-W motif packs with H242 (EGF1) and P279 (EGF2). Figure after Jensen *et al.*, 2012.

The crystal structure of the cbEGF22-TB4-cbEGF23 fragment (Lee *et al.*, 2004) and NMR experiments with TB-cbEGF pairs (Jensen *et al.*, 2005) explain the variation in calcium affinity. Key residues in the TB4-cbEGF23 and other pairs involved in hydrophobic packing interactions are absent from TB6-cbEGF32: notably an aromatic residue from the G-F/W motif in the TB domain, a hydrophobic residue in the X-G motif of the cbEGF and residues in 'loop 1' of the cbEGF (Figure 1.4B). Hyb-cbEGF domain pairs also have these key packing residues (Figure 1.4C) and therefore bind calcium with high affinity (nM K_d values) (Jensen *et al.*, 2009). Combined with the identification of a flexible linker between TB6 and cbEGF32 (Yuan *et al.*, 2002), these results suggest that calcium affinity is related to the stability of TB-cbEGF and hyb-cbEGF interfaces.

Calcium-independent interdomain interfaces in fibrillin have been less extensively studied. There is some information from the crystal structures of the cbEGF22-TB4-cbEGF23 and cbEGF9-hyb2-cbEGF10 fragments (Jensen *et al.*, 2009; Lee *et al.*, 2004). At the cbEGF22-TB interface, there are residues that form hydrophobic interactions and salt bridges (Figure 1.4D). The cbEGF9-hyb2 interface is less extensive, as the hyb domain lacks the second helix found in the TB domain, which mediates interactions at the cbEGF22-TB4 interface (Figure 1.4D and E). The dynamics of these interfaces have not been studied in depth.

Other predicted calcium-independent interfaces in fibrillin-1 are between the FUN and non-calcium binding EGF domains 1-3, between EGF3 and hyb1, between TB1 and the Pro-rich region, between the Pro-rich region and non-calcium binding EGF domain 4 and finally between cbEGF43 and the unique C-terminal region. There is no structural information about these interfaces, but structures from other proteins may provide clues. For example, in the EGF1-EGF2 pair of human Jagged-1 (Cordle *et al.*, 2008a) a tryptophan residue from the EGF1 domain packs against a histidine in EGF1 and a proline in EGF2 (Figure 1.4F). Further information about these interfaces will aid understanding the conformation and dynamics of fibrillin in microfibrils.

1.5 Fibrillin evolution

With the recent release of many genome and RNA transcript sequences, it is now possible to study the evolution of fibrillin and other elastic fibre components in depth (Piha-Gossack *et al.*, 2012; Robertson *et al.*, 2011). The evolutionary timing of the emergence of ECM components and comparisons of their protein sequences can provide valuable functional insights. Fibrillins first appeared at least 600 million years ago − fibrillin sequences have been found in animals with radial (Cnidaria) and bilateral symmetry (Bilateria), but not in other kingdoms, such as plants, fungi and protozoa (Robertson *et al.*, 2011). They are also absent from nematodes such as *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*, although they are found in other arthropods. The domain organisation of the fibrillins is extremely similar across the phylogenetic tree, from jellyfish to man. Moreover, the features of the interdomain interfaces appear relatively unchanged by evolution (Jensen *et al.*, 2012), suggesting that the conformation of fibrillin is also conserved. Microfibril-like structures have been found in a number of vertebrate and invertebrate organisms, including jellyfish (Reber-Müller *et al.*, 1995) and sea cucumber (Thurmond and Trotter, 1996). This supports the conserved role of fibrillins as microfibril-forming proteins.

1.6 Fibrillin microfibril organisation

The organisation of fibrillin in microfibrils is unknown, with various groups proposing different models. Models rationalise observations that the average repeat distance between the bead structures of microfibrils is 50-60 nm, while the extended fibrillin monomer length is ~150 nm (Section 1.2). Two general types of model have been suggested, based on a large amount of experimental evidence. Firstly, folded models, in which fibrillin monomers fold back on themselves (Figure 1.5A). Secondly, staggered models, in which the fibrillin monomers have a

largely linear conformation, but overlap in a manner that gives rise to the observed inter-bead distance (Figure 1.5B).

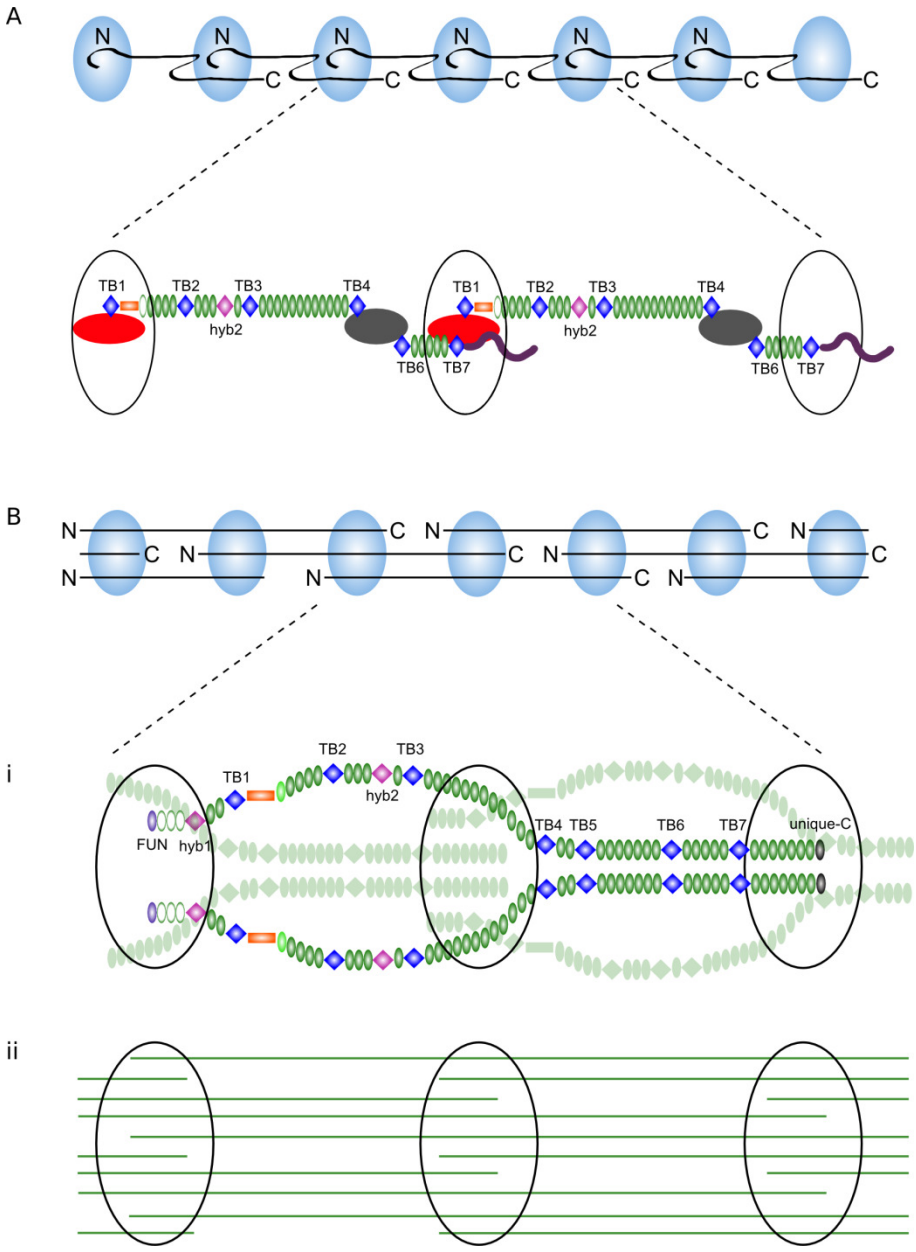


Figure 1.5 Models of fibrillin microfibril organisation. A) The pleated model (Baldock *et al.*, 2006) predicts folding of the N-terminal domains (red oval) from FUN to cbEGF2, folding of domains cbEGF23-cbEGF31 (grey oval) and curvature of domains cbEGF37-unique C-terminal domain (purple line). One fibrillin monomer spans one inter-bead distance. B) Staggered models. i) The partially staggered model of Kuo *et al.*, 2007 based on antibody epitope mapping. One fibrillin monomer spans two inter-bead distances, meaning that its conformation is not completely linear. ii) Model of Lee *et al.*, 2004. Fibrillin molecules are fully elongated, so each monomer crosses three inter-bead distances.

Initial evidence in support of the folded model came from antibody epitope mapping, which formed the basis of several studies into microfibril organisation. Monoclonal antibodies against fibrillin fragments can be used to label extracted microfibrils. Electron microscopy is then used to map the location of these antibodies on the rotary-shadowed or negative-stained microfibrils. An early study showed that antibodies directed against N- and C-terminal peptides bound to the inter-bead region, close to the beads (Reinhardt *et al.*, 1996), but could not discriminate between folded and staggered models. A later study showed patterns of antibody labelling of tensioned and untensioned microfibrils that were consistent with a folded model − distances between epitopes were shorter than would be expected in fully extended fibrillin molecules (Baldock *et al.*, 2001). Further support came from a small angle X-ray scattering (SAXS) study of overlapping multi-domain fibrillin fragments (Baldock *et al.*, 2006). Molecular modelling based on known domain structures and SAXS-derived shape information indicated that many regions adopted non-linear conformations. Based on this work, an updated model was proposed (Figure 1.5A).

This folded model is inconsistent with other antibody mapping studies and structural data. This might be explained by the methods of microfibril preparation. Collagenase, used in microfibril extraction in the 2001 and other studies, cleaves fibrillin at multiple sites, resulting in the reduction of structural integrity of microfibrils and loss of some antibody epitopes (Kuo *et al.*, 2007). The pattern of antibody labelling of guanidine-extracted microfibrils is different, showing an overlap of epitopes from regions across the fibrillin molecule. This can only be explained by a staggered arrangement of fibrillin monomers (Figure 1.5Bi).

In addition, the SAXS data are inconsistent with high-resolution structures of fibrillin domain pairs and triples. Structures of cbEGF-cbEGF, cbEGF-TB-cbEGF and cbEGF-hyb-cbEGF regions (Section 1.4) show a linear arrangement of domains. These formed the basis of an earlier version of the staggered model, which proposed that fibrillin molecules were fully elongated in microfibrils

(Figure 1.5Bii) (Lee *et al.*, 2004).The folded model also requires a kink in the structure in the TB4-TB5 region, which seems implausible based on the structures and high calcium affinity of these domains (Jensen *et al.*, 2005). Nonetheless, folding around other regions, such as the Pro-rich domain, cannot be ruled out.

The structural organisation of fibrillin has implications for microfibril elasticity. Both types of model need to predict mechanisms of extension and recoil. The folded model suggests that extension occurs through a reversal of the bending of fibrillin molecules, but does not explain how recoil could occur. By contrast, the staggered model predicts extension and recoil through partial unfolding and refolding of interdomain interfaces (Jensen *et al.*, 2012). Consistent with the folded model, extension of microfibrils extracted from tissue by applied force (Glab and Wess, 2008) or increased ionic strength (Wang *et al.*, 2009) result in an increase in bead periodicity from ~56 nm to 150-160 nm; this effect is partially reversible if the microfibrils are not extended beyond 150% of their original length (Wang *et al.*, 2009; Wess *et al.*, 1998). The degree of extension is significantly greater than predicted by staggered models (Jensen *et al.*, 2012). While the results are plausible, tissue damage may have permitted artificially large microfibril extension. At present, firm conclusions cannot be drawn about which model is the most accurate.

1.7 Fibrillin microfibril assembly

The processes underlying fibrillin microfibril assembly are only partially understood. Assembly is likely to involve interactions between fibrillin monomers, as well as between fibrillin and other ECM components. Possible steps in the pathway are illustrated in Figure 1.6.

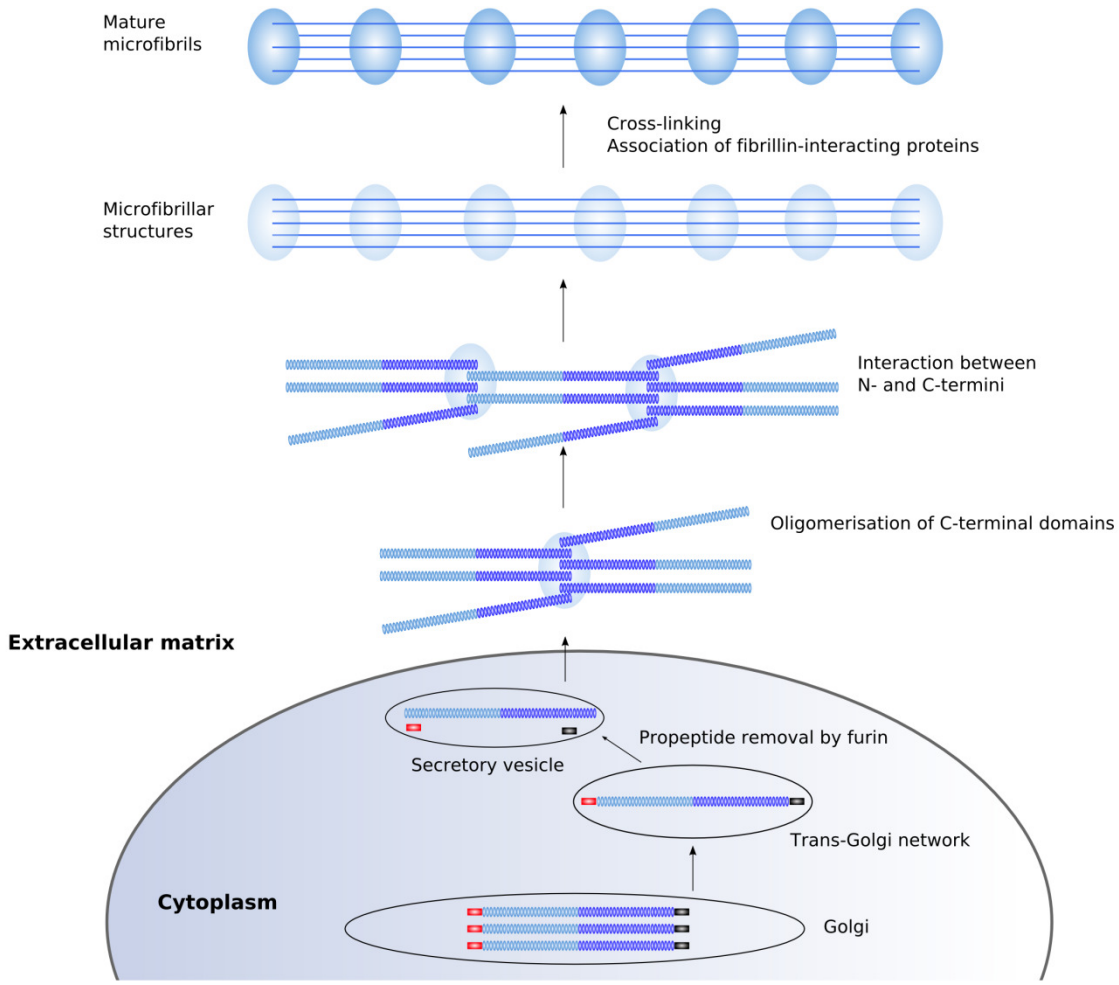


Figure 1.6 Fibrillin assembly into microfibrils. Fibrillin is translated, then translocated into the endoplasmic reticulum and processed as it passes through the secretory pathway. The molecule is folded, forms disulphide bonds, undergoes glycosylation and potentially β-hydroxylation. The N- and C-terminal propeptides are cleaved by furin near the cell surface. The C-terminal domains aggregate, forming oligomers that interact with the N-terminal domains of other fibrillin monomers with high affinity. The N-C interaction is thought to be an initial stage in assembly, although the timing and location are unknown. Fibrillin aggregates then mature into microfibrils with bead-like structures. These are only visible after several weeks when extracted from cell culture (Hubmacher and Reinhardt, 2011).

1.7.1 Furin-mediated processing of the N- and C-termini

One of the early critical stages is proteolytic processing of the N- and C-termini of 'pro-fibrillin' into fibrillin. The furin/PACE sites after the N-terminal propeptide and before the C-terminal propeptide are highly conserved, indicating their universal importance (Piha-Gossack *et al.*, 2012). Furthermore, mammalian fibroblasts process fibrillin at these sites. Two studies identified the first residue of secreted full-length fibrillin (Reinhardt *et al.*, 2000a) or an N-terminal fragment (Kettle *et al.*, 2000) as R45, the residue after the predicted N-terminal furin/PACE cleavage site. It is not known whether the N-terminal propeptide affects fibrillin assembly.

Initial evidence for C-terminal processing came from a patient with skeletal features of MFS (Milewicz *et al.*, 1995). The patient had a missense mutation that resulted in the substitution of R2726, two residues before the core furin consensus recognition site, with tryptophan. Only half of the fibrillin secreted by dermal fibroblasts from the patient was processed and unprocessed material did not incorporate into the ECM. C-terminal processing and the role of furin/PACE enzymes were confirmed using recombinantly expressed, truncated 'mini-fibrillin' constructs (Ashworth *et al.*, 1999; Raghunath *et al.*, 1999; Ritty *et al.*, 1999). Purification and N-terminal sequencing of the cleaved propeptide in the medium of cells stably expressing a C-terminal fragment confirmed cleavage after R2731 (Lönnqvist *et al.*, 1998). In addition, furin was shown to be the processing enzyme, rather than another member of the furin/PACE family, using a specific inhibitor (Wallis *et al.*, 2003).

Furin cycles between the trans-Golgi network and cell surface, and can also be secreted in an active form (Molloy *et al.*, 1999). The site of furin cleavage was inferred from experiments in which fibrillin secretion was blocked using agents that raise intracellular pH, thereby preventing passage of protein from the endoplasmic reticulum to the Golgi. A study using chinese hamster

ovary-K1 cells found that processing of recombinant protein was not blocked (Ritty *et al.*, 1999). Subsequent work using human dermal fibroblasts found that processing of endogenous fibrillin was prevented and that the cells secreted furin (Wallis *et al.*, 2003). Moreover, reversibly blocking secretion by lowering the temperature inhibited processing. Thus, it is likely that furin-mediated proteolysis occurs at or near the cell surface. Despite this, it is still not known how the C-terminal propeptide blocks fibrillin assembly or how the timing of processing links to other assembly events.

1.7.2 Roles of the N- and C-terminal regions

Interactions between the N- and C-terminal regions are thought to be important for microfibril structure and assembly. All models of microfibril organisation predict approximate co-localisation of the N- and C-terminal regions (Section 1.6). This requirement is based on a number of studies that showed interactions of recombinant N- and C-terminal fibrillin fragments *in vitro* (Figure 1.7). Indirect evidence for the N-C interaction came from a study showing that an N-terminal fragment (up to domain cbEGF6) promoted aggregation and C-terminal processing of a C-terminal fragment when they were co-expressed in an *in vitro* translation system (Ashworth *et al.*, 1999). This effect was dependent on the presence of domains TB1 to cbEGF6 (Figure 1.7A). However, a direct interaction was not assayed and the results could have been artefacts of the protein production method.

Subsequently, a direct interaction between the N- and C-terminal halves of fibrillin-1 was shown (Lin *et al.*, 2002). N- and C-terminal fragments (Figure 1.7B) produced in human embryonic kidney 293 (HEK293) cells interacted with high affinity (low nM K_d) in enzyme-linked immunosorbent assay (ELISA) and surface plasmon resonance (SPR) assays. The interaction was dependent on $CaCl_2$ and was abolished by ethylenediaminetetraacetic acid (EDTA). This study found that

equivalent N- and C-terminal fragments of fibrillin-2 did not interact with each other, but could interact with fibrillin-1 fragments. This suggested a molecular basis for the formation of heterotypic microfibrils, which contain more than one fibrillin isoform.

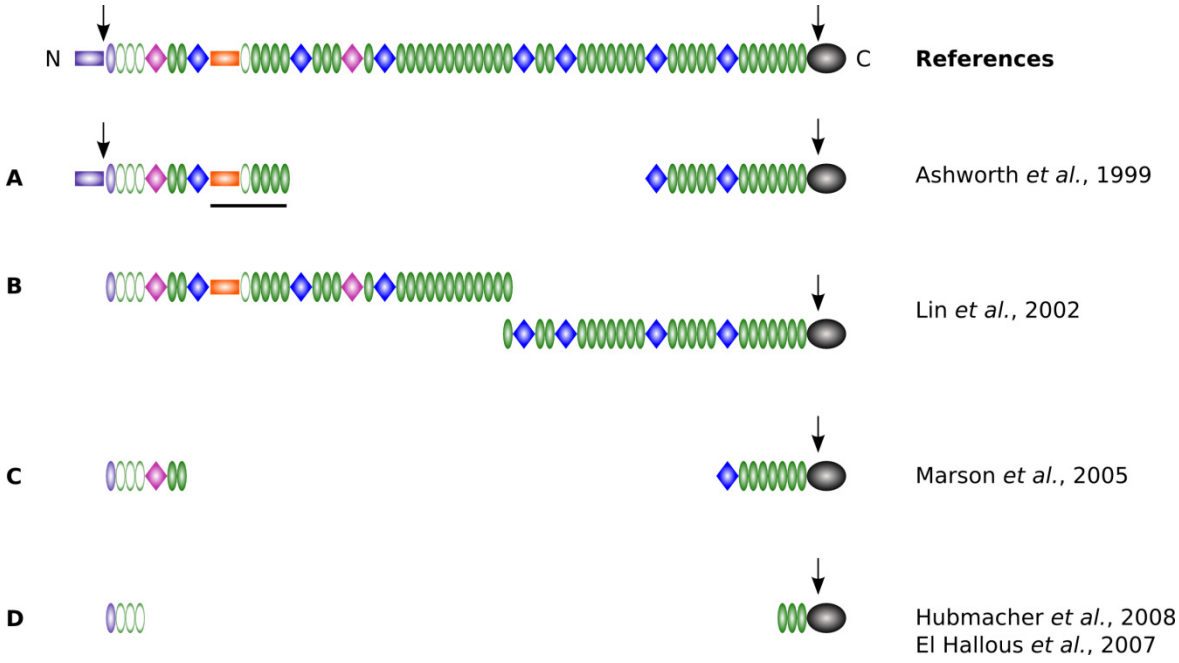


Figure 1.7 Molecular dissection of the fibrillin N-C terminal interaction. Fragments used in N-C terminal interaction studies are shown. Arrows indicate the presence of furin recognition sites. In A, the N-terminal fragment failed to induce aggregation of the C-terminal fragment when the underscored region was removed. The legend is the same as in Figure 1.2.

The interacting regions of fibrillin-1 were narrowed down to the FUN-cbEGF2 region and the TB7-unique C-terminus using smaller fragments in similar assays (Marson *et al.*, 2005) (Figure 1.7C). As for the larger fragments, the interaction was found to have a low nM K_d and was partially abrogated by EDTA. Finally, the minimal regions needed for the interaction were identified as FUN-EGF3 at the N-terminus and cbEGF41-cbEGF43 at the C-terminus (Figure 1.7D). A C-terminal fragment terminating at cbEGF40 did not interact with the N-terminal half, while deletion of only the unique C-terminal domain had no effect on binding (Hubmacher *et al.*, 2008). A fragment encompassing cbEGF41-cbEGF43 and the unique C-terminus bound to the N-terminal half,

although with much lower affinity. Data for the N-terminal region have not been published, mentioned only as an aside in a study on fibrillin-fibulin interactions (El-Hallous *et al.*, 2007), but were presented at a conference (Sacha Jensen, personal communication). Molecular details of this interaction are still unknown, as is its role in microfibril assembly.

As well as 'head-to-tail' interactions, oligomerisation of the N- and C-terminal regions may play a role in assembly. The C-terminal half of fibrillin-1 can oligomerise, dependent on the presence of domains cbEGF41-cbEGF43 (Hubmacher *et al.*, 2008). When visualised using electron microscopy, these multimers have bead-like structures that might be analogous to the beads of extracted microfibrils. Significantly, oligomers interact with apparently higher affinity with the N-terminal half of fibrillin and can block microfibril assembly in cell culture. This suggests that aggregation of the C-terminus might precede the N-C interaction in the microfibril assembly pathway.

Self-interactions of the N-terminal domains may also give rise to fibrillin oligomerisation. As well as non-covalent interactions (Marson *et al.*, 2005), the unpaired cysteine in the hyb1 domain (C204 in human fibrillin-1) could form intermolecular disulphide bonds. This residue is absolutely conserved (Robertson *et al.*, 2011), but its role is unclear. The side chain is present as a reactive free thiol in recombinant fibrillin fragments (Reinhardt *et al.*, 2000a). This is consistent with its surface-exposed location in a homology model of the hyb1 domain (Jensen *et al.*, 2009). Furthermore, fibrillin in chick aorta cultures and recombinant N-terminal fragments can form disulphide-bonded aggregates and dimers respectively (Reinhardt *et al.*, 2000a; Trask *et al.*, 1999). However, the native reaction partner or partners of C204 are unknown. Its role in assembly was also questioned by a study showing that fibrillin lacking hyb1 could still incorporate into microfibrils in mice (Charbonneau *et al.*, 2010).

1.8 Functions of ECM components that interact with fibrillin

Fibrillins interact with a large number of elastic fibre and ECM components. The functional roles of many of these interactions have already been identified and progress has been made towards mapping the binding sites of these components on fibrillin (Figure 1.8). Many fibrillin-binding components also interact with each other.

1.8.1 Microfibril structure and assembly

MAGP-1 is a core constituent of microfibrils and MAGP-2 also associates with microfibrils in some tissues (Section 1.2). Fibrillin-1 interacts directly with MAGP-1 and MAGP-2 (Hanssen *et al.*, 2004; Jensen *et al.*, 2001; Rock *et al.*, 2004). No direct role of MAGPs has been shown in microfibril assembly and they are not found in lower metazoa (Robertson *et al.*, 2011), but they are involved in elastic fibre assembly (Section 1.8.4). MAGP-1 knockout mice do not have elastic fibre defects, suggesting that they are not essential for microfibril formation or structure (Weinbaum *et al.*, 2008).

A number of ECM components appear to regulate the formation of fibrillin microfibrils, but their mechanisms of action are not understood. First, fibrillin assembly is dependent on the presence of fibronectin fibrils in higher metazoa (Kinsey *et al.*, 2008; Sabatier *et al.*, 2009). The binding site of fibronectin type III domains 6 to 9 is in the C-terminal half of fibrillin-1, -2 and -3 (Figure 1.8). In addition, the GAG heparan sulphate also regulates fibrillin assembly through a direct interaction. Heparan sulphate-conjugated proteoglycans include perlecan (Tiedemann *et al.*, 2005) and the syndecan cell surface receptors (Ritty *et al.*, 2003). There are four high affinity binding sites on fibrillin-1 for heparin, which has a similar structure to heparan sulphate (Figure 1.8) (Cain *et al.*, 2005; Cain *et al.*, 2008; Ritty *et al.*, 2003; Tiedemann *et al.*, 2001). Heparin blocks microfibril

formation in culture, but GAG synthesis is required for assembly (Tiedemann *et al.*, 2001). Several proteins of the 'A distintegrin-like and metalloprotease with thrombospondin motif' (ADAMTS) and ADAMTS-like (ADAMTSL) family also interact with fibrillin: ADAMTS10, ADAMTSL2, ADAMTSL3 and ADAMTSL6β (Figure 1.8) (Kutz *et al.*, 2011; Le Goff *et al.*, 2011; Sengle *et al.*, 2012; Tsutsui *et al.*, 2010). ADAMTS10 and ADAMTSL6β both promote microfibril assembly in cell culture.

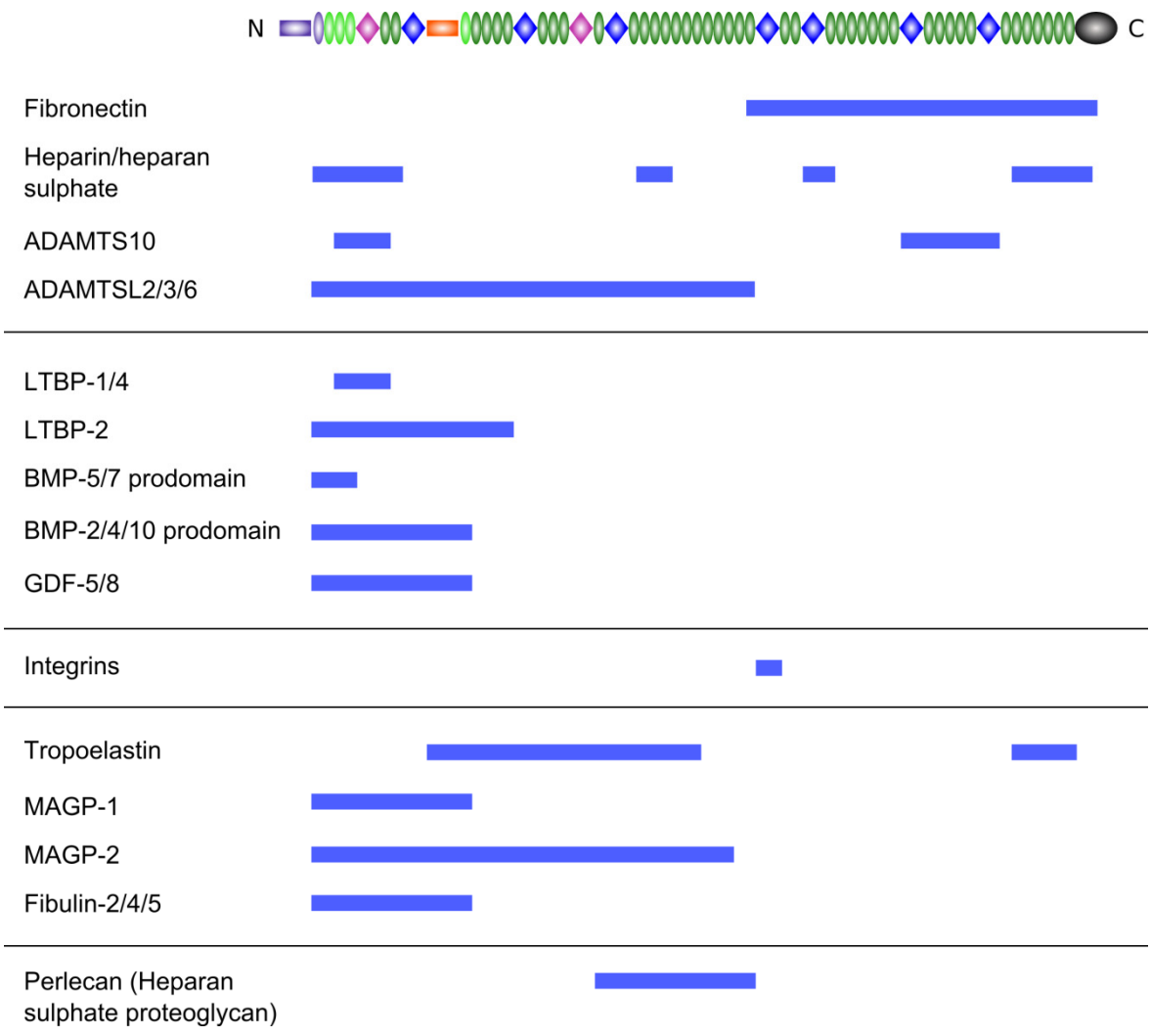


Figure 1.8 Binding sites of accessory factors that interact with fibrillin. Interaction sites are shown using blue bars for each molecule. In many cases the precise binding sites have not been identified, but have been narrowed down using fibrillin fragments. References for each interaction are given under the relevant sections in the text. GDF = growth and differentiation factor.

1.8.2 Integrins: cell adhesion and role in microfibril assembly

Integrins are transmembrane cell surface receptors that mediate cell adhesion and migration through interactions with ECM components. They play a key role in the formation of focal adhesions (Shattil *et al.*, 2010). Integrins comprise a heterodimer of α and β subunits, the precise composition of which determines their ligand-binding specificity. The molecular basis of ligand recognition has been elucidated for many integrins. For example, the five αV integrins, α5β1, α8β1 and αIIbβ3 recognise ligands containing the R-G-D (RGD) sequence, while α4β1, α4β7 and α9β1 bind to L-D-V (LDV) sequences. By contrast, binding specificity of β2-family integrins is conferred by an additional I-domain in the α subunit (Campbell and Humphries, 2011).

The TB4 domain of human fibrillin-1 has an RGD motif on a flexible loop (Lee *et al.*, 2004) and can interact with integrins α5β1, αVβ3 and αVβ6 (Bax *et al.*, 2003; Jovanovic *et al.*, 2007; Pfaff *et al.*, 1996). Tsuruga *et al.* found that the addition of an RGD-containing αVβ3 inhibitor peptide inhibited fibrillin microfibril assembly in cell culture (Tsuruga *et al.*, 2009). Integrins might also regulate microfibril assembly indirectly through other ECM components, many of which have integrin-binding sequences − for example fibronectin (Kinsey *et al.*, 2008). The RGD motif in the fibrillin-1 TB4 domain is not found outside vertebrates, suggesting that it is unlikely to have a universal role in assembly (Robertson *et al.*, 2011).

1.8.3 Growth factor regulation

Fibrillins regulate growth factor signalling in the ECM through interactions with the LTBPs, BMPs and growth and differentiation factors (GDFs). Members of the transforming growth factor-β (TGFβ) family modulate intracellular signalling pathways through interactions with cell surface receptors and are implicated in many biological processes, including development, wound healing

and immune regulation (Massagué, 1998; Tran, 2012). With the exception of LTBP-2, LTBPs sequester TGFβ in its latent form. TGFβ is secreted as a covalent complex with the latency associated peptide (LAP) and one of the LTBPs (Todorovic and Rifkin, 2012) − the large latent complex (LLC; Figure 1.9A). The LAP is covalently linked to the LTBPs through a disulphide rearrangement in the LTBP TB2 domain (Lack *et al.*, 2003; Saharinen and Keski-Oja, 2000). C-terminal domains of LTBP-1, -2, and -4, but not LTBP-3, interact with fibrillin N-terminal domains (Hirani *et al.*, 2007; Isogai *et al.*, 2003; Ono *et al.*, 2009) (Figure 1.8). The mechanism of TGFβ release and activation is not yet known, but could involve proteolytic cleavage and interactions of the LLC with integrins (Figure 1.9A) (Shi *et al.*, 2011; Todorovic and Rifkin, 2012).

In contrast to the LTBPs, BMPs and GDFs are non-covalently bound to their prodomains. The prodomains of BMP-2, -4, -5, -7, and -10 and GDF-5 can interact with fibrillin-1 and -2 (Gregory *et al.*, 2005; Sengle *et al.*, 2008a; Sengle *et al.*, 2011). They also bind the fibrillin N-terminus, but at different sites from LTBP-1 and -4 (Figure 1.8). Unlike the LTBPs, sequestration by fibrillin does not necessarily mean that BMPs and GDFs are inactive. The activity of BMP-10 is blocked by its prodomain (Sengle *et al.*, 2011) and the BMP-7 prodomain competes with certain BMP receptors for binding to BMP-7 (Sengle *et al.*, 2008b). However, the prodomain-BMP complexes of BMP-4, -5 and -7 are still active and BMP-1 can activate BMP-10 by proteolysis (Sengle *et al.*, 2011). The likely role of microfibrils is to concentrate growth factors in the ECM, keeping only some in a latent state.

1.8.4 Elastic fibre assembly

In elastic fibres, microfibrils surround a core of elastin. Tropoelastin assembles spontaneously into higher-order structures by a process called coacervation (Yanagisawa and Davis, 2010). The mechanism by which it is assembles on microfibrils and then undergoes cross-linking is not fully

understood, but is known to involve several elastic fibre components (Figure 1.9B). Several studies have shown a direct interaction of tropoelastin with fibrillin-1 and fibrillin-2 (Rock *et al.*, 2004; Trask *et al.*, 2000). Some of the MAGPs, fibulins and LTBPs may also act as adaptors that recruit tropoelastin. MAGP-1 can interact with tropoelastin, as well as fibrillin (Bashir *et al.*, 1994; Jensen *et al.*, 2001). Fibulins-2, -4 and -5 bind to the fibrillin-1 N-terminus (Ono *et al.*, 2009) and can also bind tropoelastin (Wagenseil and Mecham, 2007). While LTBPs do not bind tropoelastin, they may regulate the binding of fibulins to fibrillin, as they have overlapping binding sites (Ono *et al.*, 2009). In addition, LTBP-2 and LTBP-4 can actively recruit fibulin-5 (Hirai *et al.*, 2007; Todorovic and Rifkin, 2012) and LTBP-1 binds fibulin-4 (Massam-Wu *et al.*, 2010). Finally, LOX and LOXL cross-linking enzymes can bind fibulin-4, fibulin-5 and fibronectin, suggesting a mechanism of recruitment to fibrillin microfibrils (Yanagisawa and Davis, 2010).

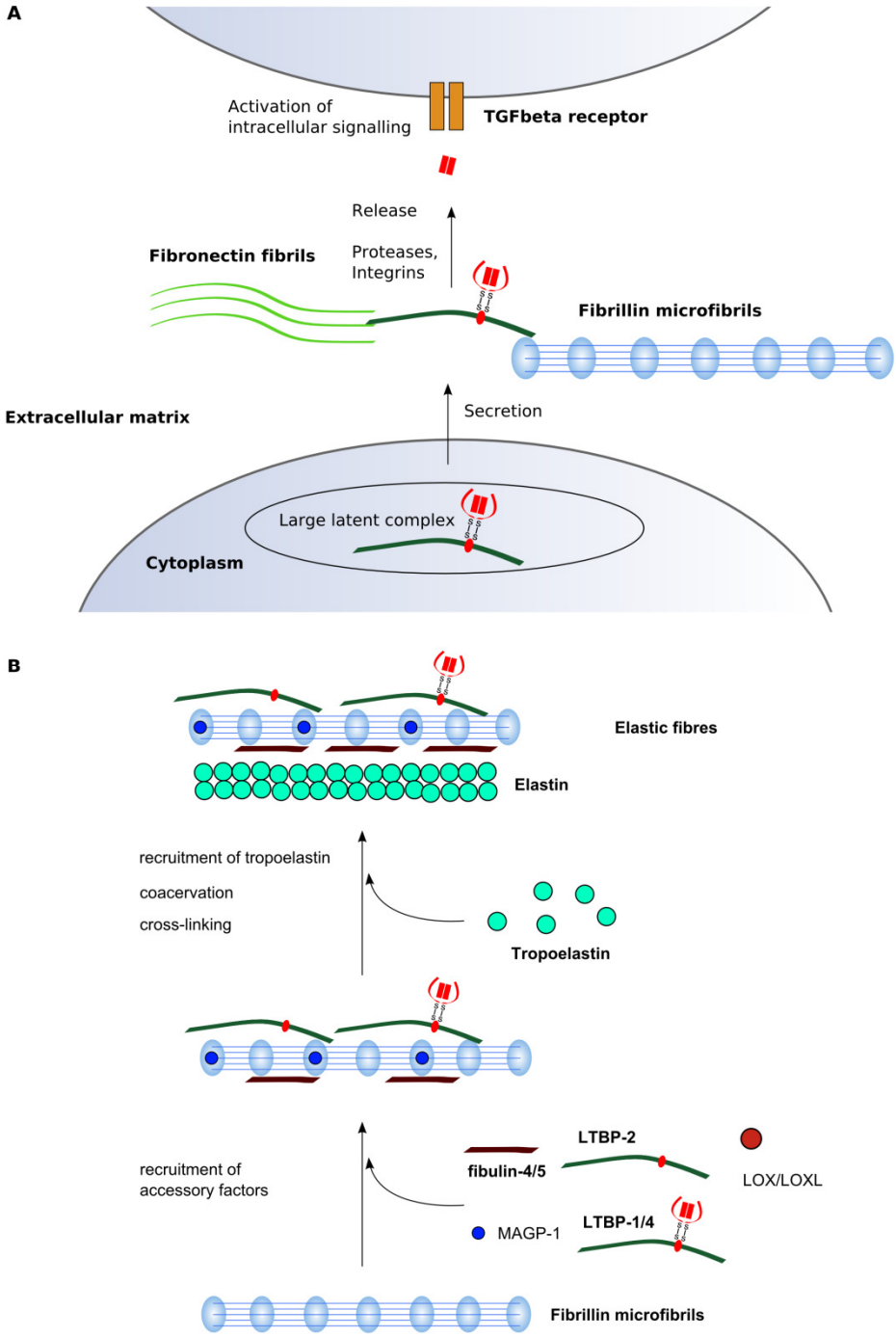


Figure 1.9 Fibrillin and elastic fibre function. A) TGFβ sequestration and release by microfibrils. TGFβ is secreted as part of the large latent complex and is then incorporated into the ECM through interactions with fibrillin and fibronectin. Mechanisms of TGFβ activation and release are not well understood. Signal transduction is initiated by binding of active TGFβ to its receptor on the cell surface. B) Elastic fibre formation. Elastin is deposited onto a scaffold of fibrillin microfibrils. Tropoelastin is recruited to fibrillin through interactions with MAGP and fibulins, the latter of which may be recruited by LTBPs. LOX and LOXL enzymes are recruited to catalyse cross-linking of elastin and microfibrils.

1.9 Fibrillinopathies: inherited connective tissue disorders linked to fibrillin

MFS and congenital contractural arachnodactyly (CCA) were the first inherited connective tissues disorder to be associated with the fibrillins (Lee *et al.*, 1991), but others have since been identified. Collectively, they are called the fibrillinopathies and many are associated with mutations in *FBN1*. There is a possible link between *FBN3* and polycystic ovary syndrome, but it is still controversial (Jordan *et al.*, 2010; Raja-Khan *et al.*, 2010).

1.9.1 Marfan syndrome

MFS is an autosomal dominant condition with an incidence of about one in 5000 to 10000 people. MFS patients have a spectrum of symptoms, with defects in the skeletal, cardiovascular and ocular systems, among others. Their height is typically above average and they have unusually flexible joints (Cañadas *et al.*, 2010a). In order to be diagnosed with MFS, patients have to satisfy the Ghent criteria, with a minimum number of symptoms (Table 1.1). Family history of MFS and mutations in the *FBN1* gene are also taken into account (Faivre *et al.*, 2012). The major cause of death in MFS patients is dissection of the ascending aorta − the arterial wall weakens, resulting in separation of the outer layers and eventually rupture of the blood vessel (Cañadas *et al.*, 2010b). Despite the general link between *FBN1* and MFS, there is little correlation between genotype and phenotype. The only exception is neonatal MFS, in which symptoms are severe very early in life, resulting in death at a young age. Patients with neonatal MFS have mutations in exons 24-32, which encode the TB3-cbEGF18 region of fibrillin-1 (Stheneur *et al.*, 2011).

Organ system	Major criteria	Minor criteria
Skeletal system	1) Chest deformities 2) Unusual body dimensions 3) Unusual rotation of wrist or thumb 4) Spinal deformities 5) Reduced elbow extension 6) Collapsed foot arches 7) Defective hip socket	1) Chest deformities (no surgery required) 2) Joint hypermobility 3) High-arched palate 4) Characteristic facial appearance
Ocular system	Ectopia lentis − detachment of the lens	1) Unusual dimensions of the eye 2) Under-development of iris or ciliary muscle
Cardiovascular system	1) Dissection of the ascending aorta 2) Dilatation of the ascending aorta	1) Mitral valve prolapse − failure of mitral valve to close properly 2) Dilatation of pulmonary artery (age under 40) 3) Calcification of mitral annulus − part of mitral valve (age under 40) 4) Dissection or dilatation of thoracic or abdominal aorta annulus (age under 50)
Pulmonary system	None	1) Spontaneous pneumothorax − collapsed lung 2) Apical blebs − air-filled lesions
Skin	None	1) Stretch marks 2) Herniae
Dura (tissue surrounding brain and spinal cord)	Expansion of dural sac	None

Table 1.1 Clinical criteria for diagnosis of MFS. Modified from Cañadas *et al.*, 2010a. In order for a patient to be diagnosed and not have a family history of MFS, patients must have major symptoms in at least two of the systems and also have minor symptoms in a third. The diagnosis can be validated by screening for mutations in the *FBN1* gene.

1.9.2 Other diseases linked to the *FBN1* gene

While MFS mutations are spread across the *FBN1* gene, some diseases are linked with mutations that affect particular fibrillin-1 domains. All patients with stiff skin syndrome (SSS) tested so far have mutations affecting the TB4 domain (Loeys *et al.*, 2010). The SSS phenotype is very different from MFS − patients have a short stature, as well as hard and thick skin, resulting in reduced joint mobility; this is attributed to excessive microfibril deposition and reduced elastogenesis. The disease appears to result from a defective interaction of fibrillin with integrins α5β1 and αVβ3, resulting in abnormalities in the ECM. Other diseases affect the TB5 domain, encoded by exons 41 and 42. Some patients with autosomal dominant Weill-Marchesani syndrome (WMS) have a deletion in exon 41 of *FBN1* and have similar symptoms to SSS patients (Faivre *et al.*, 2003). Other patients with WMS have an in-frame deletion, resulting in loss of the TB1, Pro-rich and EGF4 domains that abolishes binding of ADAMTSL proteins to fibrillin (Sengle *et al.*, 2012). Patients with acromicric and geleophysic dysplasias (AD and GD) also have missense mutations in exons 41 and 42 and their symptoms are very similar to WMS, but without defects in the eye lens (Le Goff *et al.*, 2011). The pathogenic mechanisms for WMS, AD and GD are not yet well defined.

1.9.3 Mechanisms of Marfan syndrome pathogenesis

The broad spectrum of disease phenotypes and large spread of mutations across the *FBN1* gene makes it hard to uncover a single underlying mechanism for MFS pathogenesis. Types of mutation found in patients include missense and nonsense mutations, splice-site mutations causing exon-skipping, mutations causing nonsense-mediated decay of RNA, insertions and deletions (Faivre *et al.*, 2007). As the condition is inherited in an autosomal dominant fashion, possible pathogenic mechanisms can be classified into two broad types (Figure 1.10). In the first class, disease would result from a reduced number of microfibrils (haploinsufficiency). In the second class,

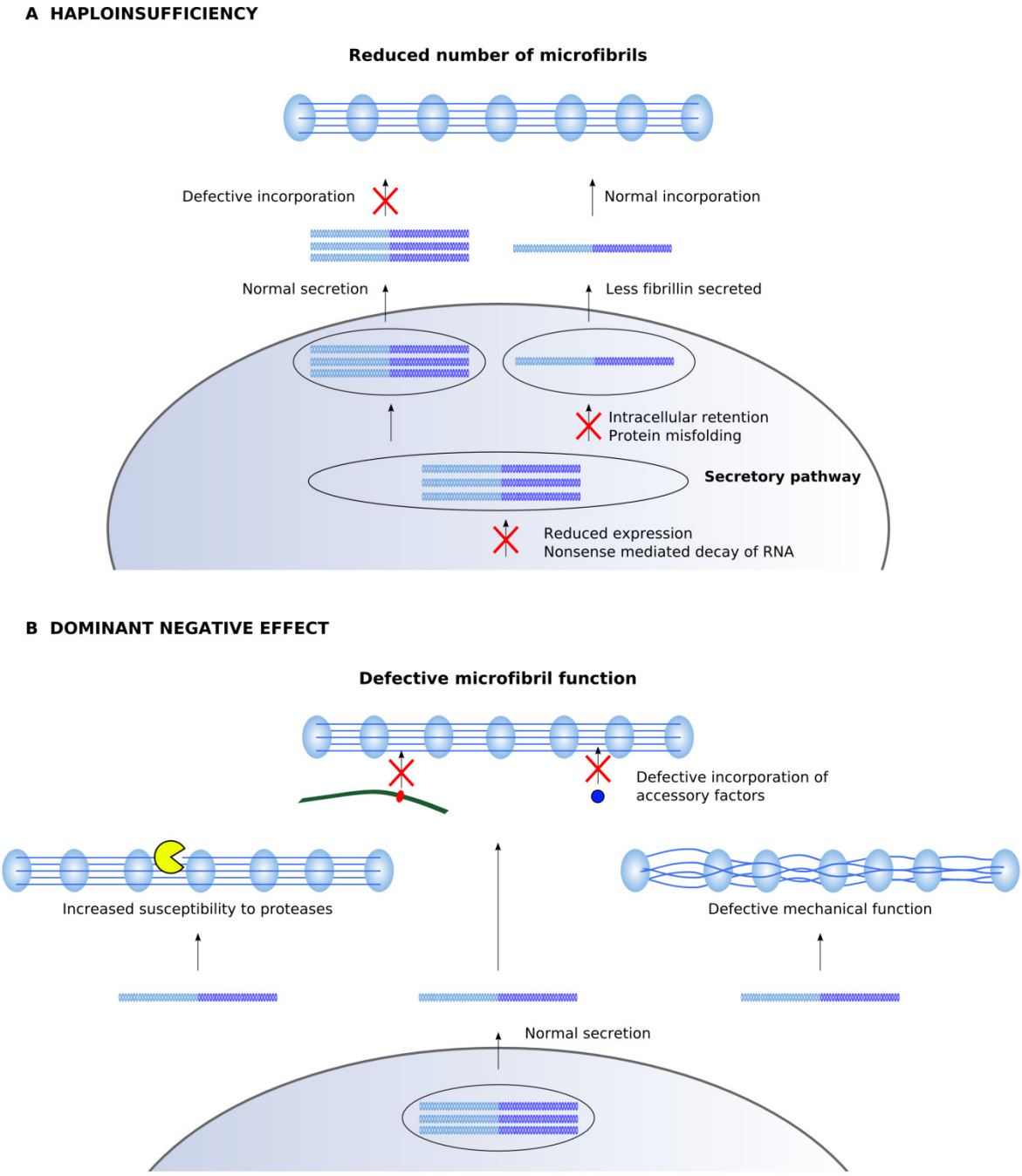


Figure 1.10 Mechanisms of MFS pathogenesis. A) Possible mechanisms of pathogenesis through haploinsufficiency, resulting in a reduced number of microfibrils in the ECM. Mutant fibrillin molecules are either secreted in small quantities or fail to incorporate into microfibrils. Normal fibrillin molecules are unaffected by these processes. B) Potential mechanisms resulting from a dominant negative effect, in which mutant fibrillin molecules are defective, can incorporate into microfibrils and impair the function of normal fibrillin molecules.

MFS would be caused by defective microfibril function (dominant negative). Fibrillin monomers could incorporate into microfibrils, but their presence would disrupt microfibril function (Hubmacher *et al.*, 2011).

The original view was that pathogenesis occurred through a dominant negative mechanism. Observations that patient tissues had a very low microfibril content (Hollister *et al.*, 1990), that extracted microfibrils had abnormal structure and that patient cells secreted normal amounts of fibrillin (Kielty *et al.*, 1994) appeared to be inconsistent with haploinsufficiency. In addition, stable expression of a mutant *FBN1* allele in a fibrillin-producing cell resulted in reduced microfibril production in cell culture (Eldadah *et al.*, 1995).

Most results from mouse models strongly suggest that haploinsufficiency is also a major contributor. Mice in which one *Fbn1* allele has a truncation (Pereira *et al.*, 1997) or is expressed at levels ~5-fold lower than normal (Pereira *et al.*, 1999) do not have any significant defects. However, mice homozygous for the mutant alleles have some MFS-like features, suggesting that a minimum amount of wildtype fibrillin is needed for normal microfibril function. Furthermore, expression of a human *FBN1* transgene with a MFS mutation on a wildtype mouse background does not cause disease, but mice with a similar mutation in one *Fbn1* allele have a MFS-like phenotype (Judge *et al.*, 2004). Nonetheless, haploinsufficiency is unlikely to be the only mechanism. A recently created mouse model, which produces a green fluorescent protein (GFP)-tagged form of truncated fibrillin that incorporated into microfibrils, has fragmented elastic lamellae (Charbonneau *et al.*, 2010). Although the C-terminal addition of GFP to fibrillin might also perturb microfibril structure, this suggests that abnormal fibrillin could disrupt microfibril function.

Several different categories of effects of MFS mutations on fibrillin secretion by cells in culture have been identified: no effect (normal secretion), reduced expression due to premature translation termination, complete intracellular retention and delayed secretion. Delayed secretion, identified by pulse-chase analysis, typically results from mutations that add or remove a cysteine residue. Where secretion is normal, reduced microfibril production is sometimes also observed. Some studies have looked at endogenous fibrillin in patient cells (Aoyama *et al.*, 1995; Halliday *et al.*, 1999; Schrijver *et al.*, 1999), whereas others used recombinant fibrillin fragments (Whiteman and Handford, 2003; Whiteman *et al.*, 2007). While examining endogenous fibrillin alone allows mutations to be studied in their native context, recombinant fragments avoid the complication of distinguishing secretion profiles of mutated and unaffected fibrillin molecules**.** Interestingly, levels of normal fibrillin were shown to vary between members of family with MFS who had premature termination codon mutations (Hutchinson *et al.*, 2003). This might explain why the symptoms of patients with the same mutation can vary in severity.

1.9.4 Role of TGFβ in disease pathogenesis

The early prevailing view that MFS resulted exclusively from mechanical and structural defects was challenged by the discovery of the involvement of TGFβ signalling. One mouse model of MFS has increased levels of active TGFβ in the lung, coinciding with defects in alveolar septation (Neptune *et al.*, 2003). The mice also have inflammation and increased expression of matrix metalloproteinases (MMPs). Furthermore, losartan, an angiotensin receptor antagonist that blocks TGFβ signalling, prevents aortic aneurysms in another mouse model (Habashi *et al.*, 2006). The importance of TGFβ was also highlighted by the discovery that Loeys-Dietz syndrome, which is associated with aortic aneurysms, is caused by mutations in the two TGFβ receptor-encoding genes (Loeys *et al.*, 2005). Elevated TGFβ levels are also found in the skin of patients with SSS

(Loeys *et al.*, 2010). The mechanisms of TGFβ action are not yet clear, but the cytokine clearly has a key role in matrix regulation, which is perturbed in disease states.

1.9.5 Structural effects of MFS-causing mutations

Several studies have examined the effects of MFS point mutations on fibrillin structure. Biochemical techniques and NMR spectroscopy can be used to assess the effects of residue substitutions on the fold of fibrillin fragments (Smallridge *et al.*, 1999; Suk *et al.*, 2004; Whiteman *et al.*, 1998; Whiteman *et al.*, 2001; Whiteman *et al.*, 2007). One of the most common changes resulting from missense mutations is the addition or removal of a cysteine residue (Whiteman *et al.*, 2006). Loss of a cysteine compromises structural integrity as it prevents formation of a disulphide bond and results in the presence of an unpaired cysteine. A cysteine not participating in the formation of a native disulphide bond could potentially interfere with the formation of other disulphides. However, such a substitution does not necessarily result in complete misfolding. Two mutations leading to C1977R and C1977Y substitutions only have local effects on interdomain packing in a fragment encompassing domains cbEGF29-cbEGF31 (Suk *et al.*, 2004). Other, more minor changes, such as the G1127S substitution in cbEGF13, may also only have local effects on structure (Whiteman *et al.*, 2001). However, G1127S does affect the efficiency of *in vitro* refolding.

Some missense mutations affect calcium binding, which is a critical indicator of cbEGF folding. Their effects can be observed by comparing calcium-binding affinities of wildtype and mutated fragments, as well as their susceptibility to proteolysis in the presence of calcium. Moreover, NMR can provide domain-specific information about calcium-binding affinities and folding by identifying and assigning key marker peaks in spectra at different $CaCl_2$ concentrations. As expected, substitutions removing a calcium-binding ligand result the in loss of calcium binding or

reduction in affinity at one site and reduced calcium-mediated protection from proteolysis. Some substitutions can also have long-range effects − perturbation of calcium binding in cbEGF13 affects folding of cbEGF12. This is dependent on the particular interdomain interactions at the cbEGF12-cbEGF13 interface − substitutions at equivalent positions in cbEGF19 only have local effects (Whiteman *et al.*, 2007).

The increased susceptibility to proteolysis resulting from some MFS mutations suggests a possible mechanism of pathogenesis (Reinhardt *et al.*, 2000b; Suk *et al.*, 2004; Whiteman *et al.*, 2007). Degradation of microfibrils by proteases could interfere with their mechanical and regulatory functions. However, many proteases tested *in vitro* are unlikely to have significant contact with microfibrils *in vivo*, if any. MMPs may play a role, as fibrillin fragments can promote MMP expression by fibroblasts in culture (Booms *et al.*, 2005). Neonatal MFS mutations also result in increased proteolytic degradation of fibrillin fragments by several MMPs (Kirschner *et al.*, 2011). However, it is still a matter for debate whether this is the primary pathogenic mechanism.

1.9.6 Marfan syndrome and microfibril assembly

A possible mechanism for functional haploinsufficiency is the failure of mutant fibrillin molecules to incorporate into microfibrils. Given the roles of the N- and C-terminal regions in assembly, it is tempting to hypothesise that mutations affecting these regions have this effect. This is supported by the failure of mutant fibrillin molecules with a C-terminal defective furin/PACE site to incorporate into the ECM (Section 1.7.1). A number of missense and nonsense mutations have also been identified in the minimal N- and C-terminal interacting regions (Figure 1.11). In the FUN-EGF3 region, there are several mutations that lead to the addition or removal of cysteines, but without a high resolution structure it is not possible to predict the effects of others (Figure 1.11A). There is some information about the functional effects of the substitutions: T101A

reduces heparin binding to an N-terminal fragment (Cain *et al.*, 2005) and N164S perturbs LTBP-4 binding (Ono *et al.*, 2009). In the cbEGF41-cbEGF43 region, the majority of substitutions are likely to affect calcium binding or interdomain packing, based on knowledge of other cbEGF structures (Figure 1.11B). There are no obvious structural explanations for the effects of two substitutions: Y2639N in cbEGF42 and S2660R in cbEGF43.

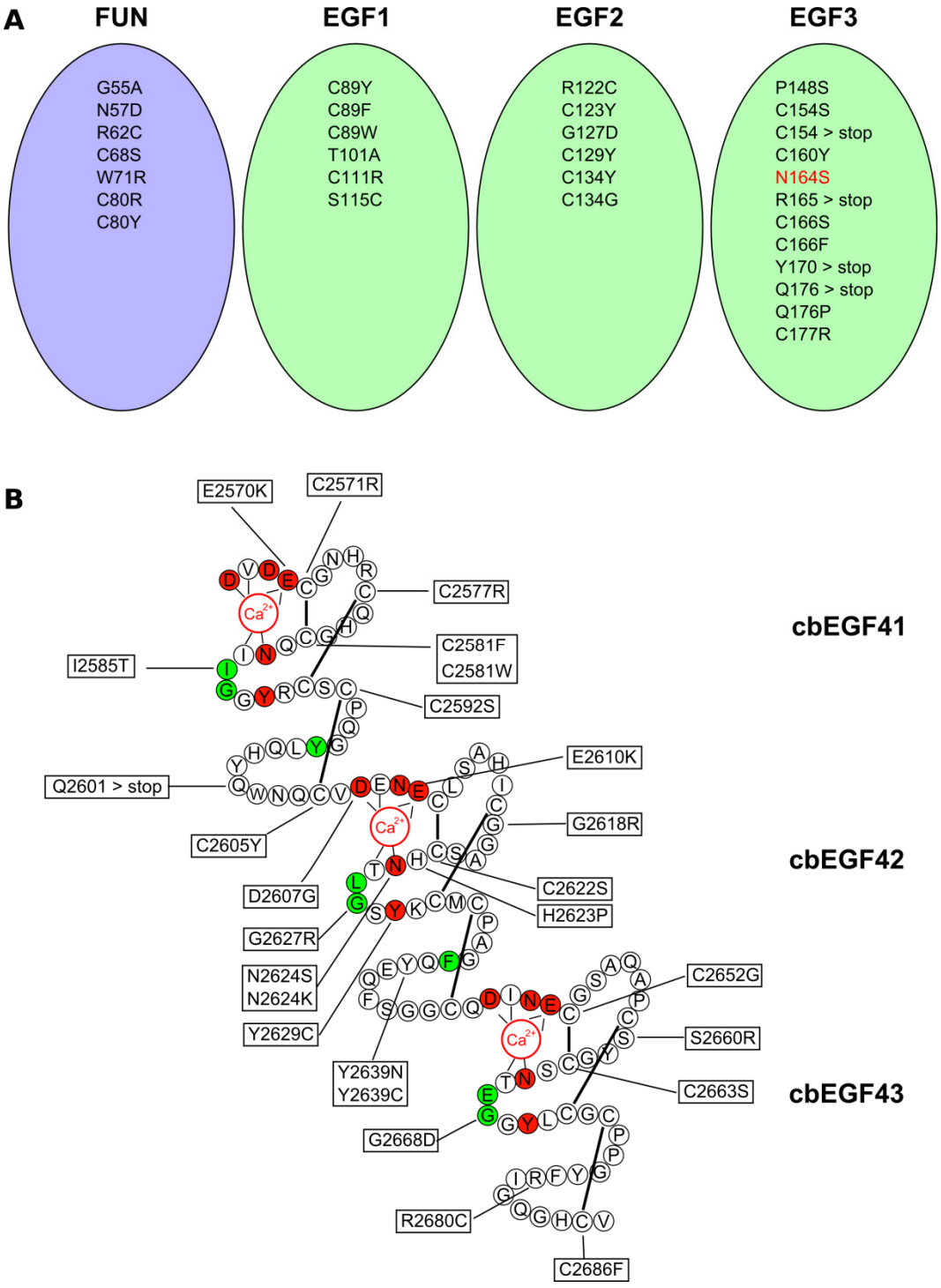


Figure 1.11 Disease-linked missense and nonsense mutations affecting the N- and C-terminal regions of human fibrillin-1. A) Effects of mutations on the FUN-EGF3 region. The structure of this region is not known, so single residue substitutions and nonsense mutations are grouped according to domain. The N164S substitution is highlighted because of its effect on LTBP-4 binding. B) Effects of mutations on the cbEGF41-43 region. Consensus calcium-binding residues and packing residues are highlighted in red and green respectively. In some cases, there is more than one mutation affecting one residue. References for mutations are given in Appendix A.

1.10 Thesis aim

The aim of this thesis is to characterise the interaction between the human fibrillin-1 N- and C-terminal domains that are important for microfibril assembly. While the minimal binding regions have been identified and narrowed to a small number of domains, there is no mechanistic information about how they interact. An important component of this work is structure determination of the FUN-EGF3 region, to characterise the fold of the FUN domain and the interdomain interfaces. Initial efforts to solve the FUN-EGF3 structure by crystallography were unsuccessful, but NMR spectroscopy showed initial promise (S. Jensen, I. Robertson, C. Redfield, P. Handford, unpublished data). Combining the structure with sequence conservation data will help to identify potentially important residues in the N-C interaction; their role can then be tested in interaction assays. The structure will also allow the prediction of effects of MFS missense mutations on the fold of the N-terminal domains. Effects of substitutions on the N-C interaction, folding and protein secretion can then be investigated. The structure is also relevant to the binding of accessory factors to fibrillin as many of them interact with the N-terminal domains.

The following areas will be covered:

1) Production of recombinant FUN-EGF3 and cbEGF41-cbEGF43 fibrillin-1 fragments and the establishment of interaction assays

2) Structure determination of a fragment corresponding to the FUN-EGF3 region and characterisation of its dynamics by NMR spectroscopy

3) Dissection of the N-C interaction, including structure- and sequence conservation-informed mutagenesis

4) Investigation of the effects of several MFS mutations on domain structure, the N-C interaction and protein secretion by fibroblast cells.

These results will provide the first insights into the structure of the fibrillin FUN-EGF3 region, as well as its interaction with the C-terminus and the effects of several MFS mutations.

Chapter 2

Materials and methods

2.1 DNA techniques

2.1.1 Polymerase chain reaction (PCR)

PCRs were carried out in 0.5 ml thin-walled PCR tubes or 0.2 ml thin-walled strip-tubes (Anachem). *Pfu* (Agilent) and *Taq* (Roche) DNA polymerases were used for insert amplification and colony screening respectively. For insert amplification, the PCR was carried out in a total volume of 50 µl with following final reagent concentrations: 1 x *Pfu* buffer (Agilent), 200 µM deoxyribonucleotides (dNTPs) and 150 nM of both forward and reverse primers. In addition, 1 ng DNA template and 1.25 U *Pfu* polymerase were added. Primers were designed to have a melting temperature (T_m) of ~60°C. T_m was calculating using the formula:

$$T_m = 2 \times (A + T) + 4 \times (C + G) \qquad 2.1$$

where A, T, C and G are the numbers of each base in the DNA molecule. Individual primer sequences are given in Appendix B. Primers (Invitrogen) were received in lyophilised form, dissolved in milliQ-purified water and stored at −20°C. 15 µM stock solutions were made immediately prior to their use in PCRs. PCRs were carried out in a Biometra T3 Thermocycler. The PCR cycle steps are summarised in Table 2.1.

Table 2.1 Temperatures and times of PCR steps used in insert amplification

Temperature (°C)	Time (s)	
95	180	
95	60	x 30
55	30	
72	120	
72	600	

10 µl of the PCR product was analysed by agarose gel electrophoresis (Section 2.1.3). The remainder was ethanol precipitated and dissolved in 30 µl milliQ-purified water for gel purification (Section 2.1.3) or restriction enzyme digestion (Section 2.1.5).

2.1.2 Ethanol precipitation

DNA was precipitated by the addition of two volumes of ice-cold 100% ethanol and 0.1 volumes of 3 M sodium acetate pH 5.0. The solution was incubated on ice for 10 minutes before being centrifuged in a bench-top centrifuge (13000 revolutions per minute [rpm]; 10 min; 25°C) to pellet the DNA. The supernatant was discarded and the pellet was washed with 70% (v/v) ethanol, before being air-dried (10 min; 25°C). The DNA pellet was then dissolved in an appropriate volume of milliQ-purified water or 10 mM Tris(hydroxymethyl)aminomethane (Tris)-HCl pH 8.0, 1 mM EDTA, depending on the subsequent application.

2.1.3 Agarose gel electrophoresis

Agarose gels comprising 1-2% (w/v) agarose with 2 µg/ml ethidium bromide were prepared, depending on the expected sizes of DNA molecules under analysis. Agarose powder (MP Biomedicals) was dissolved in 1 x TBE buffer (90 mM Tris, 90 mM boric acid, 2.5 mM EDTA), heated until the agarose dissolved, poured and left to set at room temperature, forming the gel. DNA samples were prepared in 1 x TBE buffer also containing 0.25% (w/v) bromophenol blue (BDH) and 30% (v/v) glycerol (Sigma-Aldrich) and loaded onto the agarose gel. The DNA samples were run alongside standard DNA size-markers: either a 0.5 kilobase (kb) or 1 kb ladder (New England Biolabs). Electrophoresis was performed using a Mini-Sub DNA Cell electrophoresis system (Biorad) with 1 x TBE buffer. A voltage of 80 V was applied for 45 min, or until the DNA

molecules had been sufficiently resolved. Bands on the gel were imaged using an ultraviolet (UV) transilluminator and video copy processor (UVItech).

2.1.4 Gel purification

Bands were identified on the agarose gel using a hand-held UV-transilluminator (UVItech) to minimise DNA damage. Thin slices were cut from the gel using a razor blade. DNA was purified from gel slices using the Qiagen gel purification kit, according to the manufacturer's instructions.

2.1.5 Restriction enzyme digestion of DNA

Reactions were carried out in a total volume of 30 µl with 1 to 2 µg of gel-purified DNA and 0.5-1.0 units of enzyme per µg DNA. Compatible buffers and incubation temperatures were chosen according to the manufacturer's instructions. Digested inserts were purified using the Qiagen PCR purification kit and digested vectors were gel-purified (Section 2.1.4).

2.1.6 DNA Ligation

Purified, digested inserts and vectors were ligated using the rapid DNA ligation kit (Roche) according to the manufacturer's instructions in a total volume of 20 µl. 2 µl of the reaction mixture was transformed into 100 µl *recA*- NM554 *Escherichia coli* (*E. coli*) cells (Raleigh *et al.*, 1988) also containing the pREP4 plasmid (Qiagen) (Section 2.1.13). A negative control reaction, without added insert, was also transformed into cells to confirm that the vector DNA had been completely digested.

2.1.7 PCR colony screen

Single colonies were picked using a flame-sterilised loop and transferred to 50 µl milliQ-filtered water. Picked colonies were screened for the insert by PCR (Section 2.1.1). 24 µl of a stock mix containing the components required for PCR was added to 6 µl bacterial suspension. The reagent concentrations were: 1 x PCR buffer (Roche), 200 µM dNTPs and 100 µM of each primer. 1 U *Taq* polymerase was also added per 30 µl reaction mix. The temperatures and times of PCR steps are summarised in Table 2.2.

Table 2.2 Temperatures and times of PCR steps used in colony screening

Temperature (°C)	Time (s)	
95	180	
95	60	
45	30	x 35
72	120	
72	600	

On completion of the PCR, 6 µl 6 x DNA loading buffer (Section 2.1.3) was added to the reaction mixture. 12 µl was of this mixture was loaded and separated on an agarose gel (Section 2.1.3). The sizes of PCR products, indicated by their mobility on the gel relative to size markers, were checked to confirm the presence of the correct insert. Colonies containing the insert were streaked onto 2 x yeast tryptone (YT; MP Biomedicals), 1.5% (w/v) agar (Formedium) plates containing appropriate antibiotics. Single colonies were then picked from these plates and cultured for subsequent DNA extraction and purification (Section 2.1.8).

2.1.8 DNA amplification

DNA was purified from 100 ml bacterial culture using the Qiagen DNA Midi kit. The kit uses the alkaline lysis method of extraction (Birnboim and Doly, 1979). Cells were harvested by

centrifugation (Beckman Rotor JLA 10.5; 10 min; 5000 rpm; 4°C). DNA was extracted and purified according to the manufacturer's instructions.

2.1.9 DNA quantitation

The absorbance of DNA solutions was measured at 260 nm (A_{260}) in a 100 µl quartz cuvette (Hellma) using a Shimadzu UV-mini 1240 spectrophotometer. For the measurement, the solution was diluted by an appropriate factor in a total volume of 100 µl using water or buffer. The concentration of DNA in mg/ml was calculated by dividing the A_{260} value by 20. In addition, A_{260}/A_{280} was used to assess the purity of the DNA preparation. Where appropriate, purified DNA was sent for sequencing (Section 2.1.10) and stored at −20°C.

2.1.10 Site-directed mutagenesis

Site-directed mutagenesis of pQE-30 and pKG-52 vectors was performed using the QuickChange Lightning site-directed mutagenesis kit (Agilent), according to the manufacturer's instructions. Complementary forward and reverse primers were designed to have a T_m of ≥ 78°C, calculated from the sequence as follows:

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \%\ \text{mismatch} \qquad 2.2$$

where %GC is the percentage of guanine and cytosine-containing nucleotides, N is the total number of nucleotides in the sequence and % mismatch is the percentage of mismatched nucleotides. Primer sequences are given in Appendix B.

2.1.11 DNA sequencing

DNA sequencing was performed by Source Bioscience Oxford. For pQE-30 vectors, stock pQE-30 forward and reverse primers were used. For pKG-52 vectors, custom primers were used for sequencing (Appendix B).

2.1.12 Phenol-chloroform extraction

Plasmid DNA for transfection of fibroblast cells was purified further by phenol-chloroform extraction. 100 µl of DNA solution was added to 100 µl 1 x restriction enzyme buffer 4 (New England Biolabs). 200 µl of a phenol:chloroform:isoamyl alcohol mixture (25:24:1; v:v:v), saturated with 10 mM Tris-HCl pH 8.0 1 mM EDTA (Sigma-Aldrich), was added to the DNA-buffer mixture. The resulting mixture was vortexed and centrifuged (bench-top centrifuge; 1 min; 13000 rpm; 25°C) to separate the aqueous and non-aqueous phases. The aqueous phase (top layer) was removed and added to another 200 µl of the phenol:chloroform:isoamyl alcohol mixture in a fresh tube. The process was repeated and the aqueous phase was added to 200 µl chloroform (Sigma-Aldrich). Again, the mixture was vortexed and centrifuged, and the aqueous layer was removed. DNA remaining in the chloroform-buffer mix was recovered through the addition of 200 µl 1 x buffer, followed by vortexing, centrifugation and removal of the aqueous phase. DNA was extracted from the resulting ~400 µl solution by ethanol precipitation (Section 2.1.2).

2.1.13 Production of competent cells

Stocks of competent cells *E. coli* NM554 and BL21 cells were prepared. NM554 cell stocks were made with and without the pREP4 plasmid, while only stocks of BL21 cells containing pREP4 were used. Existing laboratory glycerol stocks were streaked onto 2 x YT-agar plates, with the addition

of 25 µg/ml kanamycin (Sigma-Aldrich) for use with cells containing the pREP4 plasmid. 3 ml autoclaved 2 x YT medium (with kanamycin if appropriate) in a sterile 15 ml tube was inoculated with a single colony and cultured for at least 4 hr at 37°C, shaking at 200 rpm. 100 ml 2 x YT medium (with kanamycin if appropriate) was then inoculated with 1 ml starter culture and cultured until the absorbance at 600 nm (A_{600}) of the suspension reached ~0.5. Cells were harvested by centrifugation in pre-cooled centrifuge tubes (Beckman Rotor JLA 10.5; 10 min; 5000 rpm; 4°C). The medium was discarded and the cell pellet was gently resuspended in 50 ml transformation buffer (10 mM PIPES [piperazine-N,N'-bis(2-ethanesulfonic acid)] pH 6.6, 50 mM $CaCl_2$, 15% [v/v] glycerol) and incubated on ice for 30 min. Cells were centrifuged again and the pellet was resuspended in 5 ml transformation buffer. The suspension was aliquoted, frozen rapidly in liquid nitrogen and transferred to a −80°C freezer.

2.1.14 Transformation of competent *E. coli* cells

pQE-30 vectors were transformed into NM554 (pREP4) or BL21 (pREP4) cells for cloning and protein expression purposes respectively. For amplification of pKG-52 vectors, DNA was transformed into NM554 cells lacking the pREP4 plasmid. Competent cell stocks were first thawed on ice. 100 µl of the cell suspension was added to the DNA solution, mixed and incubated on ice for 15 min. The cells were then heat-shocked at 42°C for 90 s and replaced on ice. 100 µl sterile 2 x YT medium was added to the cells, before incubation at 37°C for 1 hr. 100 µl of the cell suspension was then plated onto 2 x YT agar plates containing 100 µg/ml ampicillin (Sigma-Aldrich). For cells containing the pREP4 plasmid, plates also contained 25 µg/ml kanamycin. The plates were inverted, incubated at 37°C overnight and stored at 4°C for up to one week.

2.2 Protein expression, purification, refolding and characterisation

Details of purification methods are given for proteins described in subsequent chapters. FUN-EGF3, cbEGF41-43 with a C-terminal BirA tag (cbEGF41-43-BirA) and cbEGF32-33 with a C-terminal BirA tag (cbEGF32-34-BirA) are introduced in Chapter 3. FUN-EGF1, EGF2-EGF3 and EGF2-EGF3 with a C-terminal BirA tag (EGF2-EGF3-BirA) are introduced in Chapter 5.

2.2.1 Protein expression in *E. coli*

250 ng pQE-30 plasmid DNA was transformed into BL21 (pREP4) cells (Section 2.1.14). Single colonies were used to inoculate 100 ml autoclaved 2 x YT medium containing 25 µg/ml kanamycin and 100 µg/ml ampicillin. The cultures were incubated overnight at 37°C, shaking at 200 rpm. 20 ml starter culture was added to 500 ml sterile 2 x YT medium with antibiotics and incubated at 37°C with shaking until the A_{600} of the cell suspension reached ~0.8. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to the cells at a final concentration of 2 mM to induce protein expression. Cells were incubated at 37°C with shaking for ~5 hr before harvesting by centrifugation (Beckman Rotor JLA 10.5; 10 min; 5000 rpm; 4°C). Cell pellets were stored at −80°C until needed.

2.2.2 Isotopic labelling for NMR spectroscopy

Cells were grown in M9 minimal medium for production of isotopically labelled protein. The medium comprised 0.6% (w/v) Na_2HPO_4, 0.3% (w/v) KH_2PO_4, 0.1% (w/v) NH_4Cl, 0.05% (w/v) NaCl, 1 x yeast nitrogen base (Formedium), 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.5% (w/v) D-glucose, 50 mg/l EDTA, 5 mg/l $FeCl_3$, 0.5 mg/l $ZnCl_2$, 0.1 mg/l $CuCl_2$, 0.1 mg/l $CoCl_2.6H_2O$ and 0.1 mg/l H_3BO_3. M9 salts (Na_2HPO_4, KH_2PO_4 and NaCl) were prepared as a 10 x stock and autoclaved. *E. coli* trace

elements (EDTA, $FeCl_3$, $ZnCl_2$, $CuCl_2$ and $CoCl_2.6H_2O$) were prepared as a 100 x stock by first dissolving the EDTA in water, adjusting the pH to 7.0 and then adding the other components, adjusting the pH with every addition. The 100 x stock was autoclaved. Stocks of all other chemicals were filter-sterilised using a 0.22 µm syringe filter (Millipore). Chemical stocks were diluted in autoclaved, deionised water to prepare the medium.

Labelling was achieved by replacing the unlabelled components with $^{15}NH_4Cl$ (99% labelled; Cambridge Isotope Laboratories) and ^{13}C-D-glucose (99% labelled; Cambridge Isotope Laboratories). The method for protein expression was the same as for unlabelled material (Section 2.2.1), with the following exceptions: first, 50 ml starter culture, grown in M9 minimal medium, was added to ~600 ml M9 minimal medium. After the addition of IPTG, cells were incubated for at least 7 hr before harvesting.

2.2.3 Nickel affinity purification of reduced protein

Frozen cell pellets (Section 2.2.1) were thawed and resuspended in lysis buffer (50 mM $NaPO_4$ pH 7.5, 5 mM β-mercaptoethanol, 6 M guanidine-HCl). Guanidine-HCl powder was added to keep the concentration at 6 M during lysis. Once the cells had been completely lysed, the mixture was sonicated for 90 s at 20 W (Jencons Ultrasonic Processor) to reduce the viscosity of the solution. The insoluble material was then pelleted by ultracentrifugation (Beckman 60Ti rotor; 30 min; 40000 rpm; 25°C). Following ultracentrifugation, the supernatant was passed through a Ni(II)-affinity column. The column was prepared by equilibrating ~3 ml of chelating fast-flow sepharose resin (GE Healthcare) with 1 M $NiSO_4$, washing with milliQ-purified water and then equilibrating with lysis buffer.

The supernatant from ultracentrifugation was passed through the column to allow His_6-tagged protein to bind. The column was then washed with ~5 column volumes lysis buffer and the protein was eluted in 20 ml lysis buffer with EDTA added to a final concentration of 50 mM. Tris-HCl pH 8.3 and dithiothreitol (DTT; Fisher) were then both added to final concentrations of 100 mM and incubated at room temperature for 1 hr. The pH was adjusted to ~2 using HCl and the solution was dialysed against 2 l 0.1% (v/v) trifluoroacetic acid (TFA) overnight using 16 mm dialysis tubing with a 7000 Da molecular weight cut-off (Medicell). Dialysis was necessary to remove DTT and guanidine before desalting by reverse phase high performance liquid chromatography (HPLC; Section 2.2.4).

Some proteins expressed in bacteria grown in M9 minimal medium required additional purification steps. During the purification of FUN-EGF3 and cbEGF41-43-BirA proteins it was noted that a large fraction of recombinant protein remained in the pellet of insoluble material after ultracentrifugation. In order to extract more protein and improve the yield, the pellet was resuspended in 30 ml lysis buffer. After an addition of 10 µl β-mercaptoethanol (increase in concentration of 0.3 mM), the pH of the solution was raised to ~9.0 using NaOH and the mixture was vortexed. The pH was then lowered to pH ~7.5 using HCl and the solution was centrifuged again. The supernatant was then pooled with the solution from the first ultracentrifugation step and loaded onto the Ni(II)-affinity column.

2.2.4 Reverse phase HPLC

Purification was performed using a reverse phase C8 semi-preparative column (Dynamax) on a Beckman Gold system. The sample pH was adjusted to ~2 and the solution was filtered where necessary before loading onto the column. A 20-80% gradient of buffer B was applied over 40 min, where A was 0.1% TFA and B was 0.1% TFA, 80% acetonitrile. The absorbance of the solution

exiting the column was measured at 280 nm (A_{280}), in order to follow the elution profile. Fractions were collected manually, frozen rapidly in liquid nitrogen and lyophilised under vacuum using an E-C Modulyo freeze dryer.

2.2.5 Prediction of protein properties and concentration measurement by spectrophotometry

The expected molecular weight, estimated extinction coefficient and estimated isoelectric point were calculated from the primary sequence using a Python script (http://www.python.org/). Extinction coefficients and molecular weights for all proteins described in this thesis are summarised in Appendix C. The molecular weight for unlabelled proteins was calculated using the following relative atomic masses: 1.00794 for H, 12.0107 for C, 14.0067 for N, 15.9994 for O and 32.0650 for S. Masses for N and C were adjusted to 15 and 13 for ^{15}N and ^{13}C labelling respectively. The extinction coefficient for absorbance of light with a wavelength of 280 nm (ε_{280}) was estimated using the method of Gill and von Hippel (Gill and von Hippel, 1989). Extinction coefficients of tryptophan (5500 M^{-1} cm^{-1}), tyrosine (1490 M^{-1} cm^{-1}) and cystine (125 M^{-1} cm^{-1}) used in the calculations were as described for the Expasy Protparam program (http://web.expasy.org/protparam/protparam-doc.html). The isoelectric point was estimated by using the bisection method. Acid dissociation constants used in the calculation were for residues in peptides in 9.8 M urea and 25°C (Bjellqvist *et al.*, 1994).

Before the concentration measurement, lyophilised protein was dissolved in milliQ-purified water. The A_{280} of the protein solution, diluted by an appropriate factor in a total volume of 100 µl, was measured using a Shimadzu UV-mini 1240 spectrophotometer as described for DNA (Section 2.1.9). The protein concentration was calculated using the Beer-Lambert law:

$$Concentration\ (M) = \frac{A_{280}}{\varepsilon_{280}} \times l \qquad 2.3$$

where l is the path length of the cuvette (1 cm). The concentration in g/l was calculated by multiplying the concentration in M by the molecular weight of the protein in g/mol.

2.2.6 *In vitro* protein refolding

HPLC-purified, reduced protein was refolded *in vitro* using an oxido-shuffling system as described previously (Knott *et al.*, 1996). In all cases, protein was refolded at a concentration of 0.2 mg/ml and in the presence of 3 mM cysteine and 0.3 mM cystine. Additional conditions varied between proteins. For N-terminal constructs (FUN-EGF3, FUN-EGF1, EGF2-EGF3 and EGF2-EGF3-BirA) the additional conditions were 100 mM Tris-HCl pH 8.3 and 50% (v/v) glycerol. For cbEGF41-43-BirA, the conditions were 100 mM Tris-HCl pH 9.0, 50% (v/v) glycerol and 50 mM $CaCl_2$. For cbEGF32-34-BirA, the conditions were 100 mM Tris-HCl pH 8.3 and 50 mM $CaCl_2$. Control refolding reactions for cbEGF32-34-BirA and cbEGF41-43-BirA contained 10 mM EDTA instead of 50 mM $CaCl_2$. Components were added in the following order: water, glycerol (where appropriate), Tris-HCl, $CaCl_2$ or EDTA (where appropriate), cysteine, protein and cystine. The solution was stirred slowly as the components were added and incubated at 4°C or 48-72 hr.

The refolding solution was then acidified to pH ~2, in preparation for loading onto the HPLC column. Refolding mixtures containing glycerol were dialysed against 0.1% (v/v) TFA for at least 5 hr (~2 l per 100 ml refolding mixture). The dialysed solution was concentrated using Vivaspin 20 concentrator columns with a 5 or 10 kDa molecular weight cut-off (Sartorius) by centrifugation (Eppendorf 5702 R centrifuge; 30 min; 3000 relative centrifugal force [RCF]; 25°C). Protein was purified from the resulting solution by reverse phase HPLC (Section 2.2.4).

2.2.7 His_6 tag removal using factor Xa

All proteins expressed from pQE-30 vectors contained a factor Xa cleavage site between the N-terminal hexa-histidine (His_6) tag and the native sequence. Lyophilised protein was first dissolved in milliQ-purified water and the concentration was measured (Section 2.2.5). To remove the His_6 tag, 1 U factor Xa (Novagen) was used per mg protein. Different conditions were used for the various constructs. For N-terminal constructs, buffer conditions were 20 mM 2-(N-Morpholino) ethanesulfonic Acid (MES) pH 6.0, 100 mM NaCl and the protein concentration was ~5 mg/ml. The pH was measured and raised to ~5.8 using HCl if required, before adding factor Xa. For cbEGF32-34-BirA, buffer conditions were 50 mM Tris-HCl pH 7.4, 100 mM NaCl and the protein concentration was ~2 mg/ml. For cbEGF41-43-BirA, buffer conditions were 50 mM Tris-HCl pH 8.3, 100 mM NaCl and the protein concentration was 1 mg/ml. Reactions were incubated at 37°C overnight.

2.2.8 Ion exchange fast protein liquid chromatography (FPLC)

N-terminal constructs were purified further using a MonoS 5/50 GL cation exchange column (GE Healthcare) on a UPC900/P900 ÄKTA basic system. Up to 1 ml of the factor Xa digest solution (Section 2.2.7) was loaded onto the column, following centrifugation or filtration to remove precipitate. His-tagged proteins were dissolved in milliQ-water and 0.5 ml of the solution was added to 0.5 ml buffer A (20 mM MES pH 6.0), before loading onto the column. Protein was eluted on a 0 to 50% gradient of buffer B (20 mM MES pH 6.0, 2 M NaCl) over 20 min, monitored by A_{280} of the solution exiting the column. Fractions were collected manually and then acidified in preparation for a final HPLC purification step (Section 2.2.4).

cbEGF32-34-BirA and cbEGF41-43-BirA were purified using a MonoQ 5/50 GL anion exchange column (GE Healthcare). For cbEGF32-34-BirA, 0.5 ml of the factor Xa digest solution was added to 0.5 ml buffer A (50 mM Tris-HCl pH 7.4, 1 mM $CaCl_2$) before loading onto the column. For cbEGF41-43-BirA, buffer A (50 mM Tris-HCl pH 9.5, 1 mM $CaCl_2$) was added to the factor Xa digest mixture in a ratio of 2:1 to prevent precipitation upon mixing. The protein was then concentrated using Vivaspin centrifugal concentrator columns (Section 2.2.6) before loading onto the MonoQ column. Protein was eluted on a 0 to 50% gradient of buffer B (50 mM Tris-HCl pH 7.4, 1 mM $CaCl_2$, 1 M NaCl for cbEGF32-34-BirA; 50 mM Tris-HCl pH 9.5, 1 mM $CaCl_2$, 1 M NaCl for cbEGF41-43-BirA), acidified and purified further by HPLC (Section 2.2.4).

2.2.9 Site-specific biotinylation of BirA-tagged proteins

Proteins expressed from the pQE-30-BirA vector had a 15-residue C-terminal recognition sequence for the BirA enzyme, a variant of the sequence identified by Schatz (Schatz, 1993). Site-specific biotinylation was carried out using the biotin-protein ligase enzyme kit (Avidity LLC). The substrate concentration in the biotinylation reaction was 40 µM. 1 µg BirA enzyme was used per 50 nmol substrate. Biomix-A (0.5 M Bicine buffer, pH 8.3) and Biomix-B (100 mM ATP, 100 mM magnesium acetate, 500 µM biotin) were both added to the solution, diluted by a factor of 10. The reaction was incubated at room temperature overnight, acidified and desalted by HPLC (Section 2.2.4). Biotinylation was confirmed by mass spectrometry (Section 2.2.11).

2.2.10 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by size using SDS-PAGE according to an established method (Laemmli, 1970). Polyacrylamide gels with 1.5 mm thickness were prepared from 29:1 acrylamide:bisacrylamide (mol:mol; Sigma-Aldrich) using the Mini-Protean III system (Biorad). The

separating gel comprised 10 or 16% (w/v) acrylamide:bisacrylamide (w/v), 0.1% (w/v) sodium dodecyl sulphate (SDS) and 0.38 M Tris-HCl pH 8.8. The percentage of acrylamide:bisacrylamide depended on the size range of proteins under analysis. Ammonium persulphate (APS) and *N*,*N*,*N*',*N*'-Tetramethylethylenediamine (TEMED) were added at final concentrations of 0.1% (w/v) and 0.1% (v/v) respectively, to polymerise the gel. A stacking gel was poured on top of the separating gel; the composition was 4% (w/v) acrylamide:bisacrylamide, 0.1% (w/v) SDS and 0.25 M Tris-HCl pH 6.8. APS and TEMED were used to polymerise the gel, at the same concentrations as for the separating gel.

Purified protein samples were diluted to 0.1 mg/ml in 2 x loading buffer and incubated at 95°C for 10 min. The buffer comprised 10% (v/v) glycerol, 1.5% (w/v) SDS, 50 mM Tris pH 6.8 and 0.2% (w/v) bromophenol blue. Samples were loaded onto the gel, assembled in the Mini Protean III apparatus (Biorad) and equilibrated with running buffer (0.1% [w/v] SDS, 30 mM Tris-HCl, 250 mM glycine pH 8.3). A potential difference of 180 V was applied across the gel for 50 min, or until the bromophenol blue had reached the bottom of the gel. The gel was stained with a solution comprising 0.125% (w/v) Coomassie R250 brilliant blue (BDH), 10% (v/v) methanol, 10% acetic acid, 80% (v/v) water for 1 hr at room temperature. Destaining was achieved through overnight incubation in 10% (v/v) methanol, 10% (v/v) acetic acid.

2.2.11 Mass spectrometry

Protein samples were prepared at a concentration of 10 µM in 50% (v/v) acetonitrile, 50% (v/v) water and 0.1% (v/v) formic acid in a total volume of 100 µl. Samples were loaded onto a Micromass LCT electrospray ionisation time of flight (ESI-TOF) mass spectrometer using a syringe pump (KD Scientific). The machine was calibrated by Dr David Staunton (Biophysical Instrumentation Facility, Department of Biochemistry) using a horse heart myoglobin sample.

Data were acquired and analysed using MassLynx v4.0 software (Waters). Spectra from 100 scans were combined to produce a single mass/charge (m/z) spectrum. Maximum entropy (MaxEnt) deconvolution was used to obtain the most likely mass spectrum from the data, with a resolution of 2 Da. Following smoothing and peak centering, the combined data were also analysed by the manual peak-picking method, using the m/z values of the major peaks of a charge series to calculate an average mass. The MaxEnt method was found to give more accurate results (data not shown). Therefore, every mass spectrometry result is reported as the mass of the major peak in a MaxEnt-deconvoluted spectrum.

2.3 Limited proteolysis

Protein was digested using Glu-C endoproteinase from *Staphylococcus aureus* V8 protease (Sigma-Aldrich). Lyophilised Glu-C was dissolved in water to give a concentration of 0.5 mg/ml and stored at −20°C. Conditions were optimised to ensure that the extent of proteolysis was incomplete, but sufficient to show differences between sets of conditions. HPLC-purified, lyophilised protein in water was diluted to 0.5 mg/ml in a solution of 50 mM Tris pH 9.5 with either 50 mM $CaCl_2$ or 10 mM EDTA. Since the two solutions have approximately the same ionic strength (*I*), no NaCl was added. *I* is calculated as follows:

$$I = \frac{1}{2} \sum c.z^2 \qquad 2.4$$

where *c* and *z* are the concentrations and charge of each ion respectively. The ionic strength of a 50 mM $CaCl_2$ solution is 150 mM, while the ionic strength of a 10 mM EDTA solution is 160 mM. A sample of the protein solution was removed before the addition of Glu-C at a final concentration of 0.05 mg/ml. Solutions with or without enzyme were incubated at 37°C for 4 hr. 15 µl sample was taken from the enzyme-containing solution at 1, 2 and 4 hr time points. Samples were added

immediately to 15 µl 2 x reducing SDS-PAGE sample buffer containing 10% (v/v) β-mercaptoethanol and boiled (Section 2.2.10). They were separated on a 16% SDS-PAGE gel (Section 2.2.10).

2.4 Pull-down interaction assay

Biotinylated, BirA-tagged proteins were immobilised on magnetic beads, which were then used to pull down potential interaction partners from solution (Section 2.4.1). Analysis was performed by SDS-PAGE and Western blotting (Section 2.4.2).

2.4.1 Pull-down

M270 streptavidin-coated magnetic beads (Invitrogen) were washed three times with Tris-buffered saline (TBS) pH 9.5, comprising 50 mM Tris-HCl pH 9.5 and 150 mM NaCl. 40 µl bead suspension was used per pull-down. Washes were performed by pelleting the beads by placing tubes on a magnetic rack, removing the supernatant and resuspending the beads in 40 µl buffer. The beads were then equilibrated with binding buffer, comprising TBS with 1% (w/v) bovine serum albumin (BSA) and 0.05% (v/v) Tween-20 through three washes. 5 µg biotinylated protein was then added to 40 µl bead suspension and incubated on a rotator (30 min; room temperature). The beads were washed three more times with binding buffer pH 9.5.

For experiments using immobilised fibrillin N-terminal fragments or appropriate BirA-tagged controls, 10 µg His-tagged protein was then added to beads and incubated on a rotator (1 hr; room temperature). The beads were washed 7 times with TBS-T (TBS with 0.05% [v/v] Tween-20) pH 9.5. 40 µl 2 x Laemmli sample buffer containing 10% (v/v) β-mercaptoethanol was then added to the beads and the sample was boiled (Section 2.2.10). The beads were pelleted, and the

supernatant was taken for analysis by SDS-PAGE and Western blotting (Section 2.4.2). For experiments using immobilised C-terminal fibrillin fragments or appropriate BirA-tagged controls, beads were washed three times with binding buffer pH 7.4 following the three washes with binding buffer pH 9.5. 15 µg His-tagged protein was then added to the beads and incubated, as described for the experiments with immobilised N-terminal fragments. The beads were then washed with TBS-T pH 7.4 and prepared for SDS-PAGE, as described above.

2.4.2 Western blotting

Pull-down supernatant samples were separated by SDS-PAGE using a 16% separating gel (Section 2.2.10). Proteins were then transferred from the gel to a nitrocellulose membrane (Whatman Protran) by electrophoresis using a Mini Trans-Blot Cell (Biorad). The transfer buffer comprised 25 mM Tris, 190 mM glycine, 20% (v/v) methanol and 0.02% (w/v) SDS, and was cooled to 4°C before use. The transfer cell was incubated on ice and a potential difference of 100 V was applied for 90 min. Following the transfer, the nitrocellulose membrane was incubated with blocking buffer for at least 1 hr at room temperature with gentle shaking. The blocking buffer was 5% (w/v) dried milk (Sigma-Aldrich) in TBS-T pH 7.4 for membranes subsequently blotted with streptavidin-horse radish peroxidase (HRP; Sigma-Aldrich). For the anti-RGS-His-HRP antibody (Qiagen), blocking buffer was prepared using the reagents provided with the antibody, according to the manufacturer's instructions.

The blots were then rinsed and solutions containing the streptavidin-HRP (1:5000 in TBS-T pH 7.4) or antibody-HRP (1:2500 in blocking buffer) were added, incubating on a rotator for 1 hr at 25°C. The blots were then washed (5 x 5 min) with TBS-T pH 7.4 and developed using the Amersham ECL Plus kit (GE Healthcare) according to the manufacturer's instructions. They were placed in a

cassette and used to expose Amersham Hyperfilm ECL (GE Healthcare) for 10 to 60 s. Films were developed by a Compact-X4 developer (Xograph Imaging systems).

2.5 Surface plasmon resonance (SPR)

SPR experiments were performed using a Biacore T100 machine using an S series CM5 chip (GE Healthcare). Surface immobilisation and binding experiments were carried out at 25°C. After initial use, chips were stored in HBS-EP buffer, comprising 10 mM 4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid (HEPES)-HCl pH 7.4, 150 mM NaCl, 3 mM EDTA and 0.005% (v/v) Tween-20.

2.5.1 Surface immobilisation

Ligand proteins were immobilised on the sensor chip surface by amine coupling using the immobilisation wizard of the Biacore T100 control software and a flow rate of 5 µl/min. The chip was first primed into HBS-EP buffer. A known quantity of lyophilised protein was dissolved in 10 mM sodium acetate pH 4.2 and diluted to 10 µg/ml in the same buffer. The optimal protein concentration for immobilisation was determined by observing the rate of increase in response when the solutions of different concentrations were injected over an uncoupled flow cell. N-hydroxyl-succinimide (NHS) and 1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide (EDC) were mixed and passed over the target flow cell for 10 min. After washing, the protein solution was injected until the response due to immobilisation reached the target level of ~1000 response units (RU). The surface was then blocked with 1 M ethanolamine for 10 min, and then washed twice with 50 mM HCl and twice with 50 mM NaOH to remove non-covalently bound protein from the surface. The reference flow cell was blocked with ethanolamine, without coupling a protein to the surface.

2.5.2 Binding experiments

The chip was primed into the running buffer, comprising TBS pH 7.4 or TBS-T pH 7.4 (Section 2.4.1). Analyte proteins were dissolved in running buffer and their concentrations were measured (Section 2.2.5). The protein was diluted to an appropriate concentration in running buffer. Several concentrations were prepared by serial dilution and tested using a modified version of the multi-cycle kinetics method from the Biacore T100 control software. The flow rate was 30 µl/min and the contact time was 120 s. The chip was regenerated between cycles using 50 mM HCl. Results were analysed using the Biacore T100 BiaEvaluation software. The baseline response level was established using 'blank' injections of running buffer and the response level of the reference flow cell was subtracted from the response level of flow cells containing immobilised protein. Data were exported from the software and imported into Microsoft Excel for further analysis and graphical representation.

2.6 NMR spectroscopy

NMR experiments were performed using three different spectrometers: home-built spectrometers with ^{1}H operating frequencies of 600 MHz or 750 MHz, with GE/Omega consoles and triple-resonance probes; a Bruker Avance 500 MHz spectrometer with a Cryoplatform, equipped with a TCI CryoProbe. With the exception of the measurement of residual dipolar couplings (RDCs; Section 2.6.9), all experiments were performed with samples in BMS-3 Shigemi tubes (Shigemi Inc.) at 25°C.

2.6.1 Sample preparation

Samples of N-terminal constructs (FUN-EGF3, FUN-EGF1 and EGF2-EGF3-BirA) were prepared in 5% D_2O/95% H_2O (v/v) or 100% D_2O, depending on the experiment, and the pH was raised to 5.4 using HCl or DCl. For three-dimensional (3D) experiments, the typical sample concentration was 1.4 mM. For other experiments, the concentrations are specified in Section 2.6.2. Samples of cbEGF41-43-BirA were prepared in 5% D_2O/95% H_2O (v/v) or 100% D_2O, depending on the experiment, and the pH was raised to 3.8 using HCl or DCl.

2.6.2 Data acquisition and processing

Data acquired using spectrometers with GE/Omega consoles were converted into binary format and processed using the NMRPipe package (Delaglio *et al.*, 1995), while data from the Bruker spectrometer were processed directly using NMRPipe without further conversion.

Parameters for NMR experiments on N-terminal constructs are described in subsequent sections. Where the same experiment was performed on multiple spectrometers, a representative set of parameters is given. Sweep widths were scaled according to the ^1H operating frequencies of the spectrometers. For processing, Gaussian and sine-bell window functions were applied to the direct and indirect dimensions respectively.

2.6.2.1 One-dimensional (1D) ^1H experiments

1D experiments were performed with protein in 5% D_2O/95% H_2O (v/v) or 100% D_2O. Pre-saturation was used to minimise the H_2O signal. The ^1H frequency, sweep width, number of scans and number of complex points were 749.994 MHz, 9090.91 Hz, 512 and 4096 respectively.

2.6.2.2 Two-dimensional (2D) homonuclear ^{1}H experiments

2D ^{1}H experiments aided the assignment of aromatic side-chain resonances. These were the 2D nuclear Overhauser effect spectroscopy (NOESY) (Jeener *et al.*, 1979), 2D total correlation spectroscopy (TOCSY) (Braunschweiler and Ernst, 1983) and 2D double-quantum filtered correlation spectroscopy (DQF-COSY) experiments (Piantini *et al.*, 1982). All experiments used protein in 100% D_2O at pH 5.4. Parameters for these experiments are listed in Table 2.3. The indirect dimension was zero-filled to 2048 complex points.

2.6.2.3 2D heteronuclear experiments

2D heteronuclear experiments were performed using ^{15}N-single and ^{13}C/^{15}N-double labelled proteins. These were the 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) (Bodenhausen and Ruben, 1980; Norwood *et al.*, 1990), 2D ^{1}H-^{15}N heteronuclear multiple quantum coherence-J (HMQC-J) (Kay *et al.*, 1989a), 2D ^{15}N-{^{1}H} heteronuclear NOE (Kay *et al.*, 1989b), 2D ^{1}H-^{15}N in-phase antiphase (IPAP)-HSQC (Ottiger *et al.*, 1998) and 2D ^{1}H-^{13}C constant time (CT)-HSQC experiments (Santoro and King, 1992; Vuister and Bax, 1992). Parameters for these experiments are given in Table 2.4. The indirect dimensions of the resulting spectra were zero-filled to 512 complex points.

Table 2.3 Experimental parameters for 2D homonuclear ^{1}H experiments

Experiment	Number of scans	Direct dimension			Indirect dimension			Mixing time (ms)
		^{1}H freq. (MHz)	^{1}H sweep width (Hz)	^{1}H complex points	^{1}H freq. (MHz)	^{1}H sweep width (Hz)	^{1}H complex points	
NOESY	96	749.9940320	7812.50	1024	749.9940320	7812.50	256	150
TOCSY	64	749.9940320	7812.50	1024	749.9940320	7812.50	350	46
DQF-COSY	64	749.9940320	7812.50	1024	749.9940320	7812.50	400	-

Table 2.4 Experimental parameters for 2D heteronuclear experiments

Experiment	Number of scans	^{1}H freq. (MHz)	^{1}H sweep width (Hz)	^{1}H complex points	^{15}N freq. (MHz)	^{15}N sweep width (Hz)	^{15}N complex points	^{13}C freq. (MHz)	^{13}C sweep width (Hz)	^{13}C complex points
^{15}N HSQC	16	749.9940320	9090.91	1024	76.0048960	2272.72	128	-	-	-
^{15}N HMQC-J	48	749.9940320	9090.91	1024	76.0048960	2272.72	450	-	-	-
^{15}N-{^{1}H} Heteronuclear NOE	96	598.5449880	7272.73	1024	60.6569490	1818.18	128	-	-	-
IPAP-HSQC	96	598.5624470	7407.4	1024	60.6587080	1945.52	128	-	-	-
^{13}C HSQC (aliphatic)	8	749.9940320	9090.91	1024	-	-	-	188.5916960	11627.91	128
^{13}C HSQC (aromatic)	128	598.5449870	7246.38	1024	-	-	-	150.5260820	4545.45	50

2.6.2.4 3D double resonance experiments

3D double resonance experiments were performed on ^{15}N-single and ^{13}C/^{15}N-double labelled protein. Spectra from these experiments were used in sequential backbone resonance assignment, as well as side-chain assignment. These were the 3D ^{1}H-^{15}N and ^{1}H-^{13}C NOESY-HSQC (Marion *et al.*, 1989), 3D ^{1}H-^{15}N TOCSY-HSQC (Marion *et al.*, 1989), 3D ^{1}H-^{15}N HSQC-NOESY-HSQC (Frenkiel *et al.*, 1990; Ikura *et al.*, 1990) and HCCH-TOCSY (Bax *et al.*, 1990). ^{15}N-edited experiments were performed using ^{15}N-labelled protein in 5% D_2O/95% H_2O (v/v), while ^{13}C-edited experiments were performed on ^{13}C/^{15}N-doubled labelled protein in 100% D_2O. The indirect ^{1}H dimensions of the NOESY and TOCSY spectra were zero-filled to 256 points, while the ^{15}N dimension of the ^{15}N-edited NOESY and TOCSY spectra was zero-filled to 128 complex points. The number of complex points in the ^{13}C dimension of the ^{13}C-edited spectra was doubled by forward linear prediction, followed by zero filling to 256 points. For the HSQC-NOESY-HSQC spectrum, both ^{15}N dimensions were zero-filled to 256 points. Experimental parameters are given in Table 2.5.

2.6.2.5 3D triple resonance experiments

3D triple resonance experiments were performed using ^{13}C/^{15}N-doubled labelled protein in 5% D_2O/95% H_2O (v/v), in order to obtain backbone and side-chain assignments. These were the 3D HNCA (Kay *et al.*, 1990), 3D HNCO (Kay *et al.*, 1990), 3D HN(CA)CO (Clubb *et al.*, 1992) and 3D (H)CC(CO)NH experiments (Grzesiek *et al.*, 1993). The number of points in the ^{13}C dimension was doubled using forward linear prediction, and then doubled again by zero filling. The ^{15}N dimension was zero-filled to 256 points. Experimental parameters are given in Table 2.5.

Table 2.5 Experimental parameters for 3D experiments

Experiment	Number of scans	^{1}H freq. (MHz)	^{1}H sweep width (Hz)	^{1}H complex points	^{15}N freq. (MHz)	^{15}N sweep width (Hz)	^{15}N complex points*	^{1}H freq. (MHz)	^{1}H sweep width (Hz)	^{1}H complex points	Mixing time (ms)
^{15}N-edited NOESY-HSQC	8	749.9940320	9090.91	512	76.0048960	2100.84	32	749.9940320	9090.91	128	75 or 150
^{15}N-edited TOCSY-HSQC	8	749.9940320	9090.91	512	76.0048960	2100.84	32	749.9940320	9090.91	128	46
HSQC-NOESY-HSQC	12	749.9940320	9090.91	512	76.0048960	2100.84	32/64	749.9940320	9090.91	-	150
		^{1}H freq. (MHz)	**^{1}H sweep width (Hz)**	**^{1}H complex points**	**^{13}C freq. (MHz)**	**^{13}C sweep width (Hz)**	**^{13}C complex points**	**^{1}H freq. (MHz)**	**^{1}H sweep width (Hz)**	**^{1}H complex points**	**Mixing time (ms)**
^{13}C-edited NOESY-HSQC	8	749.9940320	9090.91	512	188.5916960	11627.91	50	749.9940320	6666.67	128	75 or 150
HCCH-TOCSY	8	749.9940320	9090.91	512	188.5916960	11627.91	50	749.9940320	6493.51	128	15.2
		^{1}H freq. (MHz)	**^{1}H sweep width (Hz)**	**^{1}H complex points**	**^{15}N freq. (MHz)**	**^{15}N sweep width (Hz)**	**^{15}N complex points**	**^{13}C freq. (MHz)**	**^{13}C sweep width (Hz)**	**^{13}C complex points**	**Mixing time (ms)**
HNCA	48	749.9940320	9090.91	512	76.0048960	2272.72	32	188.5951420	4347.83	32	-
HNCO	8	500.0120000	6024.096	512	50.6720000	1724.14	28	125.7270000	1886.08	50	-
HN(CA)CO	24	500.0120000	6024.096	512	50.6720000	1724.14	28	125.7270000	1886.08	50	-
(H)CC(CO)NH	32	749.9940320	9090.91	512	76.0048960	2272.72	32	188.5951420	4347.83	64	17.4

*The HSQC-NOESY-HSQC experiment had two ^{15}N dimensions with different numbers of complex points: 32 in the first and 64 in the second.

2.6.3 Interpretation and analysis of NMR spectra

Peaks in multi-dimensional spectra were picked and assigned using CCPN Analysis software and the associated data model (Vranken *et al.*, 2005). Details of resonance assignment are given in Chapter 4.

2.6.4 Heteronuclear NOE experiment

Spectra with and without the ^{15}N-{^{1}H} NOE were acquired in an interleaved manner and then separated into two data sets. Peak picking, noise estimation and calculation of the NOE ratio were performed using software written by Professor Christina Redfield. Noise was estimated from the intensities of a block of points in a portion of the blank spectrum without any peaks. The estimated noise value was then used to calculate the error on the NOE ratio of all peaks with intensities exceeding a defined threshold. In order to ensure that a steady-state NOE was measured, the ^{1}H pre-saturation period was 4 s. The delay period (τ) between the 90° ^{15}N pulse and the 180° ^{1}H pulse in the INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer) portion of the pulse sequence was set to 2.75 ms or 1.375 ms for the measurement of backbone N-H or side-chain NH_2 NOE ratios respectively (Buck *et al.*, 1995).

2.6.5 Generation of distance restraints

Distance restraints were generated from the peak intensities in 3D NOESY spectra using CCPN Analysis. An arbitrary reference value was defined and an intensity (height) threshold was set for each distance bin. Restraints were separated into unambiguous and ambiguous categories; duplicate restraints were merged. They were exported in CNS ('.tbl') format using the Format Converter.

2.6.6 Measurement of J-couplings from HMQC-J spectra

The ^{15}N dimension of the HMQC-J spectrum (Section 2.6.2.3) was zero-filled to 2048 complex points. Analysis was performed in Felix 2.3 (Accelrys). The Oxford HMQC-J fitting tool was used to fit two curves to a one-dimensional ^{15}N block of points for each peak. The difference between the chemical shift values of the two peaks was used to calculate the J-coupling value. Only peaks with discernible splitting were analysed.

2.6.7 TALOS+ torsion angle prediction

Chemical shift values were exported in TALOS '.tab' format from CCPN Analysis using the Format Converter. The file was modified to include the protein sequence and was then analysed using TALOS+ (Shen *et al.*, 2009). Correction of chemical shift referencing was not used. Torsion angle restraints were generated from the output file using a Python script, including only predictions classed as 'good', multiplying the bounds by two and setting a minimum bound of 30°.

2.6.8 Identification of slowly-exchanging amide protons

The pH of ^{15}N-labelled protein at a concentration of 0.5 mM in 100% H_2O was adjusted to 3.8 and the sample was then lyophilised. Freeze-dried protein was then dissolved in 100% D_2O, the pH was checked and adjusted if necessary, and 10 HSQC spectra were queued for acquisition. Peak assignments were facilitated by HSQC spectra of protein at pH 3.8, 4.5, 5.0 and 5.4 and resonance assignments of protein at pH 5.4.

2.6.9 Measurement of residual dipolar couplings (RDCs)

2.6.9.1 Bicelle preparation

Bicelles were prepared according to an established method (Ottiger and Bax, 1999). A 10% (w/v) stock of bicelles was prepared using the ether-linked lipids 1,2-O-ditridecyl-*sn*-glycero-3-phosphocholine (D13OPC) and 1,2-dihexyl-*sn*-glycero-3-phospho-choline (DHOPC; Avanti Polar Lipids) in a ratio of 3.5:1 (mol:mol). The two lipids, in powder form, were weighed separately, dissolved in milliQ-purified water and then mixed. To produce a homogeneous mixture, the solution was vortexed and subjected to several cycles of cooling on ice and warming to room temperature. The mixture was then incubated at 4°C for two days. For RDC measurements, the mixture was diluted to 5% (w/v) and D_2O was added to a final concentration of 10% (v/v). In addition, 100 mM cetyl trimethyl ammonium bromide (CTAB; Sigma-Aldrich) was added to the mixture such that the D13OPC:DHOPC:CTAB molar ratio was 35:10:1 and pH was adjusted to 5.4. Protein was added to the solution at a final concentration of 0.5 mM and the pH was checked before acquisition of NMR data.

2.6.9.2 Monitoring bicelle alignment by 1D ^{2}H spectroscopy

The strength of alignment was monitored by observing the residual quadrupolar splitting of the ^2H_2O signal. A 1D spectrum was acquired at 35°C using a spectrometer with a ^{1}H operating frequency of 600 MHz, without a field frequency lock. The frequency was 91.8829620 Hz, the sweep width was 250 Hz, the number of complex points was 4096 and the pulse length was 1 ms. The magnitude of the splitting was checked before and after IPAP-HSQC experiments, in order to confirm that the alignment was stable.

2.6.9.3 Data acquisition and analysis

HSQC spectra at 25, 30 and 35°C were acquired to obtain peak assignments at 35°C, above the bicelle phase transition temperature. IPAP-HSQC spectra were first acquired using protein in the absence of alignment medium at 35°C, in order to obtain the isotropic peak splitting values. The same experiments were performed again in the presence of bicelles. RDC values were calculated by subtracting the isotopic peak splitting value from the anisotropic peak splitting. Peak picking and calculation of RDC values were performed using software written by Professor Christina Redfield.

2.6.10 Titrations with $CaCl_2$ and $MgCl_2$

A 0.3 mM protein sample was prepared in 5% D_2O/95% H_2O (v/v) at pH 3.8. A ^{1}H-^{15}N HSQC spectrum (Section 2.6.2.3) was acquired before adding $CaCl_2$ or $MgCl_2$. Spectra were then acquired at different concentrations of $CaCl_2$ or $MgCl_2$. $CaCl_2$ was added to the protein solution sequentially to give final concentrations of 0, 1, 5, 15, 25, 50 and 100 mM. The $MgCl_2$ concentrations tested were 0, 10 and 50 mM.

2.6.11 Protein titrations

^{15}N-labelled protein (FUN-EGF3, FUN-EGF1 or EGF2-EGF3-BirA) was titrated with unlabelled cbEGF41-43-BirA. A sample of ^{15}N-labelled protein was prepared at a concentration of 0.3 mM and Tris-HCl pH 7.5 was added to a final concentration of 5 mM. Unlabelled, HPLC-purified cbEGF41-43-BirA was dissolved in H_2O and the pH was adjusted to pH 7.5 using NaOH, before being aliquoted and lyophilised. Freeze-dried aliquots were then dissolved in the solution containing ^{15}N-labelled protein. The pH was checked before and after the addition of cbEGF41-43-

BirA, and a ^{1}H-^{15}N HSQC spectrum was acquired. A series of spectra were acquired at increasing concentrations of cbEGF41-43-BirA. More scans were acquired at higher concentrations of cbEGF41-43-BirA (256 at 0.25 mM), as the intensity of peaks in the HSQC spectrum decreased with every addition of cbEGF41-43-BirA.

2.7 NMR structure calculation

Structure calculations were performed using Xplor 3.8 and Xplor-NIH 2.29 (Schwieters *et al.*, 2003). The older Xplor version was used for prochiral swapping calculations as Xplor-NIH does not have the appropriate functionality.

2.7.1 Simulated annealing from an extended template

Initial structures were generated by simulated annealing from an extended template using Xplor-NIH. In the absence of prochiral swapping, distance restraints were modified to remove arbitrary stereospecific assignments. Restraints involving pairs of prochirals were systematically merged using a Python script. For example, HB1-HB2 prochiral pairs were replaced with HB#, taking the lower of the two upper distance bounds as the bound for the merged restraint.

Simulated annealing was performed using the internal variable module (IVM) of Xplor-NIH (Schwieters and Clore, 2001). A tutorial script for the Python interface was modified for calculations, including only distances and torsion angles as the experimentally-derived restraints. A 'soft' potential was used for distance restraints. Disulphide bonds were incorporated as distance restraints − assuming that all cysteines form disulphide bonds, pairs of S^{γ} atoms must be within 2.02 ± 0.1 Å. The annealing procedure contained an initial high temperature dynamics stage (3500 K) with 1000 steps. For this stage, the force constants for the distance and torsion

angle restraints were set to 20 kcal mol^{-1} $Å^{-2}$ and 10 kcal mol^{-1} rad^{-2} respectively. The system was then cooled from 3500 K to 100 K in steps of 25 K using torsion angle dynamics. The time step was 10 fs and the total time at each temperature was 2 ps. Force constants for the distance restraint, torsion angle restraint, bond angle and van der Waals terms were increased in a linear fashion during the calculation, while the torsion angle force constant was kept constant (Table 2.6). The scale factor for the van der Waals radius was ramped from 0.4 to 0.8. Energy minimisation was then performed in torsion angle space, followed by energy minimisation in Cartesian space using all atoms. 200 structures were calculated from a starting extended structure, using a different random 'seed' value for each calculation.

Table 2.6 Force constants for terms in simulated annealing calculations (Xplor-NIH)

Term	Start	End
Distance restraints	1 kcal mol^{-1} $Å^{-2}$	30 kcal mol^{-1} $Å^{-2}$
Torsion angle restraints	200 kcal mol^{-1} rad^{-2}	200 kcal mol^{-1} rad^{-2}
Bond angles	0.4 kcal mol^{-1} rad^{-2}	1 kcal mol^{-1} rad^{-2}
Impropers	0.4 kcal mol^{-1} rad^{-2}	1 kcal mol^{-1} rad^{-2}
Van der Waals repulsive term	0.004 kcal mol^{-1} $Å^{-4}$	4 kcal mol^{-1} $Å^{-4}$

After elimination of significant distance restraint violations (>0.3 Å), stereospecific assignments were obtained using prochiral swapping. Initial calculations before introducing prochiral swapping were performed using Xplor-NIH, as they were ~4 times faster. An Xplor script, kindly provided by Louise Walport and Christina Redfield, was modified for use in structure calculations using distance and torsion angle restraints. The cooling protocol involved an initial energy minimisation step, a 30000 step high-temperature stage (2000 K), two cooling stages and a final energy minimisation step. The first cooling stage (40000 steps) lowered the temperature from 2000 K to 1000 K, and the second stage (20000 steps) lowered the temperature to 100 K. The temperature step was 50 K. The force constants for the distance and torsion angle restraints were 1 kcal mol^{-1} $Å^{-2}$ and 5 kcal mol^{-1} rad^{-2} respectively. A torsion angle database potential (Rama) (Kuszewski *et al.*, 1996) was also included, with a force constant of 0.002 kcal mol^{-1} rad^{-2}. Prochiral swaps were

analysed in an ensemble of 100 structures and distance restraints were modified to incorporate stereospecific assignments (Section 2.7.2). These restraints were then used to generate structures using the Xplor-NIH simulated annealing protocol.

2.7.2 Analysis of prochiral swapping calculations

The total energy of every structure was calculated by Xplor. Structures were sorted according to total energy using a Python script, with the lowest energy structure giving the best agreement to the input restraints. The '.prt' files of the 50 lowest-energy structures were analysed to calculate the frequency of prochiral swaps. Prochiral pairs swapped in at least 40 structures were considered to be stereospecifically assigned. To test whether a lack of swapping of prochiral pairs represented a true stereospecific assignment, prochiral pairs exchanged in 10 or fewer structures were swapped in the distance restraints using a Python script and the calculations were repeated. Prochiral pairs swapped in at least 40 structures calculated using these modified restraints were then swapped in the original distance restraints. All others (swapped in 40 structures or fewer) were made ambiguous. Restraints involving ambiguous groups were given equal upper distance bounds if both members of the prochiral pair were present, or replaced with the wildcard (#) atom descriptor.

2.7.3 Refinement of initial structures

Refinement was carried out in two stages. First, structures from simulated annealing were refined with the Rama database potential and the hydrogen bond database potential (HBDB) (Grishaev and Bax, 2004). Second, partially refined structures were refined further using RDC restraints. A refinement protocol for Xplor-NIH was modified from an example script using the Python interface. The protocol involved a cooling stage in torsion angle space, from 3000 K to 25 K with a

temperature step of 12.5 K, followed by energy minimisation, as in the simulated annealing protocol (Section 2.7.1). There were 100 1 fs steps per temperature. The force constant for the distance restraints was kept constant, while others, including the RDC restraints, were ramped (Table 2.7). A 'hard' potential was used for the distance restraints. The RDC force constant was chosen to reflect the measurement error. The scale factor for the van der Waals radius term was ramped from 0.9 to 0.8.

Table 2.7 Force constants for terms in refinement calculations (Xplor-NIH)

Term	Start	**End**
Distance restraints	50 kcal mol^{-1} $Å^{-2}$	50 kcal mol^{-1} $Å^{-2}$
Torsion angle restraints	200 kcal mol^{-1} rad^{-2}	200 kcal mol^{-1} rad^{-2}
RDC restraints (Sani potential)	0.01 kcal mol^{-1} rad^{-2}	0.5 kcal mol^{-1} rad^{-2}
Rama potential	0.002 kcal mol^{-1} rad^{-2}	1 kcal mol^{-1} rad^{-2}
Bond angles	0.4 kcal mol^{-1} rad^{-2}	1 kcal mol^{-1} rad^{-2}
Impropers	0.1 kcal mol^{-1} rad^{-2}	1 kcal mol^{-1} rad^{-2}
Van der Waals repulsive term	0.004 kcal mol^{-1} $Å^{-4}$	4 kcal mol^{-1} $Å^{-4}$

The 10 lowest-energy structures from the structure ensemble calculated by simulated annealing were used as starting structures for refinement calculations. 20 structures were generated for each starting structure, producing a new ensemble of 200 structures. Hydrogen bonds satisfying the HBDB restraints in more than 16 of the 20 lowest-energy structures were fixed in subsequent calculations. Parameters for the alignment tensor were estimated using partially refined structures (Section 2.7.4). Detail about the derivation of alignment tensor parameters is given in Section 4.8.6. A final ensemble of structures was then generated, using the 10 lowest-energy structures from the first refinement stage as starting structures for refinement calculations including RDC restraints.

2.7.4 Fitting of RDCs to calculated structures

Experimental RDC values were compared to predicted RDC values based on calculated structures. A program written by Professor Christina Redfield was used to calculate the axial and rhombic components of the alignment tensor, as well as the three angles describing its orientation, for a given set of RDCs and an input structure. The agreement between the predicted and experimental RDCs was quantified by the Q factor (Cornilescu *et al.*, 1998):

$$Q = \left(\frac{\sum(D_{meas}-D_{calc})^2}{\sum(D_{meas})^2}\right)^{1/2} \qquad 2.5$$

where D_{meas} and D_{calc} are the experimentally measured and calculated RDCs respectively. A Q value of 0 indicates perfect agreement. The best partially refined structures had Q < 0.3.

2.7.5 Average structure and ensemble RMSD calculations

Energy-minimised, average structures were calculated from the 20 lowest-energy structures of the final structure ensemble using an Xplor-NIH script. The average root mean square deviation (RMSD) from the average structure was calculated for 20 structures in the ensemble using a Python script. RMSD is defined:

$$RMSD = \left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\right)^{1/2} \qquad 2.6$$

2.7.6 Buried surface area and domain twist/tilt calculations

Buried surface calculations were performed using the Areaimol program in the CCP4 software package (Collaborative Computational Project, 1994). Calculations were performed on the two halves of the structure forming the interface and the complete structure. Accessible surface areas were calculated using the software. The buried surface area for each half was determined by subtracting the accessible surface area in the half structure from the accessible surface area in the complete structure. The values for the two halves were then averaged. Residues involved in interdomain packing were identified by measuring inter-atomic distances in the average structure using a Python script. Packing residues were defined as those with side-chain atoms within 4 Å of atoms of residues in the opposing domain.

Twist and tilt angle calculations were performed using Mod2 software (Downing *et al.*, 1996). The β-hairpin structures in the EGF2 and EGF3 domains of human fibrillin-1 were used to define the planes of the two domains. For EGF2, the residue ranges used were G127-S130 and H133-C136; for EGF3, the residues were G158-V161 and R125-C128. A script provided by Professor Christina Redfield was used to calculate angles for all 20 structures in the final ensemble as a batch.

2.8 Fibroblast culture and transfection

2.8.1 Maintenance and culture of cells

MSU1.1 cells are transformed human dermal fibroblasts that produce extracellular fibrillin microfibrils (Cardy and Handford, 1998; Morgan *et al.*, 1991). They were cultured in Dulbecco's Modified Eagle Medium (DMEM; Invitrogen), supplemented with 10% (v/v) fetal bovine serum

(FBS; Invitrogen) and 50 U/ml penicillin, 50 µg/ml streptomycin and 2 mM glutamine (Sigma-Aldrich). This medium is hereafter referred to as complete DMEM. Cells were cultured in 75 cm^2 tissue culture flasks (T75; Greiner) at 37°C in humidified atmosphere with 5% (v/v) CO_2. They were passaged every three to four days or when 90-100% confluent (Section 2.8.2). Tissue culture work was carried out in a Scanlaf Mars Class 2 biological safety cabinet.

2.8.2 Sub-culture of cells

Cell monolayers were washed with 5 ml sterile, autoclaved phosphate-buffered saline (PBS) with 0.02% (w/v) EDTA (PBSE). They were treated with a PBSE solution containing 0.07% (w/v) trypsin (Invitrogen) and incubated (1 min; 37°C) to detach cells. Proteolysis was stopped by the addition of complete DMEM. Cells were pelleted by centrifugation (3 min; 1000 rpm; ALC centrifuge PK120, T534 rotor) and then resuspended in fresh complete DMEM. 1/10 to 1/3 of the original cells were re-seeded in new flasks.

2.8.3 Freezing, thawing and storage of cells

For freezing, cells were detached from flasks containing a ~90% monolayer (Section 2.8.2). Following centrifugation, cell pellets were resuspended in an appropriate volume of 90% (v/v) FBS with dimethyl sulphoxide (DMSO). Cells from one T75 flask were dissolved in 3 ml FBS-DMSO and split into three 1 ml aliquots, transferred to cryovials (Nunc) and frozen slowly at −80°C in a StrataCooler cryo preservation module (Stratagene). Frozen cells were stored at −80°C for up to approximately one month before being re-seeded in culture, or were transferred to liquid nitrogen for long-term storage. For re-seeding, vials were thawed rapidly by incubation in warm, sterile water at 37°C. The cell suspension was diluted in complete DMEM, pelleted by centrifugation, resuspended in medium and transferred to flasks (Section 2.8.2).

2.8.4 Stable transfection

MSU1.1 cells were transfected with circular plasmid DNA purified by phenol-chloroform extraction (Section 2.1.12). 24 µg DNA was used per transfection. DNA was added to 1.5 ml Opti-MEM (Minimal Essential Medium; Invitrogen) and 60 µl Lipofectamine-200 (Invitrogen) was added to another 1.5 ml Opti-MEM. They were incubated for 5 min at room temperature, mixed gently and incubated for a further 20 min to allow DNA-liposome complexes to form. During the second incubation period, a ~90% confluent monolayer of MSU1.1 cells in a T75 flask was washed twice with 5 ml Opti-MEM medium. The cells were then incubated with 12 ml Opti-MEM (10 min; 37°C). The Opti-MEM-DNA-lipofectamine mix was added to the flask and mixed gently into the medium. Cells were cultured in this medium for ~20 hr (Section 2.8.1). The medium was then removed and replaced with 15 ml complete DMEM. After two to three days, the cells were split 1:2 into new T75 flasks (Section 2.8.2) and re-seeded into medium containing complete DMEM containing 2.5 µg/ml puromycin (Sigma). Puromycin-containing medium was replaced every three to four days until at least 50 clones of ~3 mm diameter were visible.

2.8.5 Protein secretion assay

There were two T75 flasks from each transfection. Cells from one flask were frozen and stored (Section 2.8.3), while the second was split 1:4 into four 25 cm^2 (T25) tissue culture flasks. Cell medium and lysate samples were collected every follow passaging for three days, while cells from the fourth flask were frozen and stored. Cell medium samples were filtered with a 0.22 µm syringe filter (Millipore) to remove cell debris. 150 µl medium was then added to 4 x SDS-PAGE loading buffer and the sample was boiled (Section 2.2.10). Cell lysate samples were obtained by detaching cells (Section 2.8.2), followed by a wash with 10 ml PSBE to remove excess medium.

The cell pellet was resuspended in 0.5 ml 2 x SDS-PAGE loading buffer and boiled (Section 2.2.10). Cell medium and lysate samples were stored at −20°C.

Samples were analysed by Western blotting (Section 2.4.2). A 10% separating gel was prepared (Section 2.2.10). 40 µl and 10 µl of each medium and lysate sample were loaded respectively. Following SDS-PAGE and electrophoretic transfer, the nitrocellulose membrane was blocked for at least 1 hr at room temperature in a solution comprising PBS, 0.1% (v/v) Triton-X-100 (PBS-T) and 5% (w/v) milk. The membrane was then incubated with primary antibody diluted in blocking buffer, on a rotator at 25°C for 1 hr. The anti-Pro-rich antibody (Kettle *et al.*, 1999) was diluted 1:500, while the anti-actin antibody (A2547; Sigma-Aldrich) was diluted 1:1000. The membrane was then washed with PBS-T (5 x 5 min) and the goat anti-rabbit secondary antibody (Sigma-Aldrich) was added, diluted 1:1000 in fresh blocking buffer. Following washes with PBS-T (5 x 5 min) and a rinse with milliQ-purified water, the membrane was developed (Section 2.4.2).

Chapter 3

Production, characterisation and interaction of human fibrillin-1 N- and C-terminal fragments

3.1 Introduction

An interaction between the N- and C-terminal regions of fibrillin is thought to be a key step in the assembly of the protein into microfibrils in the extracellular matrix (Section 1.7.2). The minimal regions of human fibrillin-1 required for binding have been identified: domains FUN-EGF3 at the N-terminus and cbEGF41-cbEGF43 at the C-terminus (El-Hallous *et al.*, 2007; Hubmacher *et al.*, 2008) (Section 1.7.2). However, little is known about the molecular and structural basis of the interaction. In particular, there is no high-resolution structural information about the domain types found at the N-terminal interaction site (Section 1.3.5). Moreover, no specific residues in the N- or C-terminal regions have yet been implicated in the interaction.

The first steps in dissecting the N-C interaction are production and characterisation of fibrillin fragments corresponding to the minimal interacting regions. Large (milligram) quantities of homogeneous material are required for structural work. Prokaryotic expression, purification and *in vitro* refolding have previously been successfully employed to make other fragments of fibrillin, and other disulphide-rich glycoproteins such as Notch and Jagged for structural and functional analysis (Cordle *et al.*, 2008a; Downing *et al.*, 1996; Lee *et al.*, 2004). The functional properties of these fragments, including binding specificity, replicate the behaviour of proteins produced by mammalian expression systems, despite the absence of post-translation modifications. This is therefore the system of choice for producing N- and C-terminal fibrillin fragments.

The use of an *in vitro* refolding system means that the fold of recombinant material must be assessed to confirm that its behaviour is comparable to what would be expected for the equivalent region in native fibrillin molecules. For fragments containing cbEGFs, fold can be validated by confirming calcium binding by individual domains. This is because calcium binding by cbEGF domains is dependent on the correct formation of disulphide bonds in a C1-C3 C2-C4 C5-C6

arrangement (Handford *et al.*, 1990). Comparisons of limited proteolytic digests carried out under different conditions can be used to show calcium-dependent structure (Suk *et al.*, 2004; Whiteman *et al.*, 2007). NMR spectroscopy can provide domain-specific and quantitative information about calcium binding (Jensen *et al.*, 2005; Jensen *et al.*, 2009; Knott *et al.*, 1996). For characterisation of calcium-independent regions, NMR can be used, but more in-depth structural analysis is required.

Before this thesis project, N-terminal (FUN-EGF3) and C-terminal (cbEGF41-cbEGF43; hereafter referred to as cbEGF41-43) fragments were produced and shown to interact using a pull-down assay (S. Jensen, D. Stoddart, P. Handford, unpublished data). The interaction was not dependent on the presence of the unique C-terminal region (between the end of the cbEGF43 domain and up to the furin/PACE recognition site), consistent with published results (Hubmacher *et al.*, 2008). A C-terminal recognition sequence for the BirA enzyme was included in the FUN-EGF3 protein construct to allow site-specific biotinylation. BirA catalyses the covalent attachment of biotin to a lysine residue in the recognition sequence (O'Callaghan *et al.*, 1999). Biotinylation permits immobilisation of the BirA-tagged fragment using the high-affinity streptavidin-biotin interaction. In the pull-down assay, FUN-EGF3-BirA was immobilised on streptavidin-coated magnetic beads and used to pull down the cbEGF41-43 fragment in solution (Figure 3.1). However, the very low solubility of cbEGF41-43 at physiological pH meant that the assay buffer had to be at pH 9.5.

In this chapter, the pull-down result was first verified. Additional fragments were then produced so that the pull-down could be carried out in the reverse arrangement: FUN-EGF3 and cbEGF41-43 with a C-terminal BirA tag (cbEGF41-43-BirA). Immobilisation of cbEGF41-43-BirA allowed the assay to be carried out under more native conditions and confirmed that the fragments interacted. Appropriate negative controls are essential to eliminate the possibility of non-specific binding. Therefore, fragments with similar domain structures to FUN-EGF3 and cbEGF41-43 were

included in the assay. SPR was also used to verify the pull-down result. The FUN-EGF3 and cbEGF41-43-BirA fragments were also characterised. Unfortunately, the low solubility of cbEGF41-43-BirA, particularly in the presence of $CaCl_2$, meant that its characterisation was challenging. Nonetheless, limited proteolysis and NMR were employed to study calcium binding. 1D and 2D NMR experiments were used to examine the fold of the FUN-EGF3 fragment and to provide initial insights into its structure.

3.2 Interaction of an immobilised FUN-EGF3 fragment with cbEGF41-43 in solution

The pull-down result confirming the interaction of the N- and C-terminal fragments was first reproduced. Biotinylated FUN-EGF3 with a C-terminal BirA tag (FUN-EGF3-BirA-bio), cbEGF41-43 with an N-terminal hexa-histidine (His_6) tag (His-cbEGF41-43) and negative control fragment His-cbEGF32-34 were kindly provided by Dr Sacha Jensen (Figure 3.1A). cbEGF32-34 is a well-characterised triple cbEGF domain fragment of human fibrillin-1 (Aileen McGettrick, D.Phil. thesis, 2002). In the assay, FUN-EGF3-BirA-bio was immobilised on magnetic streptavidin-coated M270 dynabeads (Section 2.4.1). The beads were then used to pull down His-cbEGF41-43 from solution (Figure 3.1B). After extensive washing, beads were boiled in reducing SDS-PAGE loading buffer (Section 2.2.10). Samples were analysed by gel electrophoresis and subsequent Western blotting (Section 2.4.2; Figure 3.1C). Western blots demonstrated that cbEGF41-43 interacted specifically with FUN-EGF3-BirA. His-cbEGF41-43 did not bind to the beads alone and the negative control cbEGF32-34 fragment did not interact with FUN-EGF3-BirA-bio. Higher molecular weight bands were observed in both blots, running at roughly double the size of the monomeric proteins. The origin of these bands was not clear.

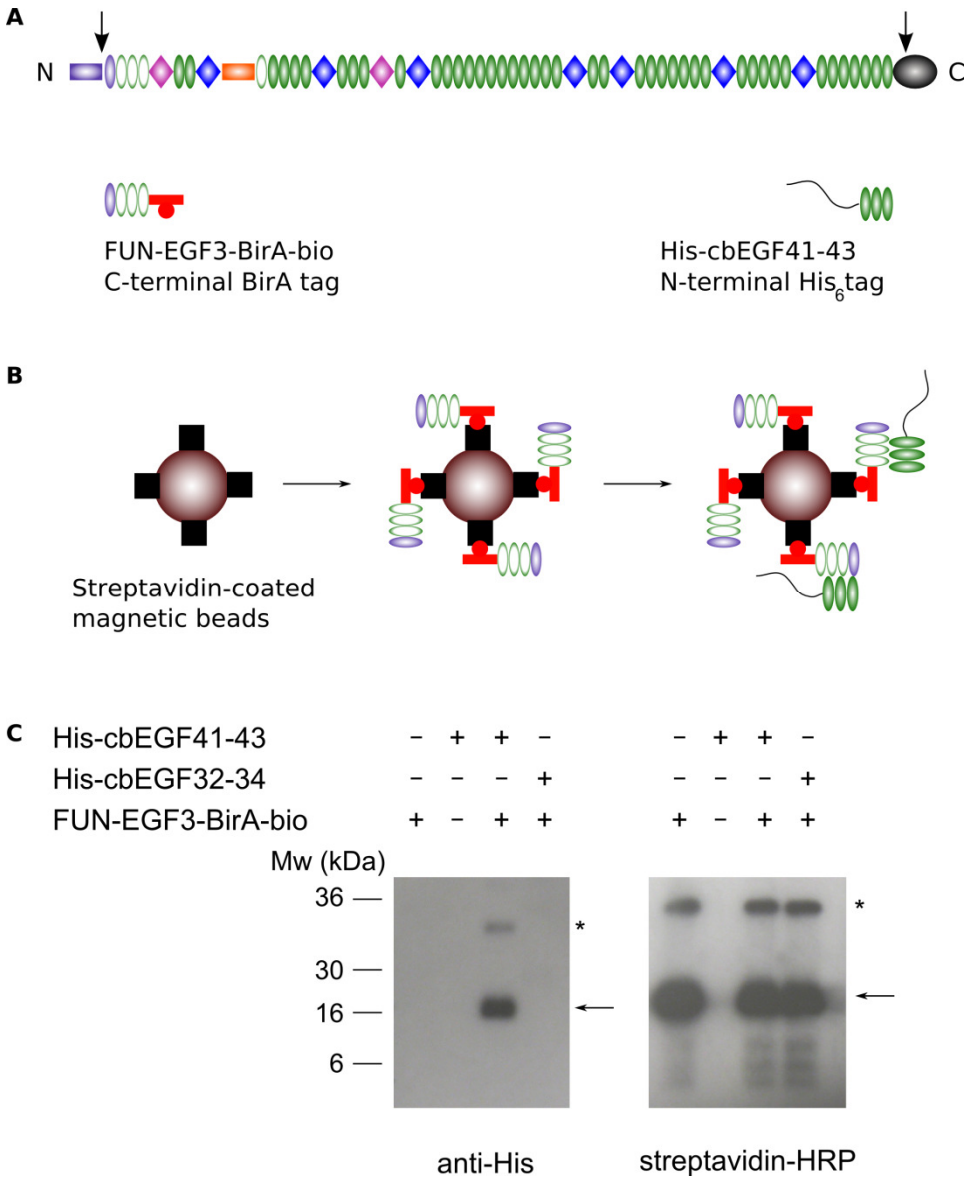


Figure 3.1 Pull-down assay using an immobilised, biotinylated FUN-EGF3-BirA fragment (FUN-EGF3-BirA-bio). A) Illustration of the N- and C-terminal fragments encompassing the minimal regions required for binding: FUN-EGF3 and cbEGF41-43. For the purposes of immobilisation and detection, FUN-EGF3 had a C-terminal BirA tag and cbEGF41-43 had an N-terminal His_6 tag. B) FUN-EGF3-BirA-bio was immobilised on beads coated with multiple streptavidin molecules and was used to pull down cbEGF41-43. C) Western blots of supernatants from bead samples after treatment with reducing SDS-PAGE sample buffer. Samples were run on 16% SDS-PAGE gels; protein was then transferred electrophoretically from gels to nitrocellulose membranes. Biotinylated and His-tagged proteins were detected using streptavidin-HRP and an anti-His-HRP antibody respectively (bands marked with arrows). The positions of molecular weight (Mw) markers are indicated on the left-hand side. There were bands in both blots with lower than expected mobility (marked with *).

3.3 Development of a pull-down assay using immobilised cbEGF41-43

Although the experiment using immobilised FUN-EGF3 showed a specific interaction, the non-physiological assay conditions raise questions about its validity. The assay buffer pH of 9.5 was necessary to prevent precipitation of the cbEGF41-43 fragment, but is unrepresentative of conditions under which the N-C interaction occurs *in vivo*. Therefore, an alternative scheme was devised with immobilised cbEGF41-43 and FUN-EGF3 in solution (Figure 3.2), with the aim of preventing cbEGF41-43 precipitation at a more physiologically relevant pH. This required the production of three additional proteins: FUN-EGF3 with an N-terminal His_6 tag (His-FUN-EGF3), as well as cbEGF41-43 and cbEGF32-34 with C-terminal BirA tags (cbEGF41-43-BirA and cbEGF32-34-BirA respectively).

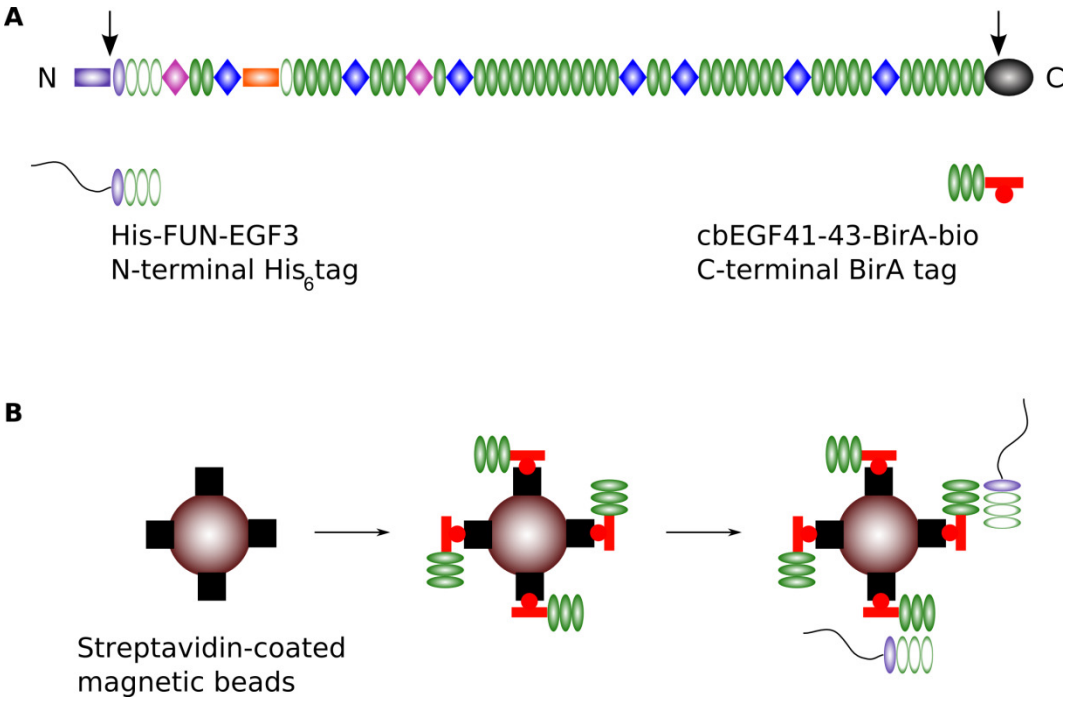


Figure 3.2 Pull-down assay using an immobilised, biotinylated cbEGF41-43-BirA fragment (cbEGF41-43-BirA-bio). A) Fragments required for this assay: His-FUN-EGF3 and cbEGF41-43-BirA-bio. B) As in Figure 3.1, the bead-immobilised fragment pulls down the interaction partner from solution.

3.4 Expression, purification and refolding of fibrillin fragments

A pQE-30 plasmid (Qiagen; Figure 3.3) encoding the FUN-EGF3 fragment without a C-terminal BirA tag was already available and was kindly provided by Dr Sacha Jensen.

3.4.1 Cloning of constructs encoding cbEGF41-43-BirA and cbEGF32-34-BirA

DNA sequences encoding cbEGF41-43 and cbEGF32-34 regions were amplified from human fibrillin-1 cDNA by PCR (Section 2.1.1; see Appendix B for primer sequences). *Bam*HI and *Sal*I restriction sites were introduced by the forward and reverse primers respectively. This allowed cloning of the sequences into a version of the pQE-30 vector containing a sequence encoding the BirA tag (Figure 3.3). There is also a sequence in the vector before the insert that encodes an N-terminal His_6 tag. The forward primer sequence was designed to introduce a factor Xa recognition site (I-E-G-R) in the protein between the His_6 tag and the first residue of the fibrillin sequence, with a Ser-Ala spacer in between. The protease factor Xa cleaves the peptide bond after the arginine residue in the recognition site (Knott *et al.*, 1996). The cloned region of the plasmid was sequenced to verify the presence of the correct insert (Section 2.1.11).

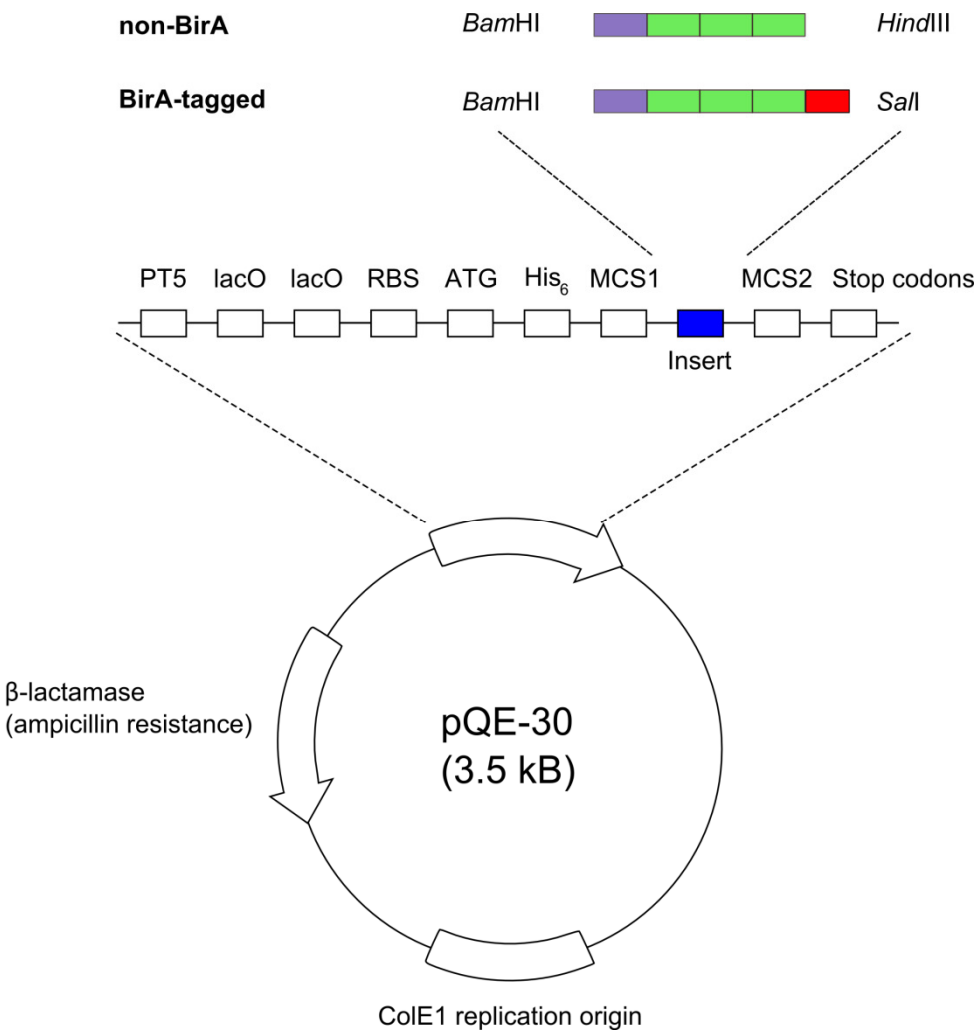


Figure 3.3 pQE-30 plasmid for protein expression in *E. coli.* The pQE-30 vector contains a ColE1 replication origin and a gene encoding β-lactamase, which degrades ampicillin and therefore confers resistance to the antibiotic on bacteria transformed with the plasmid. For expression of non-BirA-tagged proteins, sequences were amplified with *Bam*HI and *Hind*III restriction enzyme sites in the forward and reverse primers respectively. For expression of BirA-tagged proteins, sequences were cloned into a modified version of pQE-30, also encoding the 15-residue BirA recognition peptide, using *Bam*HI and *Sal*I restriction sites. PT5 = T5 promoter; lacO = lac operator (LacI binding site); RBS = ribosome binding site; ATG = start codon; His_6 = hexa-histidine tag-encoding sequence; MCS1/2 = multiple cloning site. Adapted from the QiaExpressionist manual (Qiagen). The lacO sequence means that the protein expression is inducible in the presence of the pREP4 plasmid, which encodes the LacI repressor protein.

3.4.2 Protein expression, purification and refolding

Protein fragments were expressed and purified with N-terminal His_6 tags, with a factor Xa recognition site inserted to allow His_6 tag removal (Figure 3.4A; Section 3.4.1). Protein was expressed in BL21 cells also containing the pREP4 plasmid, which contains a gene encoding the LacI protein. Protein expression was induced by IPTG, which alleviates LacI repression of the *lac* operon. The protocol for protein production is summarised in Figure 3.4B; more details are given in Section 2.2.

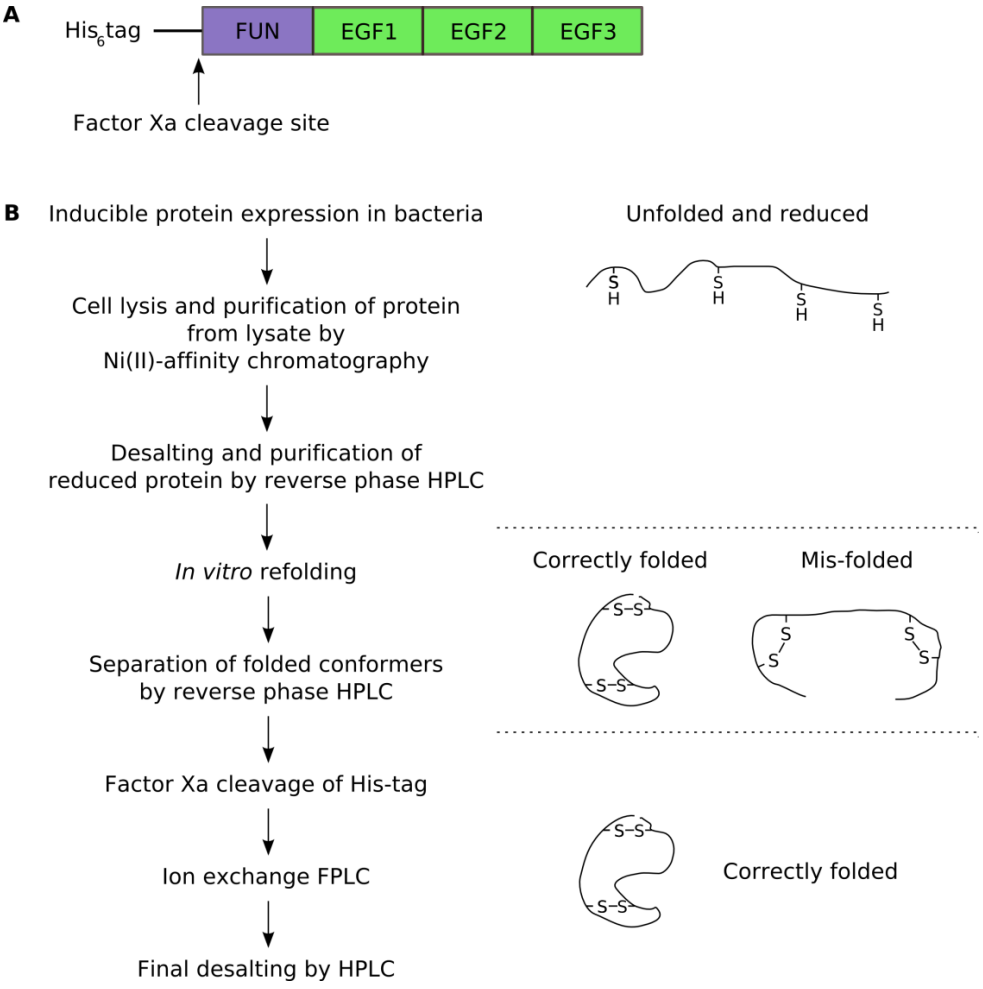


Figure 3.4 Protein expression, purification and refolding. A) Example of a protein construct: FUN-EGF3 without a BirA tag. B) Steps in purification. Proteins were purified in their reduced, denatured form, then desalted and refolded *in vitro* using an oxido-shuffling system. Folded conformers were then separated using reverse phase HPLC and anion or cation exchange FPLC, followed by a second HPLC purification step.

3.4.2.1 Purification and refolding of FUN-EGF3

Purification and refolding conditions for the FUN-EGF3 fragment were established prior to this project, although modifications were made to improve the yield (Section 2.2). Representative HPLC and FPLC chromatograms are shown in Figure 3.5. Reduced material eluted as a single peak on a gradient of acetonitrile (Figure 3.5A). Freeze-dried, desalted protein was then dissolved in deionised water and refolded using an oxido-shuffling system. In addition to the standard conditions used for refolding (0.2 mg/ml protein, 100 mM Tris pH 8.3, 3 mM cysteine, 0.3 mM cystine), glycerol was added to a final concentration of 50% (v/v) (Section 2.2.6).

The predicted isoelectric point of the unfolded form of FUN-EGF3 is ~8.4 (Section 2.2.5), suggesting that its solubility may be at a minimum in the refolding mixture. A buffer pH of 8.3 is typically used because it matches the approximate acid dissociation constant (pK_a) of thiol groups in cysteine side-chains and therefore should facilitate optimal disulphide bond formation through a shuffling mechanism. Glycerol, like some other small molecules, can significantly improve the solubility of some proteins during refolding (De Bernardez Clark *et al.*, 1999). It has previously been used in refolding of some fibrillin multi-domain fragments (Jensen *et al.*, 2009) and was essential for efficient refolding of FUN-EGF3. Following incubation at 4°C, the refolding mixture was acidified to pH ~2 and dialysed against 0.1% (v/v) TFA to reduce the concentration of glycerol (Section 2.2.6). This ensured that the solution viscosity was sufficiently low for protein concentration by ultrafiltration and HPLC purification (Section 2.2.4).

The HPLC elution profile of refolded FUN-EGF3 showed a sharp, major peak with right and left shoulders (Figure 3.5B). The major peak eluted earlier than the reduced material (Figure 3.5A), as expected. Analysis by SDS-PAGE suggested that these shoulders most likely corresponded to disulphide-bonded aggregates or misfolded material (data not shown). The major species was

collected and freeze-dried (Section 2.2.4). At this stage, some of the material was treated with factor Xa to remove the His_6 tag (Section 2.2.7). His-FUN-EGF3 is needed for the pull-down assay (Section 3.3), but FUN-EGF3 is better suited for NMR analysis, as the His_6 tag is unstructured and is likely to give rise to intense, sharp peaks in NMR spectra. Factor Xa-treated FUN-EGF3 and His-FUN-EGF3 were both purified further by cation exchange FPLC (Section 2.2.8) and a final HPLC step (Figure 3.5C-F). His-FUN-EGF3 (Figure 3.5C) eluted at a higher NaCl concentration than FUN-EGF3 (Figure 3.5E) on the FPLC − mostly likely resulting from the additional positive charges contributed by histidine residues in the His_6 tag. The major species from these purification steps were collected and the final products were freeze-dried (Section 2.2.4).

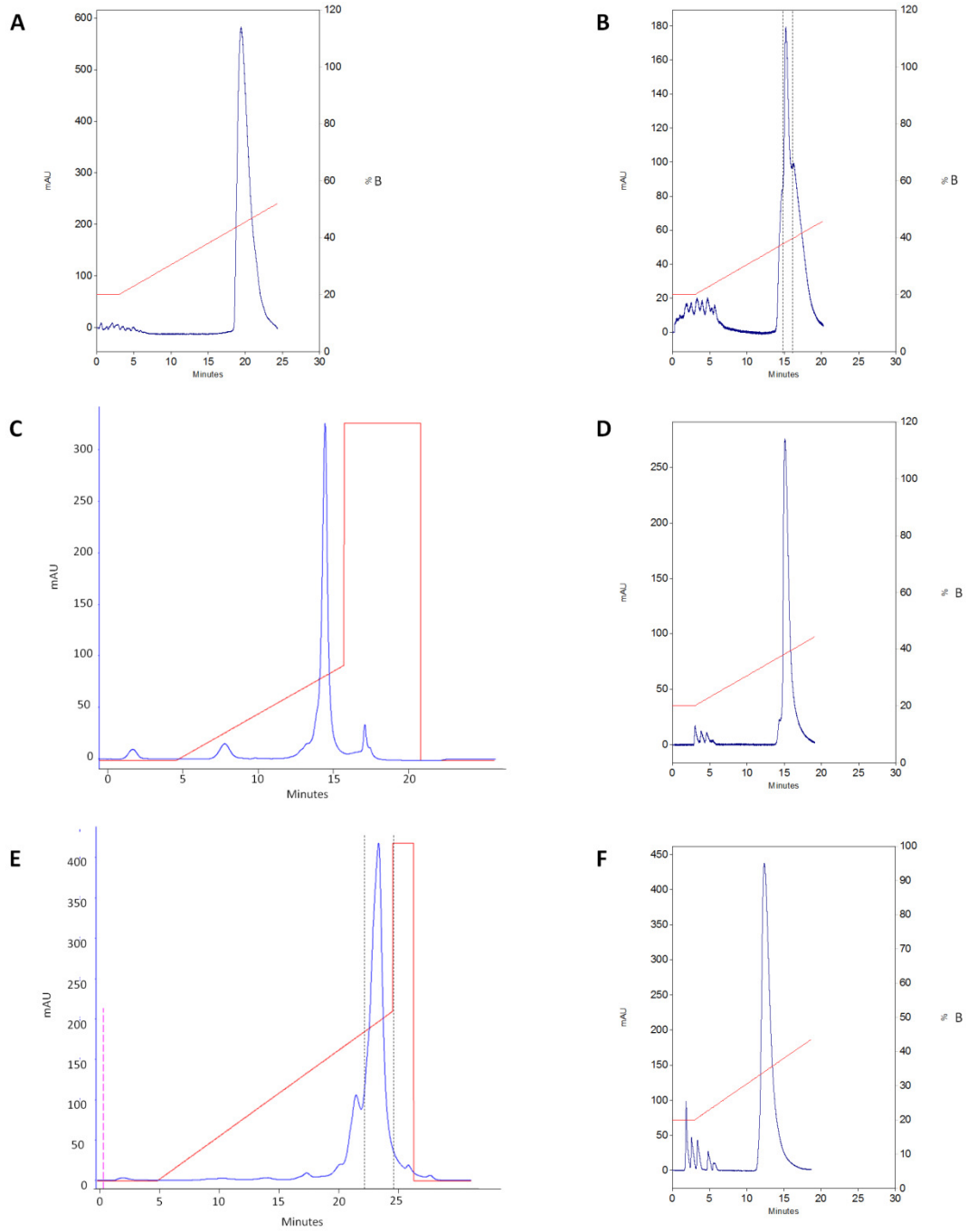


Figure 3.5 HPLC and FPLC chromatograms of FUN-EGF3. Blue and red lines indicate the absorbance at 280 nm and % of solution B respectively at a given time point; mAU = milli absorbance units. Dashed lines indicate fraction boundaries where there were significant additional species. Solution B was 80% acetonitrile, 0.1% TFA for reverse-phase HPLC and 20 mM MES pH 6.0, 2 M NaCl for MonoS FPLC purification. A) HPLC chromatogram of reduced FUN-EGF3. B) HPLC chromatogram of refolded protein. C) FPLC chromatogram of FUN-EGF3 treated with factor Xa. D) Final HPLC chromatogram of FUN-EGF3. E) FPLC chromatogram of His-FUN-EGF3. E) Final HPLC chromatogram of His-FUN-EGF3.

3.4.2.2 SDS-PAGE and mass spectrometry analysis of FUN-EGF3 and His-FUN-EGF3

The final products were analysed by SDS-PAGE (Section 2.2.10) and electrospray ionisation time of flight mass spectrometry (Section 2.2.11). Samples were run on SDS-PAGE under non-reducing and reducing conditions − without or with β-mercaptoethanol in the loading buffer respectively. β-mercaptoethanol should reduce all of the disulphide bonds. Therefore, comparing the mobility of the proteins under these two conditions provides limited information about their fold, as well as their size. The SDS-PAGE gel showed that His-FUN-EGF3 and FUN-EGF3 each comprised a single species, did not form disulphide-bonded oligomers or aggregates and was likely to have a homogeneous fold (Figure 3.6). Electrospray mass spectrometry confirmed that the protein fragments had the expected molecular weights (Table 3.1). The results given were the masses of the major species in maximum entropy-deconvoluted mass spectra (Section 2.2.11).

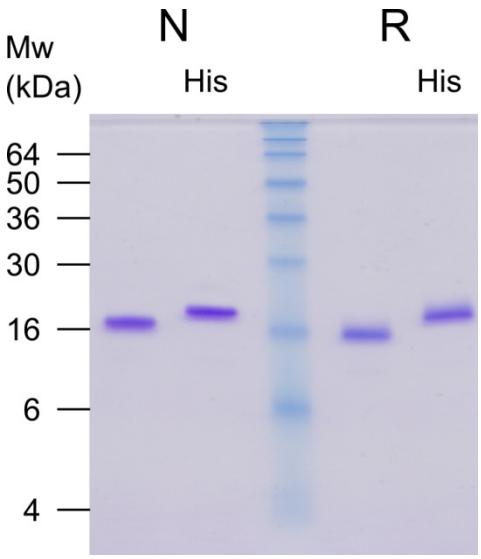


Figure 3.6 16% SDS-PAGE gel of FUN-EGF3 and His-FUN-EGF3 stained with Coomassie R250. Non-reducing and reducing sample conditions (absence and presence of β-mercaptoethanol are denoted by 'N' and 'R' respectively). His-FUN-EGF3 is labelled 'His'. The protein ladder is SeeBlue (Life Technologies); molecular weights (Mw) of markers are shown on the left-hand side.

Table 3.1 Mass spectrometry results for FUN-EGF3 and His-FUN-EGF3

Protein	Expected molecular weight (Da)	Measured molecular weight (Da)
FUN-EGF3	14082	14082
His-FUN-EGF3	15936	15938

3.4.2.3 Purification and refolding of cbEGF41-43-BirA

The purification protocol of cbEGF41-43-BirA was similar to that of FUN-EGF3 (Section 2.2). However, the HPLC elution profile of reduced cbEGF41-43-BirA was broader (Figure 3.7A). SDS-PAGE under reducing conditions confirmed that the right shoulder contained cbEGF41-43-BirA (not shown) and thus it was collected together with the main peak. Freeze-dried, reduced material was refolded by the same method as for FUN-EGF3 with two exceptions: the pH of the refolding buffer was 9.0, necessary to improve protein solubility, and $CaCl_2$ was added to the refolding mixture at a concentration of 50 mM (Section 2.2.6). Calcium ions should promote *in vitro* refolding by stabilising the calcium-dependent structure. The importance of calcium was demonstrated by comparing the HPLC traces of cbEGF41-43-BirA refolded with 50 mM $CaCl_2$ or 10 mM EDTA (Figure 3.7B). The profile was slightly broader for material refolded with EDTA than with $CaCl_2$, suggesting greater conformational heterogeneity.

Factor Xa cleavage of the His_6 tag was carried out at pH 8.3 in the absence of $CaCl_2$ in order to maintain protein solubility (Section 2.2.7). Proteolytic cleavage was incomplete at pH 9.5 when incubated overnight at 37°C (data not shown). The protein was then purified by anion exchange FPLC (Section 2.2.8). The peak on FPLC chromatogram was relatively broad (Figure 3.7C). It was established that the major fraction (denoted by dashed lines in Figure 3.7C) was likely to be the correct species using HPLC and SDS-PAGE analysis (data not shown; Joanne McNaught-Davis, Part II Project 2011). This fraction was collected and subjected to a final HPLC purification step (Figure

3.7D). For the purposes of the pull-down assay, some of the final product was biotinylated using the BirA enzyme (Section 2.2.9) and then desalted by HPLC (not shown).

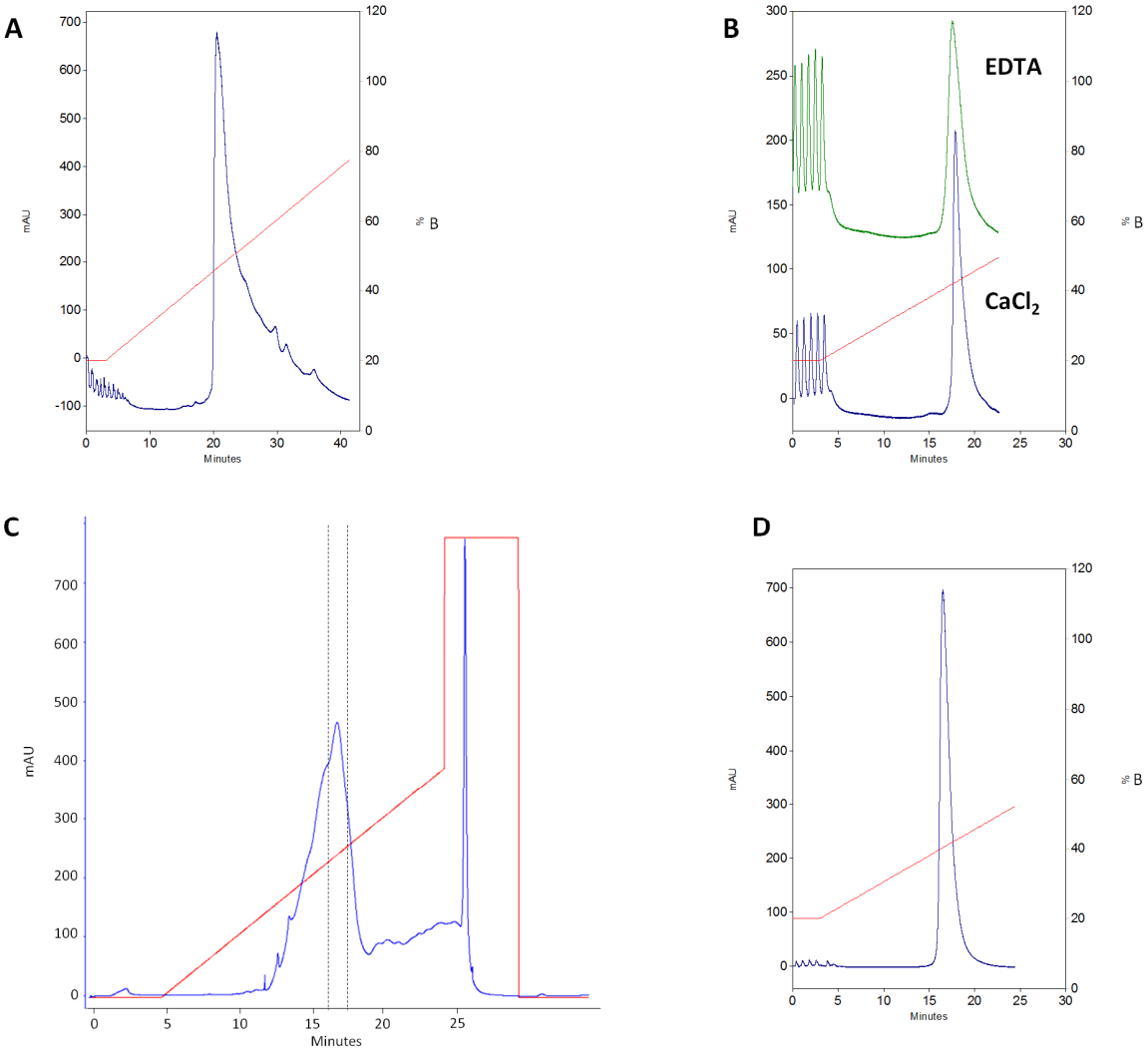


Figure 3.7 HPLC and FPLC chromatograms of cbEGF41-43-BirA. A) HPLC chromatogram of reduced material B) HPLC chromatogram of protein refolded in the presence of 50 mM $CaCl_2$ (blue trace) or 10 mM EDTA (green trace). C) FPLC chromatogram of factor Xa-treated cbEGF41-43-BirA. D) Final HPLC chromatogram after FPLC purification. HPLC purification conditions were as described in Figure 3.5. Buffer A for the MonoQ FPLC purification was 50 mM Tris pH 9.5, 1 mM $CaCl_2$ and buffer B comprised buffer A with 1M NaCl.

3.4.2.4 SDS-PAGE and mass spectrometry analysis of cbEGF41-43-BirA

The final purification product and the biotinylated protein (cbEGF41-43-BirA-bio) were also analysed by SDS-PAGE (Section 2.2.10) and mass spectrometry (Section 2.2.11). The presence of a single band on the gel under non-reducing and reducing conditions suggested that cbEGF41-43-BirA comprised a single species of the correct size (Figure 3.8). Mass spectroscopy confirmed the expected molecular weights of non-biotinylated and biotinylated material (Table 3.2).

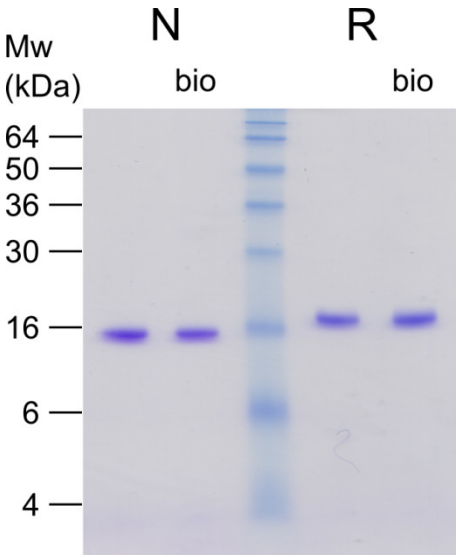


Figure 3.8 16% SDS-PAGE gel of cbEGF41-43-BirA and cbEGF41-43-BirA-bio under non-reducing (N) and reducing (R) conditions. The lanes for cbEGF41-43-BirA-bio are labelled 'bio'.

Table 3.2 Mass spectrometry results for cbEGF41-43-BirA

Protein	**Expected molecular weight (Da)**	**Measured molecular weight (Da)**
cbEGF4143-BirA	15305	15304
cbEGF41-43-BirA-bio	15531	15536

3.4.2.5 Purification, refolding and characterisation of cbEGF32-34-BirA

The purification protocol for non-BirA tagged cbEGF32-34 was established previously (Aileen McGettrick, D.Phil. Thesis, 2002) and was applied to cbEGF32-34-BirA. Results of purification and characterisation are shown in Appendix D.

3.5 Initial characterisation of FUN-EGF3 fragment by NMR spectroscopy

Very little is known about the structure of the FUN-EGF3 region of fibrillin (Section 1.3.5). As the domains do not contain calcium-binding consensus sequences, calcium binding cannot be used as a measure of folding. Initial NMR experiments were performed to provide insights into the structure of FUN-EGF3. The fragment is well suited for study by solution NMR spectroscopy because of its relatively small size (~14 kDa) – smaller molecules tumble faster, giving rise to slower relaxation, resulting in sharper peaks in spectra and a better signal-to-noise ratio. Even before performing detailed structural analysis, NMR spectra can provide 'fingerprints' of protein folding. The resonance frequency of a nucleus with non-zero spin is sensitive to the chemical structure of the protein and its fold, both of which influence its local magnetic environment.

NMR spectra of FUN-EGF3 were acquired at pH 5.4 (Section 2.6). This condition was chosen for consistency with other initial studies on N-terminal fragments (Ian Robertson, Sacha Jensen, Christina Redfield, personal communication). It maximises the solubility of the fragment and acidic pH is often more favourable for NMR studies as it minimises the rate of exchange of H^N nuclei with solvent, thereby also minimising line-broadening effects. A 1D ^1H spectrum of FUN-EGF3 in 100% D_2O displayed elements of a folded protein (Section 2.6.2.1; Figure 3.9A). The peaks were well dispersed and had chemical shifts consistent with a folded protein. For example, several peaks arising from $^1H^\alpha$ nuclei were downfield of the H_2O peak – an indicator of β-sheet

secondary structure. There were peaks in the methyl region with chemical shifts of less than 0.5 parts per million (ppm). Resonances arising from H^N protons were largely absent from this spectrum because of its acquisition in 100% D_2O, as the resonance frequencies for $^2H^N$ nuclei are very different from the frequency used to excite protons. However, this condition permits superior signal-to-noise acquisition for the protein as the H_2O content of the sample is negligible.

The ^1H-^{15}N HSQC experiment is particularly useful for assessing protein folding. The resulting spectrum should contain a peak for every non-proline residue, as well as a pair of peaks for each of the NH_2 groups in the side chains of glutamine and asparagine (Bodenhausen and Ruben, 1980; Norwood *et al.*, 1990). Moreover, $^1H^N$ and ^{15}N chemical shifts are particularly sensitive to folding. FUN-EGF3 was labelled with ^{15}N by expressing the protein in bacteria grown in minimal medium, supplemented with $^{15}NH_4Cl$ (Section 2.2.2). Mass spectrometry confirmed close to 100% labelling (Table 3.3) and a ^1H-^{15}N HSQC spectrum of the labelled protein was acquired (Section 2.6.2.3; Figure 3.9B). 154 peaks were expected from the sequence and 163 were counted in the spectrum. However, a number of peaks had weak intensities and may have resulted from a small fraction of misfolded material in the sample. The spectrum showed well-dispersed peaks, indicating that the protein was folded. There was also a pair of peaks arising from an NH_2 side-chain with unusual ^1H shifts (5.05 and 6.09 ppm) that may have resulted from the proximity of the residue to an aromatic ring. Acquisition of spectra between pH 4.0 and 10.0 in 0.5 pH unit intervals showed that pH had no significant effect on the protein fold (data not shown).

Table 3.3 Mass spectrometry result for ^{15}N-labelled FUN-EGF3

Expected molecular weight (Da)	Measured molecular weight (Da)
14269 (100% labelling)	14268

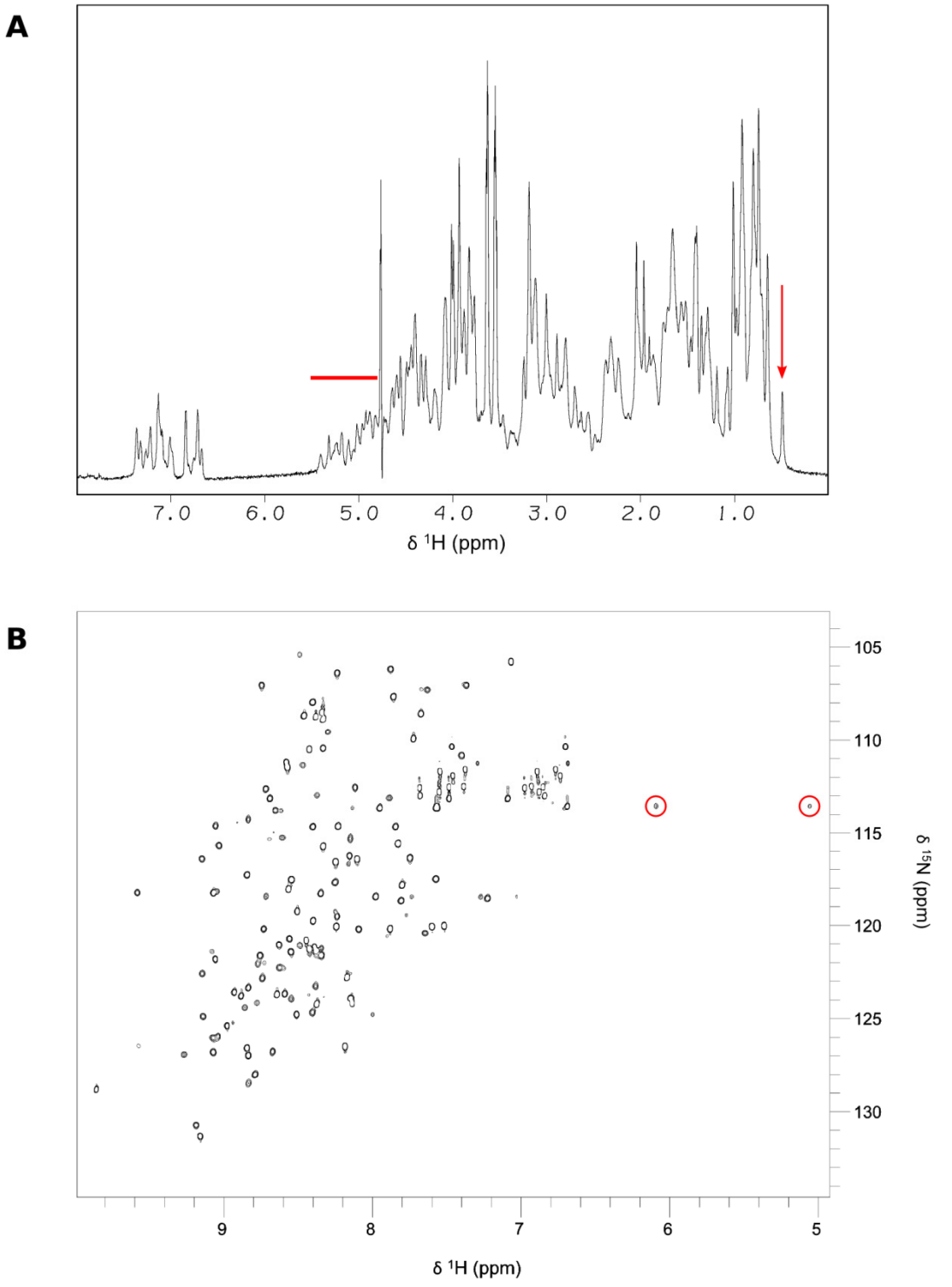


Figure 3.9 NMR spectra of FUN-EGF3. A) 1D ^{1}H spectrum of 0.5 mM protein at pH 5.4 in 100% D_2O, acquired at 25°C. The spectrum was rendered in Felix 2.3 (Accelrys). The bar indicates $^{1}H^{\alpha}$ resonances downfield of the H_2O peak (at 4.75 ppm) and the arrow highlights an upfield-shifted ^{1}H-methyl peak. B) 2D ^{1}H-^{15}N HSQC spectrum of 1.5 mM ^{15}N-labelled FUN-EGF3 at pH 5.4 in 95% H_2O/5% D_2O [v/v], acquired at 25°C). The spectrum was rendered in CCPN Analysis 2.1.5. The red circles highlight signals from an NH_2 group with unusual ^{1}H chemical shifts.

3.6 Characterisation of cbEGF41-43-BirA

Two methods were used to examine the effect of calcium ions on the structure of cbEGF41-43-BirA. The very low solubility of the fragment required the use of unusual buffer pH conditions (pH 3.8 or 9.5). It was therefore only possible to obtain qualitative rather than quantitative information about calcium binding.

3.6.1 Limited proteolysis

The effect of $CaCl_2$ on limited proteolysis of cbEGF41-43-BirA by endoproteinase Glu-C was investigated. Limited proteolysis is a useful assay of folding as the protein structure determines which potential cleavage sites are accessible to proteases. Previous studies have shown calcium-mediated protection of fibrillin or Notch fragments from proteolysis by Glu-C, an enzyme that cleaves peptide bonds C-terminally to glutamate residues (Cordle *et al.*, 2008b; Whiteman *et al.*, 2001). cbEGF41-43-BirA contains several potential sites for Glu-C cleavage (Figure 3.10A). The fragment was digested with Glu-C (Section 2.3) in the presence of either 50 mM $CaCl_2$ or 10 mM EDTA. Samples taken after 1, 2 and 4 hr were analysed by SDS-PAGE (Figure 3.10B). In the presence of $CaCl_2$, minimal digestion was observed, with the full-length fragment and minor degradation products present after 4 hr. By contrast, significant digestion occurred in the presence of the metal ion-chelating agent EDTA. Three fragments with different mobilities were observed, with approximate sizes of 4, 6 and 10 kDa. The difference between the two conditions provided evidence of calcium-dependent structure in cbEGF41-43-BirA.

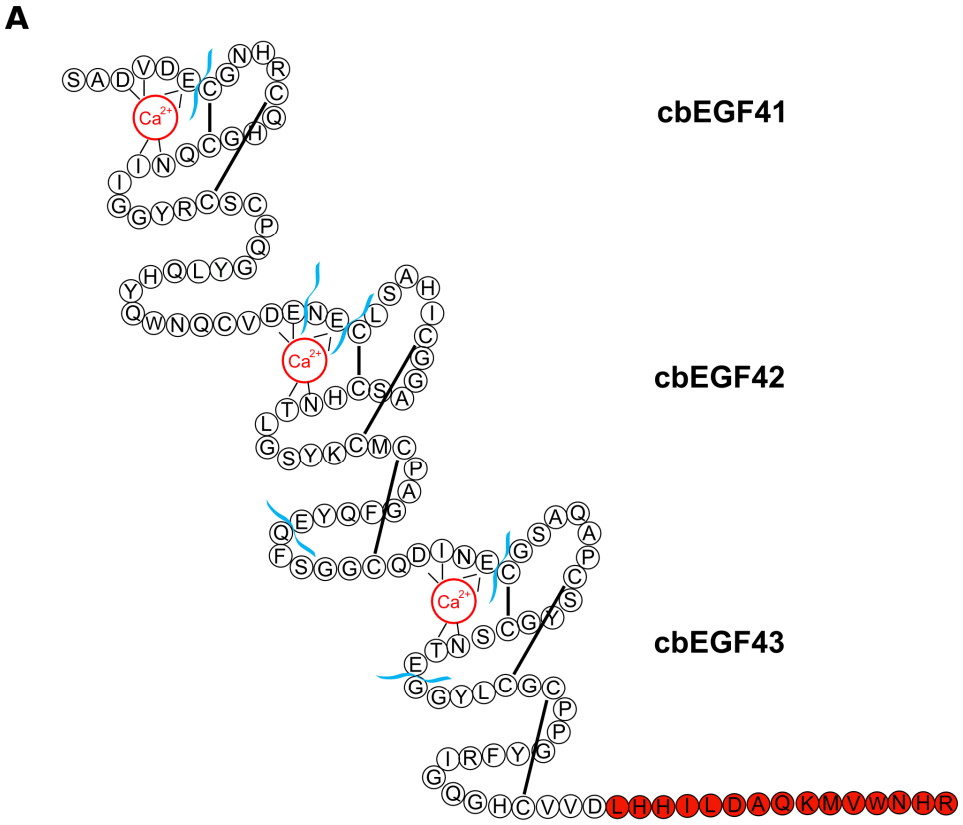


Figure 3.10 Effect of $CaCl_2$ on cbEGF41-43-BirA digestion by endoproteinase Glu-C. A) Potential major sites of peptide bond cleavage by Glu-C in cbEGF41-43-BirA, indicated by blue lines. Residues of the C-terminal BirA tag are coloured red. B) 16% SDS-PAGE gel of Glu-C-treated samples in the presence of either 50 mM $CaCl_2$ or 10 mM EDTA, buffered at pH 9.5, at 37°C. Samples were collected at 1, 2 and 4 hrs and immediately boiled after addition of reducing SDS-PAGE loading buffer. For the '0' time point, a sample was incubated without enzyme for 4 hr alongside the Glu-C-treated samples. This was to demonstrate the stability of the protein in the absence of Glu-C under these conditions.

3.6.2 NMR spectroscopy

NMR spectroscopy can be used to measure calcium binding by cbEGF domains by following peak shifts and intensity changes resulting from calcium-induced structural changes. The effect on the NMR spectrum depends on the affinity for calcium binding, which is related to the rate of exchange between the calcium-bound and unbound forms of the protein. High-affinity calcium binding means that exchange between the unbound and bound form is slower than for lower affinity binding. Depending on how the rate of exchange compares with the timescale of the NMR experiment, the exchange process can manifest itself in different ways for calcium binding. Slow exchange results in two peaks representing the bound and unbound form, the relative intensities of which depend on the bound fraction. Fast exchange results in a single peak with a chemical shift that is an average of the chemical shift of the individual species and varies with the bound fraction. The exchange rate can also be intermediate between these two extremes, resulting in a combination of shifts and intensity changes.

Spectra of cbEGF41-43-BirA were acquired at several $CaCl_2$ concentrations. Initial experiments were carried out with protein buffered at pH 9.5 and a series of 1D ^1H spectra were acquired. No peaks or intensity changes were observed, even with the addition of $CaCl_2$ to a concentration of 5 mM (data not shown). Similar experiments were then carried out at pH 3.8. This condition is not ideal for calcium binding studies as the pH is close to the pK_a of aspartate and glutamate side-chains, which are involved in calcium binding. However, it was necessary to maintain the protein in soluble form. 1D ^1H spectra showed some peak shifts and intensity changes on addition of $CaCl_2$ (data not shown).

1D experiments only provide limited information about structure because of peak overlap. It is also difficult to assign peaks to particular resonances using these spectra. Several studies have

employed 2D ^{1}H-^{1}H NOESY experiments to study fibrillin fragment structure. In particular, cross-peaks between nuclei in aromatic side-chains are good markers for calcium binding because of their involvement in interdomain packing and their proximity to calcium-binding sites. Unfortunately, the large number of aromatic residues in cbEGF41-43-BirA is likely to complicate analysis of titration data. Therefore, cbEGF41-43-BirA was labelled with ^{15}N, using the same method as for FUN-EGF3 (Section 3.5). Again, mass spectrometry confirmed labelling (Table 3.4).

Table 3.4 Mass spectrometry result for ^{15}N-labelled cbEGF-41-43BirA

Expected molecular weight (Da)	Measured molecular weight (Da)
15498 (100% labelling)	15498

^{1}H-^{15}N HSQC spectra of labelled cbEGF-41-43-BirA were acquired at a range of $CaCl_2$ concentrations at pH 3.8 (Section 2.6.10; Figure 3.11A). Peaks in the spectrum of the protein without added $CaCl_2$ were not well dispersed, suggesting that it was not well structured. Adding $CaCl_2$ resulted in peak shifts and intensity changes, but high concentrations (greater than 10 mM) were needed to see significant effects; the low pH is likely to be responsible for the apparently low affinity. In particular, peaks arising from glutamine and asparagine side chains became more dispersed. This is consistent with their expected proximity to the calcium-binding sites in cbEGF41-43-BirA. A titration with $MgCl_2$ was also performed, in order to demonstrate that the effect of $CaCl_2$ did not result from a change in ionic strength, rather than calcium binding. There were differences between spectra of protein with 50 mM $CaCl_2$ and the same concentration of $MgCl_2$. Therefore, it can be concluded that cbEGF41-43-BirA binds calcium, but the conditions required for the experiment prevented quantitative or domain-specific information from being obtained.

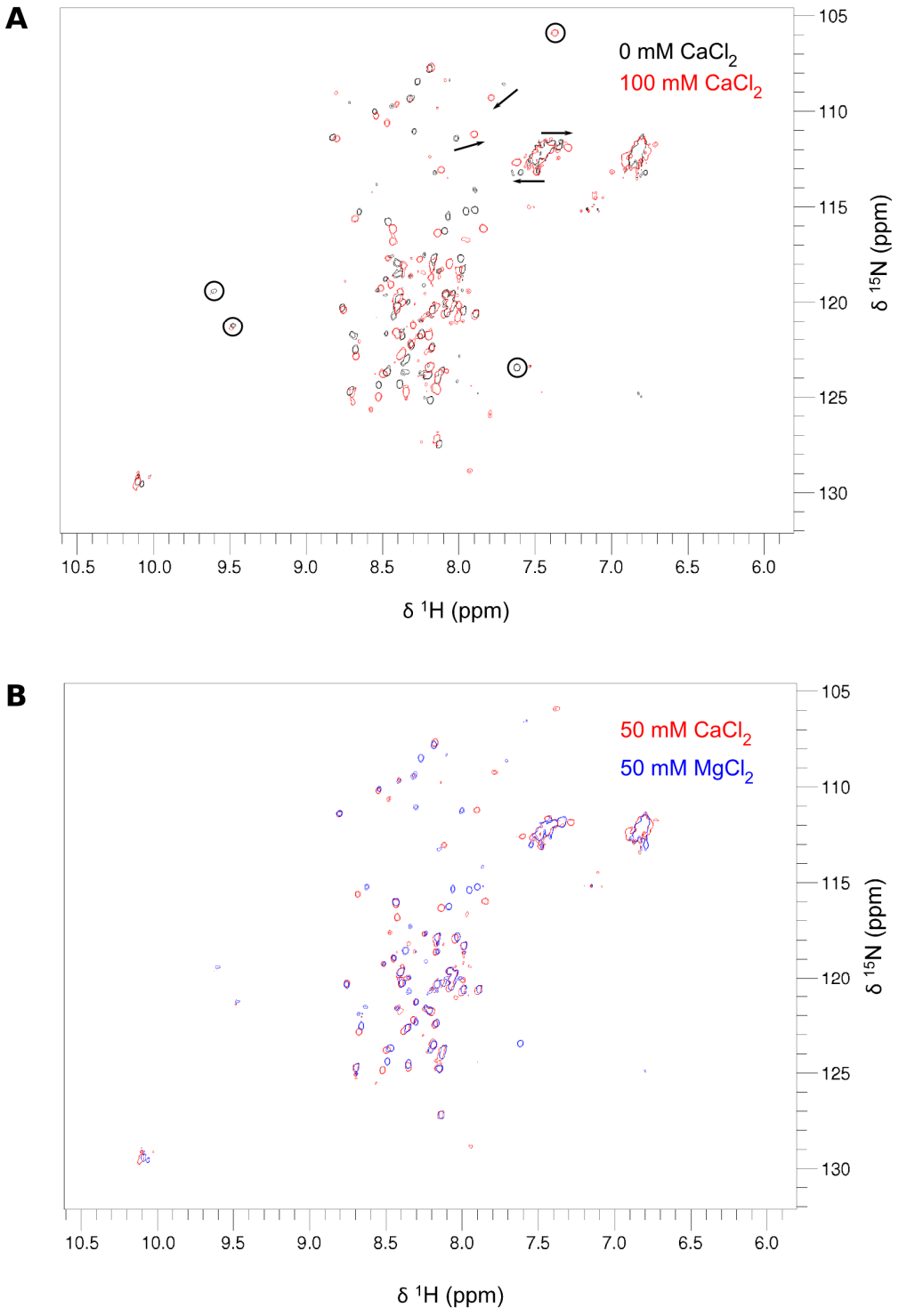


Figure 3.11 Effect of $CaCl_2$ on NMR spectra of ^{15}N-labelled cbEGF41-43-BirA. Measurements were carried out with protein at pH 3.8 in 95% H_2O/5% D_2O (v/v), without added NaCl at 25°C. A) 2D ^{1}H-^{15}N HSQC spectra of protein with no added calcium and 100 mM $CaCl_2$, plotted on the same axes. Several peak shifts and intensity changes identified by a $CaCl_2$ titration are highlighted with arrows and circles respectively. B) HSQC spectra of protein with 50 mM $CaCl_2$ or $MgCl_2$, plotted on the same axes.

3.7 Interaction of immobilised cbEGF41-43-BirA with His-FUN-EGF3

3.7.1 Pull-down assay

Pull-down experiments were performed using the scheme in Figure 3.2 (Section 2.4). cbEGF41-43-BirA-bio or negative control fragment cbEGF32-34-BirA-bio were immobilised on streptavidin-coated beads in binding buffer at pH 9.5. After washing off excess unbound protein, the beads were transferred to binding buffer at pH 7.4 and then incubated with His-FUN-EGF3. The Jagged-1 DSL-EGF3 fragment with an N-terminal His_6 tag (His-DSL-EGF3), kindly provided by Dr Mouli Chillakuri, was used a negative control. DSL-EGF3 comprises the DSL (Delta/Serrate/Lag2) domain and three non-calcium binding EGF-like domains (Cordle *et al.*, 2008a) − similar to the expected structure of FUN-EGF3. Analysis by Western blotting demonstrated binding of His-FUN-EGF3 to bead-immobilised cbEGF41-43-BirA-bio (Figure 3.12A). Binding was not dependent on pH, as His-FUN-EGF3 also bound to cbEGF41-43-BirA-bio at pH 9.5 (data not shown). Moreover, His-FUN-EGF3 did not bind non-specifically to the beads alone and there was minimal binding to cbEGF32-34-BirA-bio. His-DSL-EGF3 did not bind significantly to cbEGF41-43-BirA-bio either. Therefore, the pull-down demonstrated a specific interaction between FUN-EGF3 and cbEGF41-43-BirA.

The dependence of the interaction on calcium was also investigated. His-FUN-EGF3 was incubated with immobilised cbEGF41-43-BirA-bio in buffer containing either 2 mM $CaCl_2$ or 10 mM EDTA and binding was compared. Initial experiments showed unexpectedly that EDTA did not abrogate binding (data not shown). To eliminate the possibility that calcium remained bound to a high-affinity site in cbEGF41-43-BirA during purification, protein was pre-treated with 50 mM EDTA at pH 9.5, acidified and desalted by HPLC, freeze-dried and then immobilised in the presence of EDTA in the pull-down. An interaction was still observed under these conditions (Figure 3.12B).

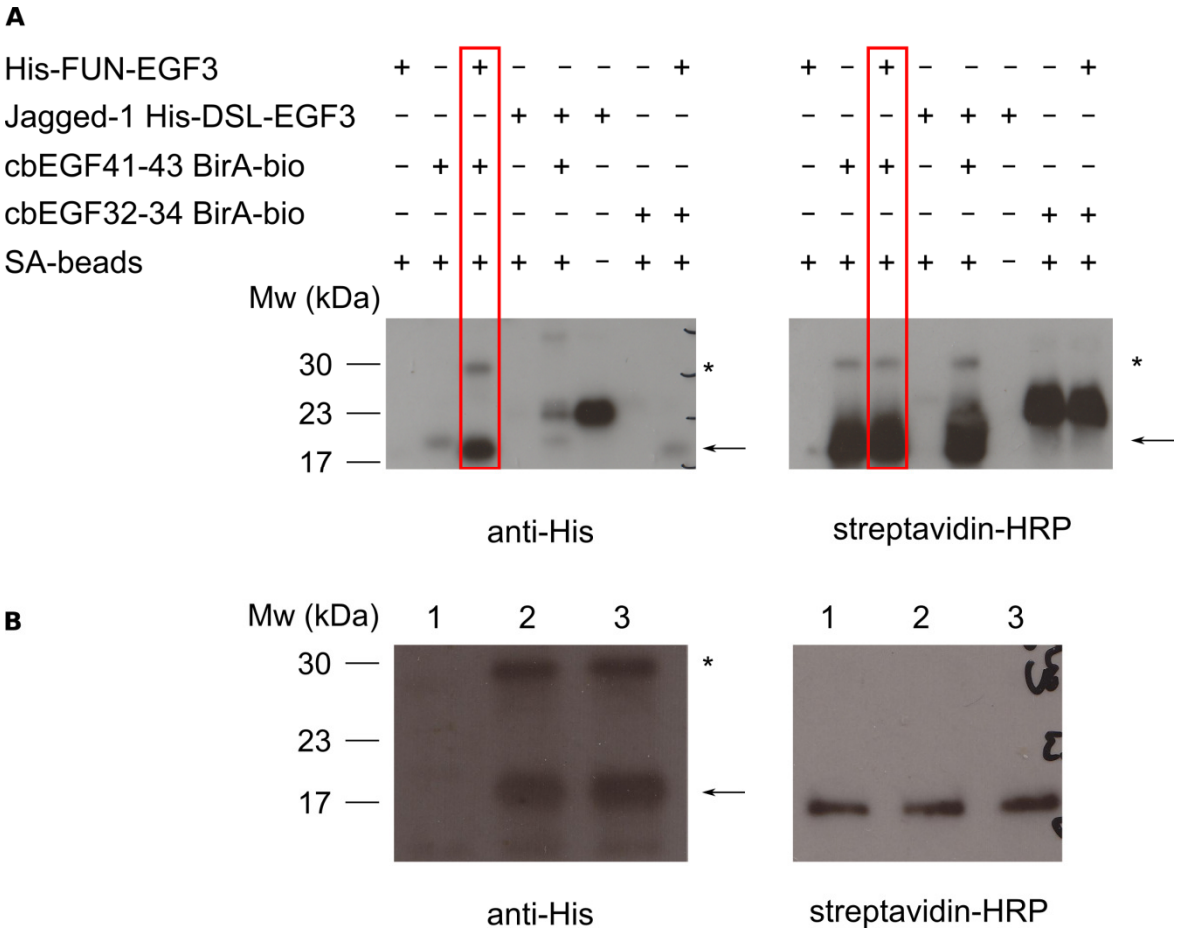


Figure 3.12 Pull-down experiments using immobilised cbEGF41-43-BirA-bio and His-FUN-EGF3 in solution. A) Western blots of supernatant samples after boiling beads in reducing SDS-PAGE loading buffer. cbEGF41-43-BirA-bio or cbEGF32-34-BirA-bio (negative control) were immobilised on streptavidin-coated beads in buffer at pH 9.5. Beads were incubated with His-FUN-EGF3 or Jagged-1 His-DSL-EGF3 (negative control) in buffer at pH 7.4. 100 ng His DSL-EGF3 was also loaded to show the expected position of the protein. Red boxes highlight the sample from the pull-down in which bead-immobilised cbEGF41-43-BirA-bio was incubated with His-FUN-EGF3. B) Calcium-independence of the interaction, using cbEGF41-43-BirA pre-treated with EDTA, as described in the text. 1 = cbEGF41-43-BirA-bio + beads only; 2 = cbEGF41-43-BirA-bio + His-FUN-EGF3 incubated in buffer containing 2 mM $CaCl_2$; 3 = cbEGF41-43-BirA-bio + His-FUN-EGF3 incubated in buffer containing 10 mM EDTA. The positions of molecular weight markers (Mw) are indicated on the left-hand side. Arrows indicate bands that ran at the expected molecular weights of His-FUN-EGF3 and cbEGF41-43-BirA-bio. * indicates lower mobility bands of unknown origin, as described for the pull-down experiment in the other orientation (Section 3.2).

3.7.2 Surface plasmon resonance-based interaction assay

In an attempt to obtain more quantitative data about the interaction between FUN-EGF3 and cbEGF41-43, SPR was employed. SPR can be used to detect interactions between macromolecules on a surface, which result in a change in refractive index (Figure 3.13). Light undergoes total internal reflection in a prism when its angle of incidence is greater than a critical angle, producing a so-called evanescent field. If the base of the prism is covered with a metal such as gold, the evanescent field becomes coupled to an electromagnetic wave at the surface of the metal (Stenberg *et al.*, 1991). Absorption by the surface occurs at a particular angle of the incident light and a maximum is observed when the minimum intensity of reflected light is measured. This angle is determined by the refractive index of the medium close to the surface. The Biacore T100 machine used here detects the change in refractive index as changes in the angle required for surface plasmon resonance. Importantly, the change, measured as arbitrary response units (RU) is proportional to the size and number of molecules in contact with the surface.

In an SPR-based interaction assay, one interaction partner is immobilised on the surface and the other is flowed over the surface in solution − the ligand and analyte respectively. The response measured for a reference flow cell is subtracted from the response measured for the flow cell containing the immobilised interaction partner; this corrects for bulk changes in refractive index. Here, for reasons already given (Section 3.3), cbEGF41-43-BirA was immobilised on the surface of the sensor chip. Two possibilities for immobilisation were explored: first, through the biotin-streptavidin interaction; second, using direct, covalent coupling of protein to the chip surface. In both cases, a Biacore S-series CM5 sensor chip was used. The chip surface comprises dextran functionalised with carboxyl groups, which can be covalently coupled to amine groups on a protein if the carboxyl groups are first activated using NHS and EDC (Section 2.5.1).

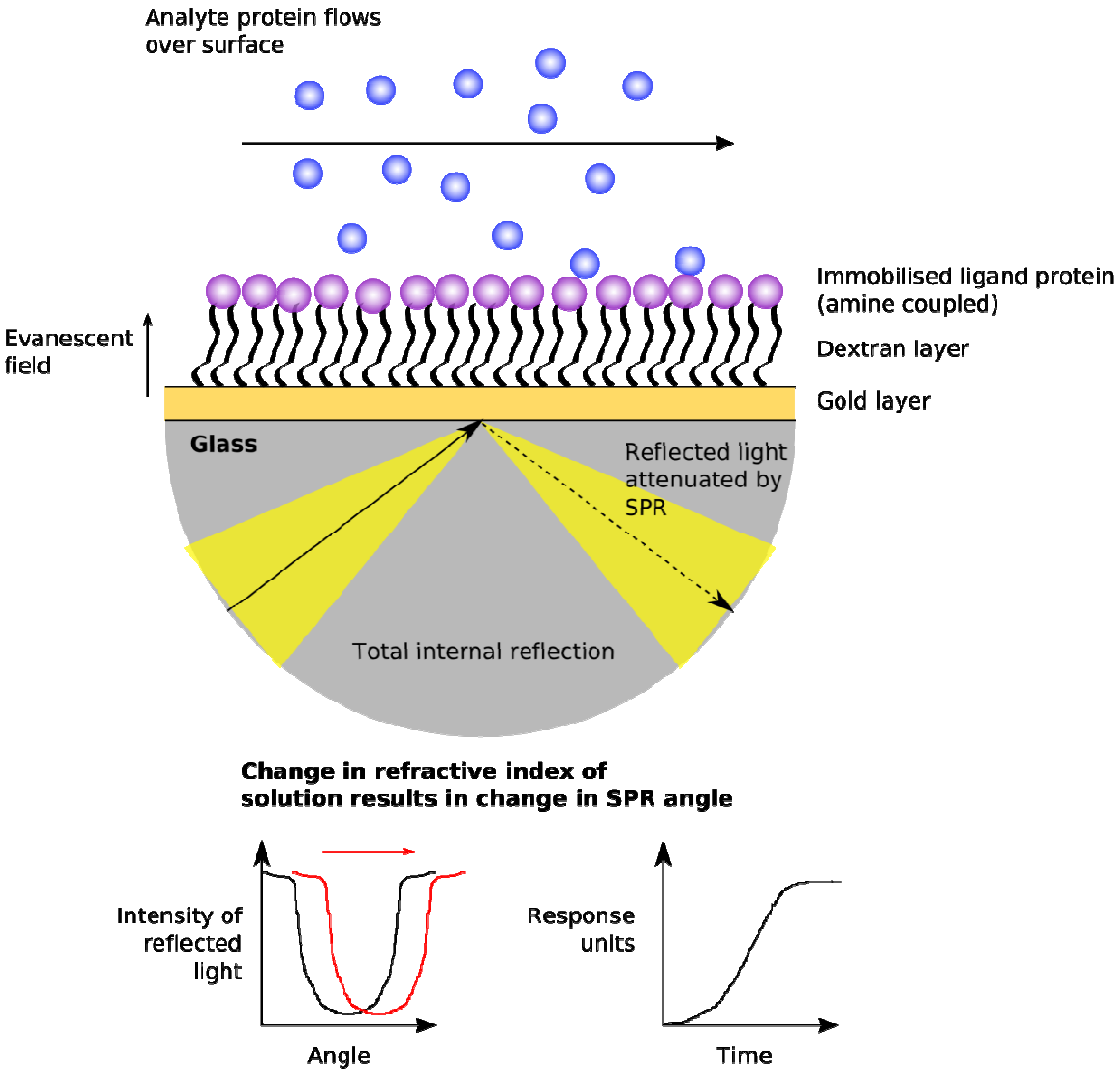


Figure 3.13 Illustration of the SPR-based protein-protein interaction assay. One protein, the ligand, is immobilised on the sensor chip surface through covalent coupling. This surface sits on top of a gold layer that is responsible for the SPR effect. Light passing through a glass prism at the critical angle undergoes total internal reflection. Where the angle of incidence is close to the angle required for SPR, the maximum amount of light is absorbed by the surface and the minimum amount is reflected. A change in refractive index, resulting from a change in composition of the solution above the surface or binding of an analyte molecule, is detected through the change in the angle at which SPR occurs.

While the streptavidin-biotin method offers the potential advantage of immobilising biotinylated protein molecules in similar orientations, it was found to be unsuitable for the interaction between cbEGF41-43-BirA and FUN-EGF3. FUN-EGF3 stuck non-specifically to streptavidin on the reference flow cell and the harsh conditions required for complete regeneration of the chip after the addition of FUN-EGF3 (50 mM HCl) resulted in a reduction in the baseline response level (data not shown). Instead, cbEGF41-43-BirA and negative control protein cbEGF32-34-BirA were coupled directly to the chip (Section 2.5.1). Several concentrations of FUN-EGF3 were sequentially flowed over the surface, with regeneration steps in between, and a concentration-dependent response was observed in the flow cell with immobilised cbEGF41-43-BirA (Section 2.5.2; Figure 3.14A). There was also a response on the cbEGF32-34-BirA flow cell, albeit smaller (Figure 3.14B). Addition of 0.05% (v/v) Tween-20 detergent to the buffer, also used in the pull-down assay, had no effect on the response profiles (data not shown). The LTBP-1 EGF3-cbEGF15-BirA fragment, kindly provided by Ian Robertson, was used as a negative control in solution and showed almost no response at equivalent concentrations on the flow cell containing cbEGF41-43-BirA (Figure 3.14B). Experiments testing the calcium dependence of the interaction were inconclusive, but EDTA did not abrogate binding (data not shown).

Unfortunately it was not possible to obtain quantitative information about the affinity of the interaction. Curve fitting to determine on and off rates for binding produced values that varied with concentration (not shown). The response profile also suggested that the interaction was heterogeneous, as there appeared to be fast and slow dissociation processes. Analysis of response levels at equilibrium also produced variable results. Moreover, the unknown stoichiometry of the interaction means that caution must be exercised in interpreting these data. Despite the variability of the data, the method has potential for use in a semi-quantitative manner. For example, response levels for fragments with single residue substitutions could be compared with unmodified protein, provided that they are tested on the same chip.

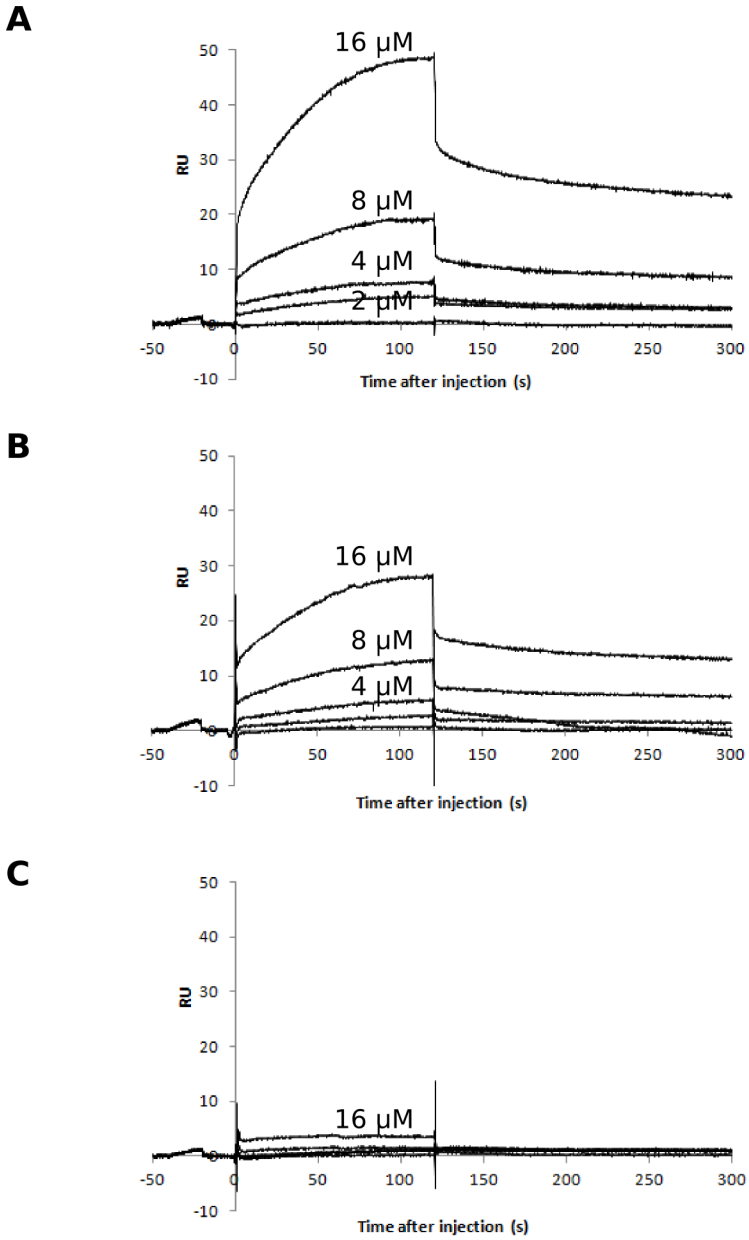


Figure 3.14 Results of the SPR-based interaction assay. ~800 Response units (RU) of cbEGF41-43-BirA and cbEGF32-34-BirA were immobilised on separate flow cells on the sensor surface (Section 2.5.1). The reference flow cell was blocked with ethanolamine. Freeze-dried aliquots of protein were dissolved in running buffer comprising 50 mM Tris pH 7.4, 150 mM NaCl, 2 mM $CaCl_2$ and diluted to the desired concentration. Representative blank-subtracted Biacore sensograms are shown for a series of concentrations (0, 2, 4, 8 and 16 µM) of protein solution. The analyte sample was injected at time 0 at a flow rate of 30 µl/min and the contact time was 120 s. A) FUN-EGF3 as the analyte and cbEGF41-43-BirA as the ligand. B) FUN-EGF3 as the analyte and cbEGF32-34-BirA as the ligand. C) LTBP-1 EGF3-cbEGF15 as the analyte and cbEGF41-43-BirA as the ligand.

3.8 Discussion

In this chapter, fibrillin N- and C-terminal fragments were produced, corresponding to the previously identified minimal interacting regions. FUN-EGF3 and cbEGF41-43 fragments had been made before the start of this project. However, an interaction could only be demonstrated if the fragments were incubated in buffer at pH 9.5 because of the low solubility of the cbEGF41-43 fragment. A new cbEGF41-43 fragment with a C-terminal BirA tag was cloned, expressed, purified and refolded *in vitro*. The aim was to perform the interaction assay under more physiologically relevant conditions by immobilising cbEGF41-43-BirA using the biotin-streptavidin interaction. A negative control fragment, cbEGF32-34-BirA, was also produced for use in interaction assays. In addition, FUN-EGF3 without a BirA tag was expressed and purified for the purpose of the pull-down assay and initial characterisation of its structure by NMR spectroscopy.

The FUN-EGF3 and cbEGF41-43-BirA fragments were characterised using several methods. While the prokaryotic expression and refolding system has the advantage of producing large quantities of homogeneous protein, it is important to verify that the material has a native-like fold. SDS-PAGE and mass spectrometry demonstrated that the fragments had the expected sizes and were homogeneous under non-reducing, denaturing conditions. While the structure of FUN-EGF3 is unknown, NMR spectroscopy suggested that it was folded. 1D and 2D spectra had well dispersed peaks with chemical shifts that differed from random coil values. This indicated that FUN-EGF3 is a promising candidate for structure determination by NMR.

For cbEGF41-43-BirA, folding was assessed by examining the effect of calcium. The cbEGF domains are predicted to bind calcium ions as they contain the appropriate consensus residues (Section 1.3.1). However, the insolubility of the fragment between pH ~4.0 and ~9.5 in the presence of $CaCl_2$ presented a significant obstacle to its characterisation and required the use of

unusual conditions. This behaviour is consistent with the role of these domains in aggregation of the fibrillin C-terminus, which is functionally important for assembly (Hubmacher *et al.*, 2008). Protection from proteolysis by Glu-C in the presence of $CaCl_2$ provided evidence of calcium-dependent structure. As potential cleavage sites are spread across the molecule, this did not provide domain-specific information.

The effect of $CaCl_2$ was investigated further using NMR spectroscopy, which can give atomic resolution information if peak assignments are available. Cross-peaks between nuclei in aromatic side-chains in 2D homonuclear ^{1}H NOESY spectra are often used as markers in calcium titrations (Jensen *et al.*, 2005; Knott *et al.*, 1996). The absence of resonance assignments for cbEGF41-43-BirA and the large number of aromatic residues in the construct meant this was impractical. Therefore, titrations were performed using HSQC spectra of ^{15}N-labelled protein. The calcium-binding affinity was likely to be significantly reduced at pH 3.8 compared with physiological conditions. However, the protein remained soluble, and peak shifts and intensity changes were observed upon addition of $CaCl_2$, but not $MgCl_2$. Unfortunately, it was still not possible to obtain domain-specific information. The absence of an effect of $CaCl_2$ at pH 9.5 appeared to be inconsistent with the proteolysis data. However, calcium binding may result in oligomerisation, rendering the protein 'invisible' to solution state NMR. Further investigation of the aggregation properties of cbEGF41-43-BirA is needed.

Following their characterisation, the interaction of FUN-EGF3 and cbEGF41-43-BirA was demonstrated using the pull-down assay. As expected, immobilising biotinylated cbEGF41-43-BirA on streptavidin-coated magnetic beads allowed the assay to be carried at pH 7.4. The use of negative control proteins cbEGF32-34-BirA-bio and Jagged-1 His-DSL-EGF3 demonstrated the specificity of the interaction. Moreover, the interaction has now been shown with both immobilised FUN-EGF3 and cbEGF41-43. The apparent independence of the interaction on

calcium in the pull-down assay cannot be directly compared with previous studies, as they used larger fragments (Section 1.7.2). It is possible that, while cbEGF41-43-BirA is essential for the interaction, calcium dependence is conferred by other regions of the fibrillin. One study using a C-terminal fragment encompassing the TB7-unique C-terminus found that the interaction with N-terminus could only be partially abrogated by EDTA (Marson *et al.*, 2005).

SPR was used to probe the interaction further. Binding of FUN-EGF3 to immobilised cbEGF41-43-BirA was observed, but there was also a smaller response on the flow cell with immobilised cbEGF32-34-BirA. This did show a degree of specificity, but meant that the results needed to be interpreted cautiously. The analyte control fragment LTBP-1 EGF3-cbEGF15 did not show any significant binding to cbEGF41-43-BirA or cbEGF32-34-BirA, indicating that the non-specific response was due to FUN-EGF3, rather than the properties of the immobilised fragments. In addition, the apparent heterogeneity of the interaction profile meant that traces could not be reliably interpreted in a quantitative fashion. The pull-down is likely to have shown greater specificity because of the inclusion of BSA as a blocking agent in the binding buffer (Section 2.4.1). Multimerisation of cbEGF41-43-BirA-bio through immobilisation on the beads may have also enhanced the affinity, as has been previously shown for the Notch-Jagged interaction (Jemima Cordle, D.Phil. Thesis, 2005). In addition, the available binding sites may be different in the two assays because of the different methods of cbEGF41-43-BirA immobilisation. Despite the lack of quantitative information about binding, the assay could potentially provide semi-quantitative information when comparing modified fragments. This is explored further in Chapter 5.

In summary, fragments corresponding to the minimal regions needed for binding in the fibrillin N-C terminal interaction were produced and characterised. The interaction was been confirmed using pull-down and SPR assays. Importantly, NMR spectra of the FUN-EGF3 fragment were of suitable quality for structure determination. This is the subject of Chapter 4.

Chapter 4

Structure and dynamics of
the FUN-EGF3 fragment

4.1 Introduction

Structural information is critical for understanding the organisation and assembly of fibrillin in microfibrils. There are already high-resolution structures of fibrillin fragments containing cbEGF, TB and hyb domains (Section 1.3). However, there are no structures of domains in the N- and C-terminal regions, including the FUN and first three non-calcium-binding EGF-like domains (FUN-EGF3). FUN-EGF3 is the minimal region required for binding to the fibrillin C-terminus (Chapter 3). These domains are also implicated in the interaction of fibrillin with accessory factors in the ECM, for example the BMPs (Section 1.8). Although the EGF-like domains are expected to have the standard disulphide bonding pattern of C1-C3, C2-C4, C5-C6, the fold of the FUN domain is unknown. There is also no information about how these domains pack together. Interdomain packing interactions determine the relative orientations of domains and therefore the overall conformation of fibrillin (Section 1.4). They are also likely to influence interactions with other molecules whose binding sites lie across multiple fibrillin domains.

Attempts to crystallise the FUN-EGF3 fragment were unsuccessful (Section 1.10). The quality of 1D and 2D spectra of the protein indicated that structure determination using NMR spectroscopy was a promising alternative (Section 3.5). Structures of the cbEGF32-cbEGF33, cbEGF12-cbEGF13 and TB6 fibrillin fragments were solved previously using NMR (Downing *et al.*, 1996; Smallridge *et al.*, 2003; Yuan *et al.*, 1997). NMR has the additional advantage of providing information about dynamics. Structure determination employs restrained molecular dynamics calculations to fold a molecule *in silico* using restraints derived from NMR experiments and defined parameters for covalent geometry. Restraints are incorporated such that deviations from allowed values introduce energy penalties. The system moves to fold the molecule into a conformation that best satisfies the input restraints − a structure with a low energy. Multiple calculations are performed,

resulting in an ensemble of low-energy structures. This is different from X-ray crystallography, which produces a single model using the electron density derived from X-ray diffraction patterns.

Structure determination by NMR requires the acquisition of many datasets. A flowchart for the process is shown in Figure 4.1.

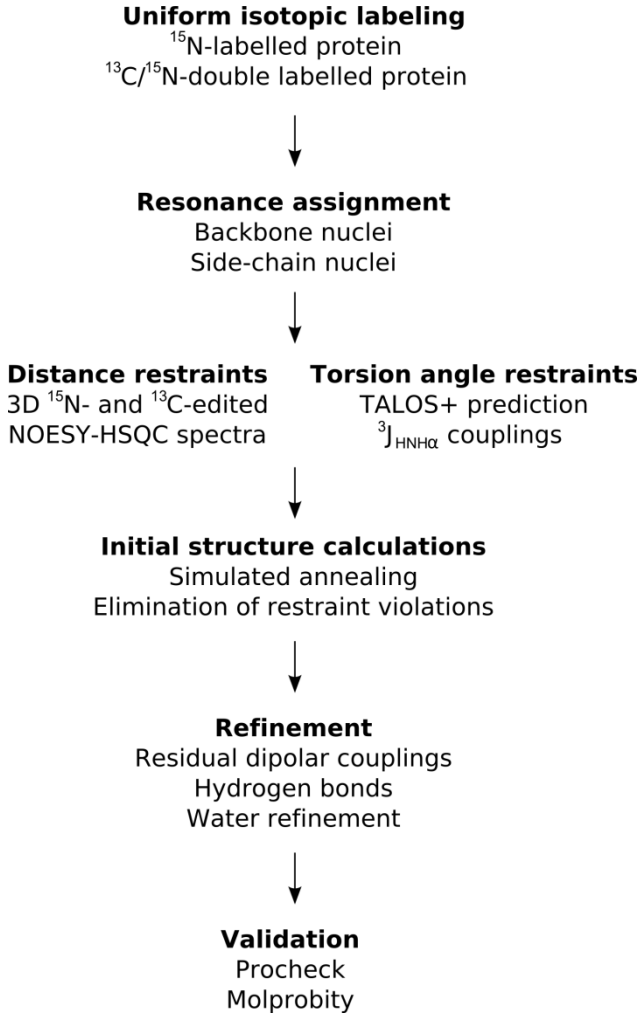


Figure 4.1 NMR structure determination. The FUN-EGF3 fragment was labelled with ^{15}N and ^{13}C. NMR datasets were collected and analysed to assign resonances from ^{1}H, ^{15}N and ^{13}C nuclei in the protein. Assignments allowed analysis of other datasets: NOESY spectra provided inter-^{1}H distance restraints; measurement of scalar coupling constants from HMQC-J spectra, provided φ angle restraints. The TALOS+ program also predicted φ and ψ torsion angles from chemical shifts. Restraints were implemented in simulated annealing calculations to calculate initial structures. Structures were refined with hydrogen bond and residual dipolar coupling restraints. They were then validated using Procheck and Molprobity.

4.2 Isotopic labelling of FUN-EGF3 for structure determination

Chemical shift assignments are needed to interpret NMR data in terms of structure. In theory, ^{1}H NMR alone can be used to define protein structure, but in practice, chemical shift degeneracy, resulting in peak overlap in spectra, makes this very challenging. Multidimensional NMR experiments are extremely powerful for assignment because cross-peaks in spectra are spread across several dimensions. Sequential assignment and triple resonance methods (Section 4.3) require isotopic labelling. H, C, N and O atoms make up the peptide backbone, but none of their naturally most abundant isotopes except ^{1}H has a nuclear spin of 1/2. Fortunately, the nuclei of ^{13}C and ^{15}N isotopes have spin 1/2 and can be readily incorporated into proteins. ^{15}N single-labelling of FUN-EGF3 was described previously (Section 3.5). Double-labelling of FUN-EGF3 with ^{13}C and ^{15}N was achieved by expressing protein in cells grown in minimal medium supplemented with ^{15}N-NH_4Cl and ^{13}C-labelled D-glucose (Section 2.2). Mass spectrometry of the final product again confirmed labelling (Table 4.1).

Table 4.1 Mass spectrometry result for $^{13}C/^{15}N$-labelled FUN-EGF3

Expected molecular weight (Da)	Measured molecular weight (Da)
14835 (100% labelling)	14830

4.3 Assignment of resonances in FUN-EGF3

Resonance assignment was carried out in several stages (experiments illustrated in Figure 4.2). All assignment experiments were performed with protein samples at pH 5.4 in 100% D_2O or 5% D_2O/95% H_2O (v/v) (Section 2.6.1). Sequential assignment was first performed using data for ^{15}N-labelled protein. 3D ^{15}N-edited NOESY-HSQC and TOCSY-HSQC spectra (Marion *et al.*, 1989) were acquired, as well as an HSQC-NOESY-HSQC spectrum (Frenkiel *et al.*, 1990; Ikura *et al.*, 1990)

(Section 2.2.6.4). The TOCSY spectrum was used to obtain spin system information, while cross-peaks in the NOESY allowed spin systems to be linked through sequential NOEs (Figure 4.3). Cross-peaks in the HSQC-NOESY-HSQC spectrum gave further information about sequential connections between spin systems with similar $^{1}H^{N}$, but different ^{15}N chemical shifts (not shown).

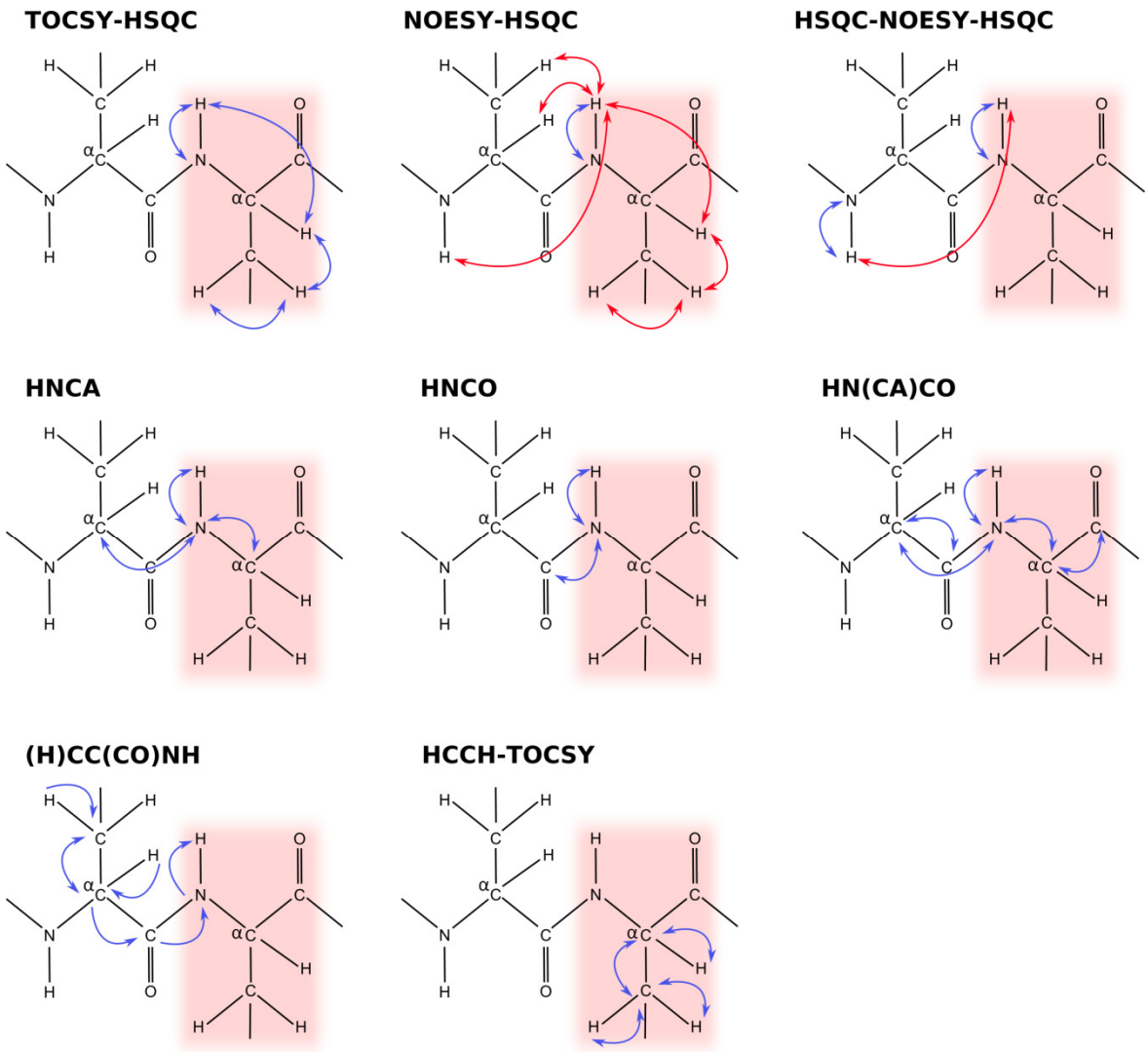


Figure 4.2 3D NMR experiments used in chemical shift assignment of FUN-EGF3. Blue and red arrows represent the transfer of magnetisation between nuclei through bonds (J coupling) and through space (cross-relaxation) respectively. Residue *i* of the polypeptide chain is highlighted in red. ^{15}N-edited TOCSY-HSQC: correlates protons with the $^{1}H^{N}$ and ^{15}N nuclei of the same residue. ^{15}N-edited NOESY-HSQC: correlates $^{1}H^{N}$ and ^{15}N nuclei with ^{1}H nuclei in close proximity to $^{1}H^{N}$ (< ~6.0 Å). HSQC-NOESY-HSQC: correlates ^{15}N nuclei of residues *i+1* and *i-1* with $^{1}H^{N}$ and ^{15}N nuclei of residue *i*. HNCA: correlates $^{13}C^{\alpha}$ of residues *i* and *i-1* with $^{1}H^{N}$ and ^{15}N of residue *i*. HNCO: correlates ^{13}C' of residue *i -1* with $^{1}H^{N}$ and ^{15}N of residue *i*. HN(CA)CO: correlates ^{13}C' of residues *i* and *i-1* with $^{1}H^{N}$ and ^{15}N of residue *i*. (H)CC(CO)NH: correlates aliphatic ^{13}C nuclei of residue *i-1* with $^{1}H^{N}$ and ^{15}N of residue *i*. HCCH-TOCSY: correlates aliphatic ^{1}H and ^{13}C nuclei of the same residue.

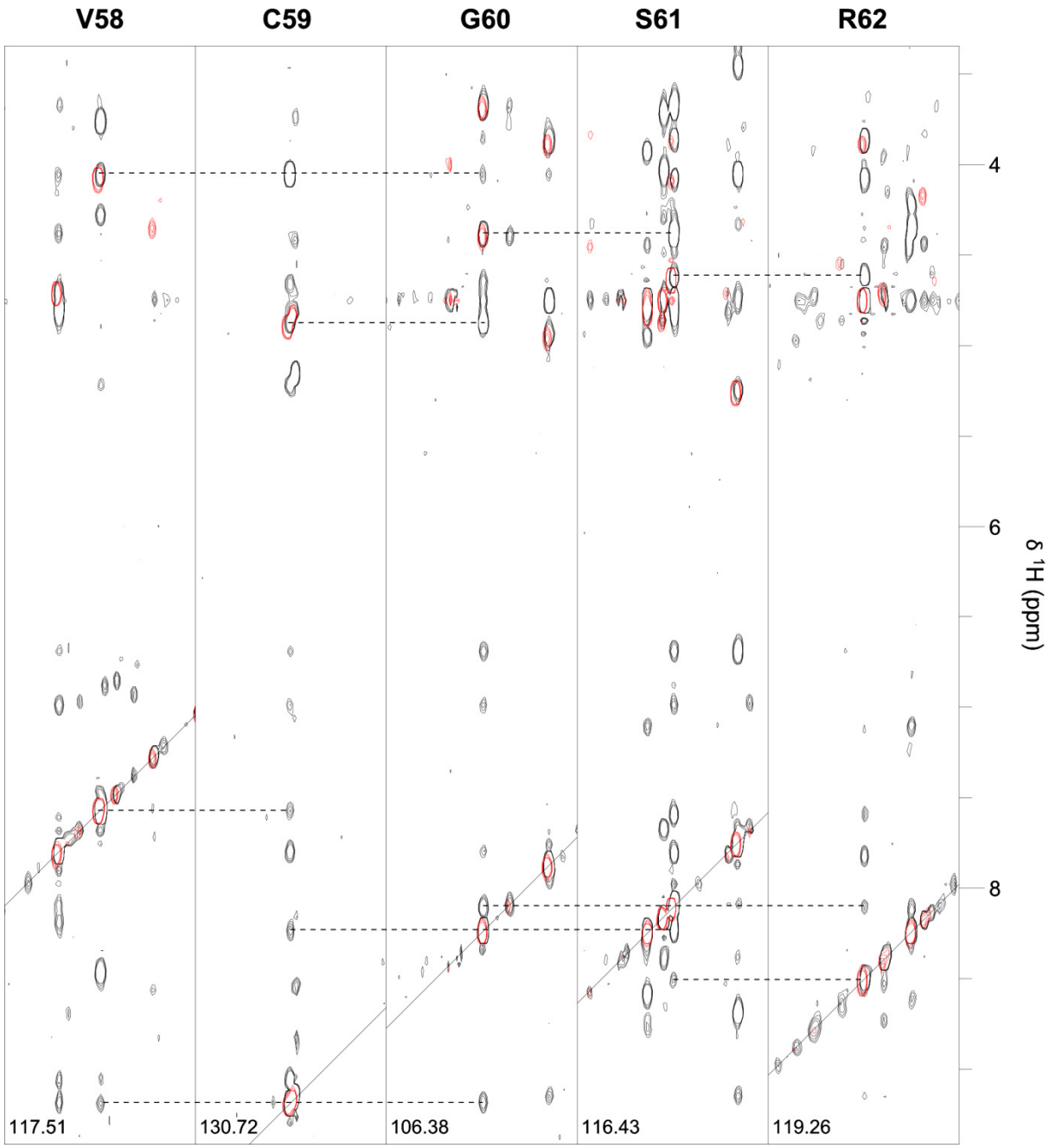


Figure 4.3 Sequential assignment of ^{15}N-labelled FUN-EGF3. Representative strips of 2D planes of 3D spectra for residues V58-R62 (FUN domain). Each strip is centred on the $^{1}H^{N}$ chemical shift of one residue. The ^{15}N chemical shift for the residue is given in the bottom-left hand corner of the strip. Contours for 3D ^{15}N-edited NOESY-HSQC (mixing time 150 ms) and TOCSY-HSQC spectra (mixing time 46 ms) are coloured black and red respectively. Peak coincidence indicated that the cross-peak in the NOESY arose from an interaction between $^{1}H^{N}$ of residue *i* with a ^{1}H nucleus in its own side chain. Dashed lines show sequential links between residues. For clarity, only connections between the $^{1}H^{N}$-$^{1}H^{N}$ (7-9 ppm) and $^{1}H^{\alpha}$-$^{1}H^{N}$ (4-5 ppm) cross-peaks are shown. The plot was generated in CCPN Analysis and annotated in Inkscape (http://inkscape.org/).

Backbone assignments were confirmed and extended with triple resonance spectra of double-labelled FUN-EGF3 (Section 2.6.2.5). The HNCA experiment (Kay *et al.*, 1990) correlates the $^{1}H^{N}$ and ^{15}N resonances of residue *i* with $^{13}C^{\alpha}$ resonances of residues *i* and *i-1* (Figure 4.2) Cross-peaks showed sequential connections between residues, confirming results of the 3D NOESY-HSQC and TOCSY-HSQC analysis (Figure 4.4A). HNCO (Kay *et al.*, 1990) and HN(CA)CO experiments (Clubb *et al.*, 1992) (Figure 4.2) provided additional evidence of sequential links and added to chemical shift assignments (Figure 4.4B). Both spectra had cross-peaks between $^{1}H^{N}$ and ^{15}N resonances of residue *i* and the $^{13}C'$ resonance of residue *i-1*, but the HN(CA)CO spectrum had an additional cross-peak to the $^{13}C'$ resonance of residue *i*.

Side-chain assignments are important for validating and extending backbone assignments, as well as for structure determination. ^{1}H side-chain assignments were initially obtained from NOESY-HSQC and TOCSY-HSQC spectra. A number of triple resonance experiments can provide ^{13}C side-chain assignments. The 3D (H)CC(CO)NH experiment (Grzesiek *et al.*, 1993) was particularly useful for FUN-EGF3. It correlates aliphatic ^{13}C side-chain resonances of residue *i-1* with the $^{1}H^{N}$ and ^{15}N resonances of residue *i* (Figure 4.2). Cross-peaks in the spectrum provided partial side-chain assignments (Figure 4.5A). Initial side-chain assignments were confirmed and extended using the 3D HCCH-TOCSY experiment (Bax *et al.*, 1990) (Section 2.6.2.4; Figure 4.2). Correlation of ^{1}H and ^{13}C chemical shifts allowed definitive assignment of side-chain resonances where the type of ^{1}H chemical shift was ambiguous. Furthermore, cross-peaks demonstrated connections between nuclei in the same residue (Figure 4.5B). Side-chain assignments of aromatic residues were also assisted by a 2D ^{1}H-^{13}C HSQC spectrum (Santoro and King, 1992; Vuister and Bax, 1992) (Section 2.6.2.3) centred on the aromatic ^{13}C resonances, as well as 2D NOESY and TOCSY spectra (Section 2.6.2.2; data not shown).

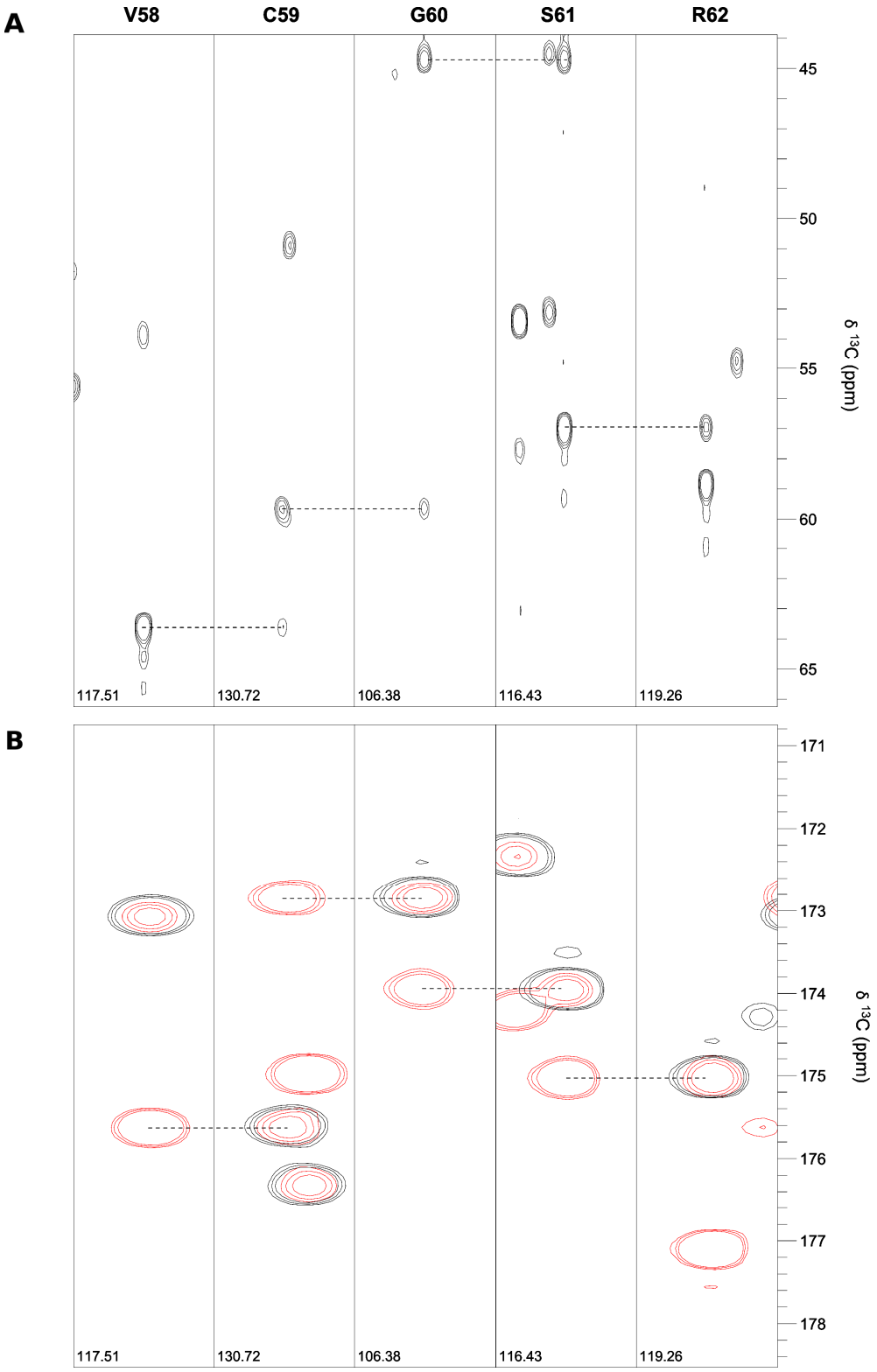


Figure 4.4 Assignment of the FUN-EGF3 backbone using triple resonance methods. Example strip plots, in the same format as Figure 4.3. A) Strips of the 3D HNCA spectrum showing sequential links between $^{1}H^{N}$-^{15}N-$^{13}C^{\alpha}$ cross-peaks. B) Strips of the 3D HNCO (black contours) and HN(CA)CO (red contours) spectra. Sequential links between $^{1}H^{N}$-^{15}N-$^{13}C'$ cross-peaks are indicated.

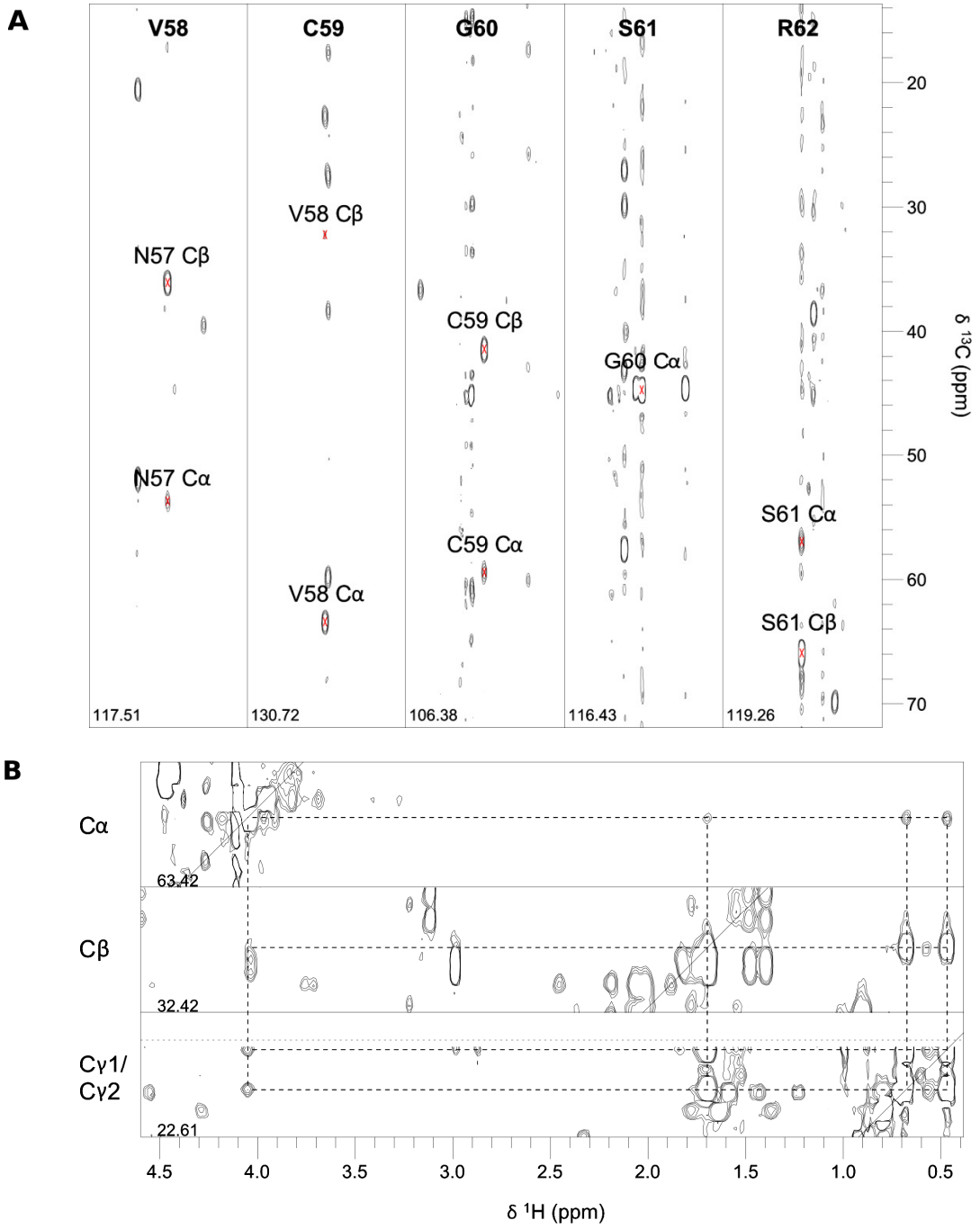


Figure 4.5 Side-chain assignment of FUN-EGF3. A) Strip plot of 3D (H)CC(CO)NH spectrum in the same format as Figure 4.3. Labelled cross-peaks are marked with red crosses. B) Strips of 2D ^{13}C planes of a 3D HCCH-TOCSY spectrum for resonances of residue V58. Each strip is centred on the ^{1}H chemical shift of the highlighted nucleus; the ^{13}C chemical shift of the nucleus bonded to this nucleus is given in the bottom left-hand corner of the strip. Dashed lines indicate connections between diagonal and cross-peaks, showing that they arise from resonances belonging to the same spin-system. The chemical shifts of $^{13}C^{\gamma1}$ and $^{13}C^{\gamma2}$ are very similar, so cross-peaks to $^{1}H^{\gamma1}$ and $^{1}H^{\gamma2}$ are shown in the same plane (bottom strip). Assignments are not stereospecific, so labels are arbitrary.

Following an iterative process of assignment using multiple datasets, near-complete backbone and side-chain assignments were obtained. Backbone $^{1}H^{N}$ and ^{15}N assignments are shown as labelled cross-peaks in the ^{1}H-^{15}N HSQC spectrum (Figure 4.6). Table 4.2 shows statistics for the completeness of assignment. Assignments were obtained for every residue, but some expected resonances were absent from spectra, presumably resulting from broad lines or weak intensities of cross-peaks. The majority of missing assignments were from aromatic side chains. Backbone ^{15}N nuclei of proline residues were also not assigned. The extent of resonance assignment was sufficient for assignment of cross-peaks in 3D NOESY-HSQC spectra and prediction of torsion angles from chemical shifts. A complete list of assignments is given in Appendix E.

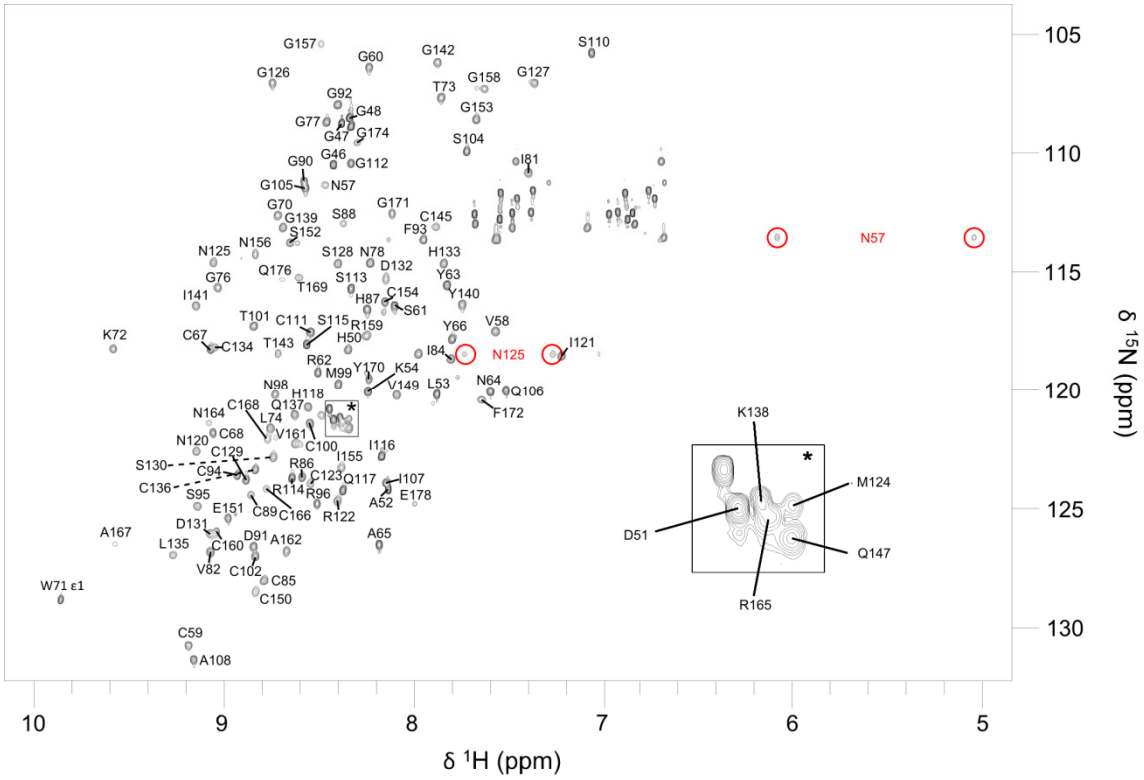


Figure 4.6 ^{1}H-^{15}N HSQC of FUN-EGF3 with assignments. This is the same spectrum as in Figure 3.9B. Cross-peaks arising from backbone amides are labelled with the residue number. The W71 side-chain $^{1}H^{\varepsilon 1}$-$^{15}N^{\varepsilon 1}$ cross-peak is also marked. For clarity, assignments are not shown for overlapping peaks or for most of those arising from NH_2 side-chain groups. The N57 and N125 NH_2 side-chain peaks are highlighted in red as they have unusual chemical shifts. The boxed region marked '*' is expanded in the bottom right-hand corner because of crowding in the centre of the spectrum.

Table 4.2 Completeness of chemical shift assignment for FUN-EGF3. Statistics are given for the 134 native residues in the FUN-EGF3 construct (R45-E178). The table was compiled in CCPN Analysis.

Category	Available	Assigned	% Assigned
All Residues	134	134	100.0
Amide	258	242	93.8
Backbone	526	510	97.0
Backbone non-^{1}H	402	388	96.5
Side Chain ^{1}H	586	569	97.1
Side Chain non-^{1}H	338	245	72.5
Type ^{13}C'	134	134	100.0
Type $^{13}C^{\alpha}$	134	134	100.0
Type $^{13}C^{\beta}$	111	110	99.1
Type $^{13}C^{\gamma}$	84	60	71.4
Type $^{13}C^{\delta}$	56	45	80.4
Type $^{13}C^{\varepsilon}$	18	11	61.1
Type $^{13}C^{\zeta}$	16	3	18.8
Type $^{13}C^{\eta}$	1	1	100.0
Type $^{1}H^{N}$	124	122	98.4
Type $^{1}H^{\alpha}$	157	157	100.0
Type $^{1}H^{\beta}$	202	200	99.0
Type $^{1}H^{\gamma}$	97	97	100.0
Type $^{1}H^{\delta}$	84	82	97.6
Type $^{1}H^{\varepsilon}$	41	28	68.3
Type $^{1}H^{\zeta}$	4	4	100.0
Type $^{1}H^{\eta}$	1	1	100.0
Type ^{15}N	134	120	89.6
Type $^{15}N^{\delta}$	13	8	61.5
Type $^{15}N^{\varepsilon}$	20	7	35.0
Type $^{15}N^{\zeta}$	3	0	0.0
Type $^{15}N^{\eta}$	16	0	0.0

4.4 Preliminary analysis of backbone dynamics

Assignments allowed additional datasets to be interpreted in terms of primary sequence. The ^{15}N-{^{1}H} heteronuclear NOE experiment was used to provide insights into the fast-timescale dynamics (ps-ns) of the backbone. It measures the effect of the cross-relaxation between $^{1}H^{N}$ and ^{15}N on the magnetisation of the ^{15}N nucleus, which is sensitive to the motion of the H-N bond vector (Kay *et al.*, 1989b). The experiment therefore requires initial direct polarisation of ^{15}N, different from

many protein NMR experiments that polarise ^{1}H first. Two spectra were acquired in an interleaved manner: one with ^{1}H pre-saturation and a second without pre-saturation (Section 2.6.4). The spectrum acquired without pre-saturation acted a 'blank' reference. The heteronuclear NOE ratio for each peak was calculated as follows:

$$Heteronuclear\ NOE = \frac{I_{sat}}{I_{unsat}} \qquad 4.1$$

where I_{sat} and I_{unsat} are the peak intensities for the spectra acquired with and without pre-saturation respectively. A lower heteronuclear NOE ratio means that the H-N bond vector for a given residue is more dynamic. Plotting the heteronuclear NOE ratio against residue number gave insights into the profile of dynamics across FUN-EGF3 (Figure 4.7).

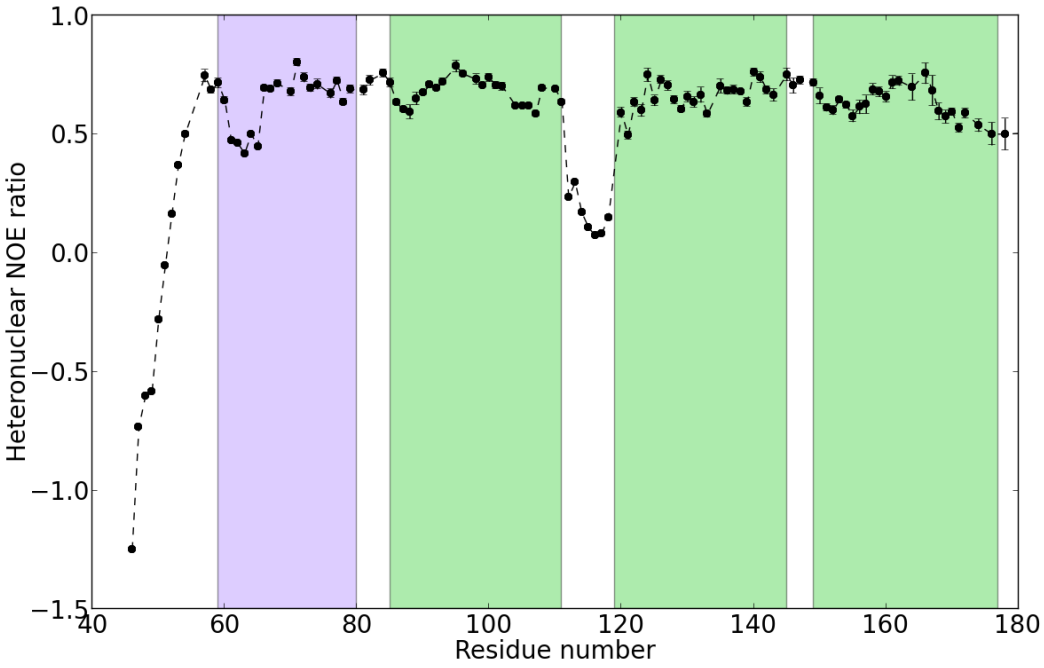


Figure 4.7 Results of ^{15}N-{^{1}H} heteronuclear NOE experiment for FUN-EGF3. The NOE ratio is plotted against residue number. The points are connected by dashed lines. Error bars show the standard deviation of the NOE ratio, calculated using an estimate of the noise in the spectrum (Section 2.6.4). Domains are indicated by coloured boxes, starting and ending at the first and last cysteine residues respectively. The FUN domains and EGF-like domains are coloured lilac and green respectively.

There was significant variation in the heteronuclear NOE ratio across the FUN-EGF3 fragment. Several flexible regions were identified. Residues at the N-terminus (R45-K54) all had heteronuclear NOE ratios less than 0.60. There was also a dip in values in the FUN domain – residues S61-A65 had values less than 0.50. This region may correspond to a flexible loop. There were significant differences in the results for residues at the three interdomain interfaces. While the four residues in each of FUN-EGF1 and EGF2-EGF3 interdomain linker sequences appeared to be rigid, the longer linker between the EGF1 and EGF2 domains was flexible. All 7 residues between the last cysteine of EGF1 (C111) and the first cysteine of EGF2 (C119) had NOE ratios less than 0.30. This hinted that the FUN-EGF1 and EGF2-EGF3 regions may be independent structural units. Finally, residues in the C-terminus of EGF3, after the fifth cysteine residue (C168) also had lower NOE ratios (less than 0.60). In the full-length fibrillin molecule, this region is likely to pack with the hyb1 domain and so its flexibility in the FUN-EGF3 fragment was unsurprising. The results of the heteronuclear NOE experiment had implications for structure determination, particularly for the implementation of residual dipolar coupling (RDC) restraints during refinement (Section 4.8.5).

4.5 NOE assignment and restraint generation

The major restraints in NMR structure calculations are inter-^{1}H distances derived from the intensities of cross-peaks in 2D and 3D NOESY spectra. Therefore, correct assignment of cross-peaks is essential for accurate structure determination. Assigning these peaks also helps to confirm backbone and side-chain resonance assignments. Examples of NOE assignments in 3D spectra of FUN-EGF3 are given in Figure 4.8. In some cases, reciprocal cross-peaks between resonances confirmed assignments. If assignments were unclear, peaks were left unassigned until completion of initial rounds of structure calculations (Section 4.7).

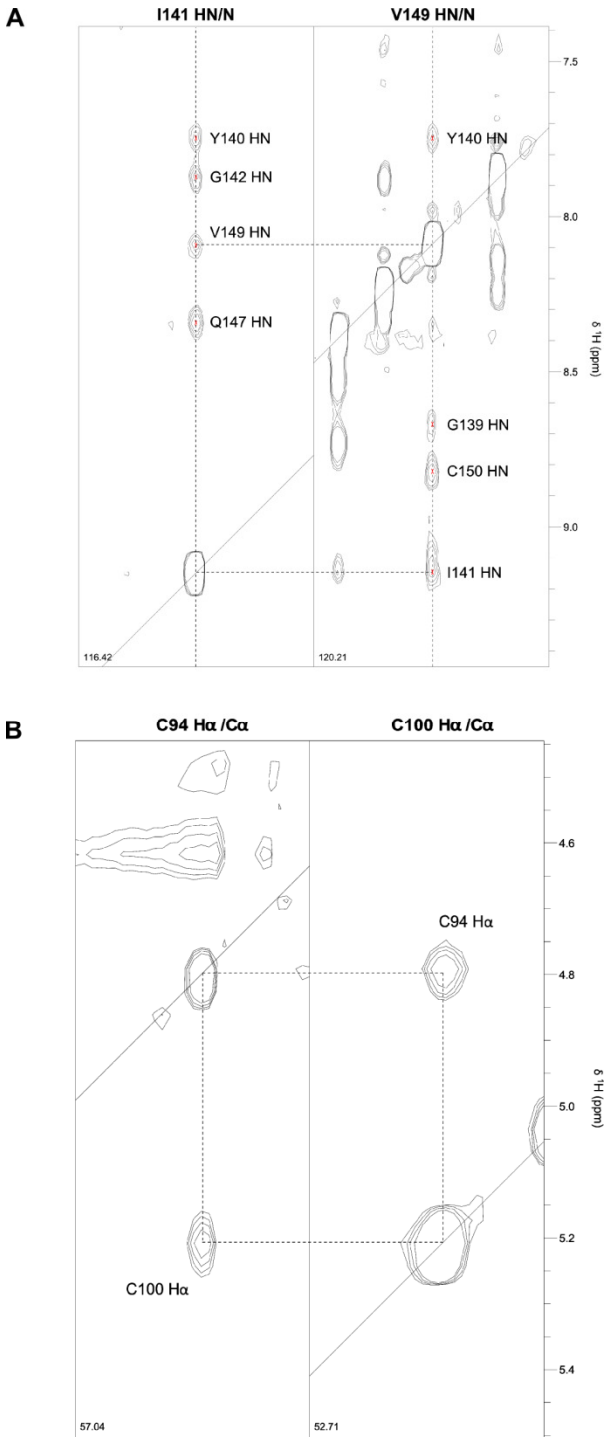


Figure 4.8 Cross-peak assignment of FUN-EGF3 NOESY spectra. A) Strip plot of the 3D ^{15}N-edited NOESY-HSQC spectrum (mixing time 150 ms) for residues I141 and V149 at the EGF2-EGF3 interdomain interface, showing sequential and long-range NOEs between $^{1}H^{N}$ nuclei. Connections between residues, identified by labelled cross-peaks, are indicated by dashed lines. B) Strip plot of the 3D ^{13}C-edited NOESY-HSQC spectrum (mixing time 150 ms) for C94 and C100, two residues in the predicted β-hairpin of the EGF1 domain. Connections, resulting from NOEs between $^{1}H^{\alpha}$ nuclei are indicated as in part A.

NOE assignments alone, without calculated structures, provided initial qualitative insights into the structure of the FUN-EGF3 fragment. Strikingly, no NOEs were found between residues in the FUN-EGF1 and EGF2-EGF3 regions (data not shown). However, NOEs were identified at the interfaces between FUN and EGF1, and between EGF2 and EGF3 (example in Figure 4.8A). This was consistent with the results of the heteronuclear NOE experiment, which suggested that the two halves of the molecule might behave as independent structural units (Section 4.4). Further evidence is required, but the lack of an observed interface between EGF1 and EGF2 strongly supported this assertion.

Cross-peak intensities in NOESY spectra were converted into distance restraints for use in structure calculations (Section 2.6.5). In theory, the intensity of a cross-peak should be proportional to $1/r^6$, where r is the internuclear distance (Brünger *et al.*, 1986). Variation in peak intensities resulting from variable dynamics means that it is generally not possible to calculate distances with high precision. A common approach is to use distance binning, whereby cross-peaks are grouped into broad categories according to their intensities, using thresholds for each bin. Here, the initial upper distance limit bins were 2.8, 3.5, 5.0, 6.0 and 8.0 Å. These values were used as the upper bounds in the distance restraints, while the target distance and lower distance bounds were both set to 0.

Intensity thresholds were defined separately for each NOESY spectrum. NOESY spectra with mixing times of 75 ms were used for restraint generation. Comparison of peak intensities between spectra with 75 and 150 ms mixing times indicated a contribution of spin diffusion in the spectrum with the longer mixing time (data not shown). The thresholds for the ^{15}N-edited NOESY-HSQC spectrum were estimated initially using the intensities of cross-peaks between nuclei in the side chain of W71 as references (not shown), as distances within the aromatic ring should not be variable. For the ^{13}C-edited NOESY-HSQC, estimates were obtained using intensities for $^{1}H^{\alpha}$-$^{1}H^{\alpha}$

cross-peaks between residues on opposite sides of the predicted β-hairpin in the EGF1 domain (not shown).

4.6 Backbone torsion angles

Torsion angle restraints are commonly employed in structure determination. Angles can be determined experimentally or predicted using chemical shifts on the basis of information in structure databases. The three torsion angles that define the polypeptide backbone are φ, ψ and ω (Figure 4.9A).

4.6.1 Measurement of $^3J_{HNH\alpha}$ couplings

φ angle restraints can be obtained from the 3-bond J coupling constants between H^N and H^α in the same residue ($^3J_{HNH\alpha}$). The value of this coupling constant depends on the φ torsion angle, as defined by the Karplus equation (Figure 4.9B). $^3J_{HNH\alpha}$ values were measured using the HMQC-J experiment with ^{15}N-labelled FUN-EGF3 (Kay *et al.*, 1989a) (Section 2.6.6). The 2D HMQC-J spectrum contains a cross-peak for every ^{15}N-^{1}H group in the protein, as in the 2D HSQC spectrum. However, cross-peaks are split in the ^{15}N dimension − all peaks are doublets, except those arising from glycine residues, which are doublets of doublets.

The magnitude of the splitting was measured by fitting two peaks to the doublet (Figure 4.9C). The difference between the ^{15}N chemical shifts at the peak maxima was used to calculate the $^3J_{HNH\alpha}$ value. Analysis was only carried out for peaks assigned to non-glycine residues with discernible splitting. For a given residue, a value of > 8 Hz indicates that the φ torsion angle lies in the β-sheet region of the Ramachandran plot (−120 ± 40°). A value of < 6 Hz is ambiguous, as the φ torsion angle could be positive or negative (Figure 4.9B). Values of 6-8 Hz were discarded, as

they usually result from motional averaging in dynamic regions. Therefore, torsion angle restraints were only included initially for residues with $^{3}J_{HNH\alpha} > 8$ Hz.

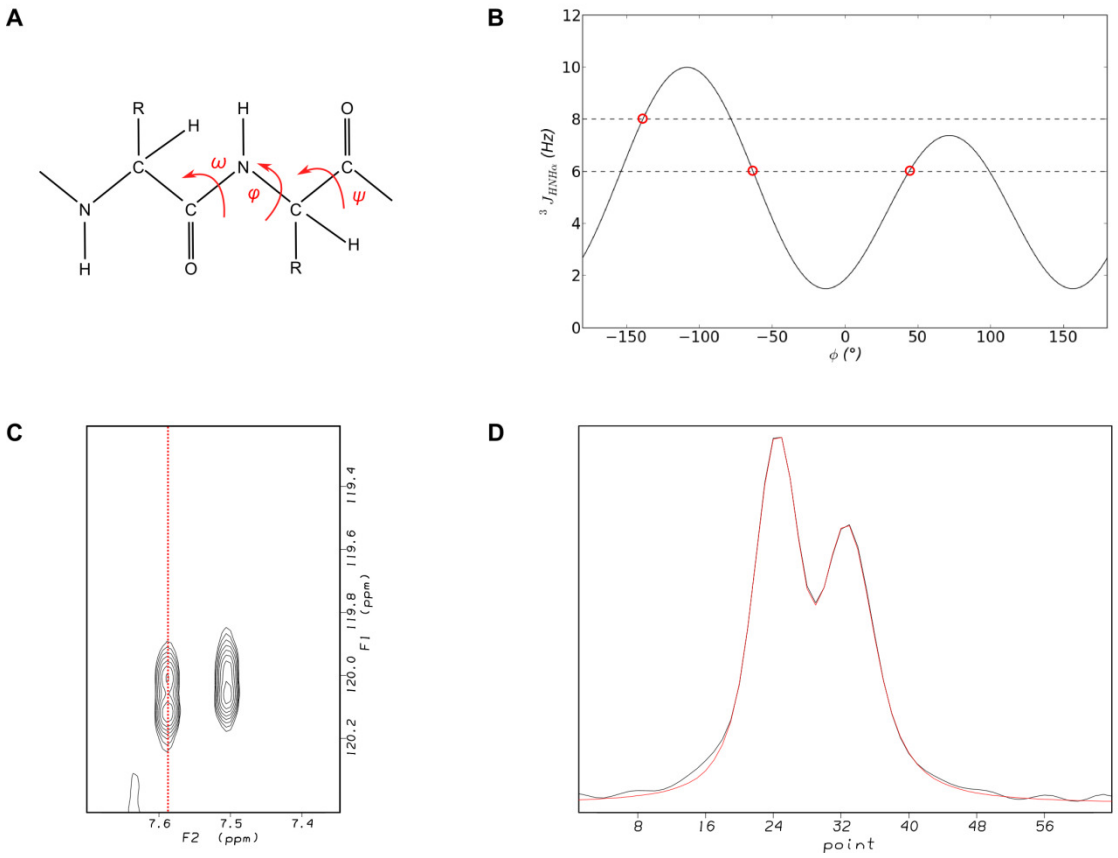


Figure 4.9 Backbone torsion angles in FUN-EGF3. A) Illustration of the *φ, ψ* and *ω* torsion angles in the polypeptide backbone. B) Karplus plot of the dependence of the $^{3}J_{HNH\alpha}$ coupling constant on the *φ* angle. The Karplus equation is an empirically derived formula: $J = A\cos^2\vartheta + B\cos\vartheta + C$. Here, A = 7.13, B = −1.31, C = 1.56 and $\vartheta = \varphi$, with an offset of 60° (Habeck *et al.*, 2005). Horizontal dashed lines are also plotted at 6 and 8 Hz; red circles indicate their points of intersection with the Karplus plot at *φ* angles in allowed regions of the Ramachandran plot. The graph illustrates the ambiguity of the *φ* angle for $J_{HNH\alpha} < 6$ Hz, as positive and negative *φ* angles can give rise to the same value. C) Fitting of 2D ^{1}H-^{15}N HMQC-J data to obtain $^{3}J_{HNH\alpha}$ values. An example for one peak is shown in one region of the spectrum. D) 1D slice through the ^{15}N (F1) dimension (marked by red dotted line in C). The actual and fitted curves are shown in red and black respectively. Fitting was performed using the 'Oxford fitting auto-HMQC' procedure in Felix 2.3 (Accelrys).

4.6.2 TALOS+ prediction

The TALOS+ program predicts *φ and ψ* angles using chemical shifts on the basis of values from proteins with known structures (Shen *et al.*, 2009). The input for the program is a set of chemical shifts ($^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, $^{1}H^{N}$, ^{15}N and $^{1}H^{\alpha}$) for a residue triplet, with one residue either side of the residue of interest. TALOS+ identifies comparable triplets with similar secondary chemical shift values in a database and uses their torsion angles to make a prediction. A prediction is only considered 'good' if the values from the database are consistent with one another. TALOS+ predictions were made for FUN-EGF3 (Section 2.6.7; Figure 4.10). Predictions were only classified as good for 55% (74/134) residues. Furthermore, N-terminal residues (up to G49) and EGF1-EGF2 linker residues were classified as 'dynamic', consistent with their low heteronuclear NOE values (Section 4.4). The torsion angle predictions were used as the target values in restraints for structure calculations and the errors, multiplied by a factor of two, were used as the upper and lower bounds; minimum upper bounds were set to 30°. Restraints from the $^{3}J_{HNH\alpha}$ (Section 4.6.1) replaced TALOS+ restraints where they were available. TALOS+ predictions were also used to resolve the ambiguity of *φ* angle restraints for $^{3}J_{HNH\alpha} < 6$ Hz.

4.6.3 *Cis vs. trans* proline residues

The partial double bond character of the peptide bond means that the *ω* torsion angle can only be one of two possible values, 0 (*cis*) or 180° (*trans*). The *cis* configuration is rare, only occurring at some peptide bonds N-terminal to proline residues (Wedemeyer *et al.*, 2002). As the *ω* angle is fixed in the structure calculations, it is important to establish first whether proline residues in the protein adopt the *cis* or *trans* configuration. Two methods were used to check this for FUN-EGF3. Firstly, cross-peaks between the proline $^{1}H^{\delta}$ side-chain resonances and the $^{1}H^{\alpha}$ resonance of the *i-1* residue showed that all proline residues adopted the *trans* configuration (Figure 4.11). This

result was verified using the program Promega, which predicts the probability of each proline being in the *cis* configuration based on chemical shifts (Shen and Bax, 2010) (data not shown).

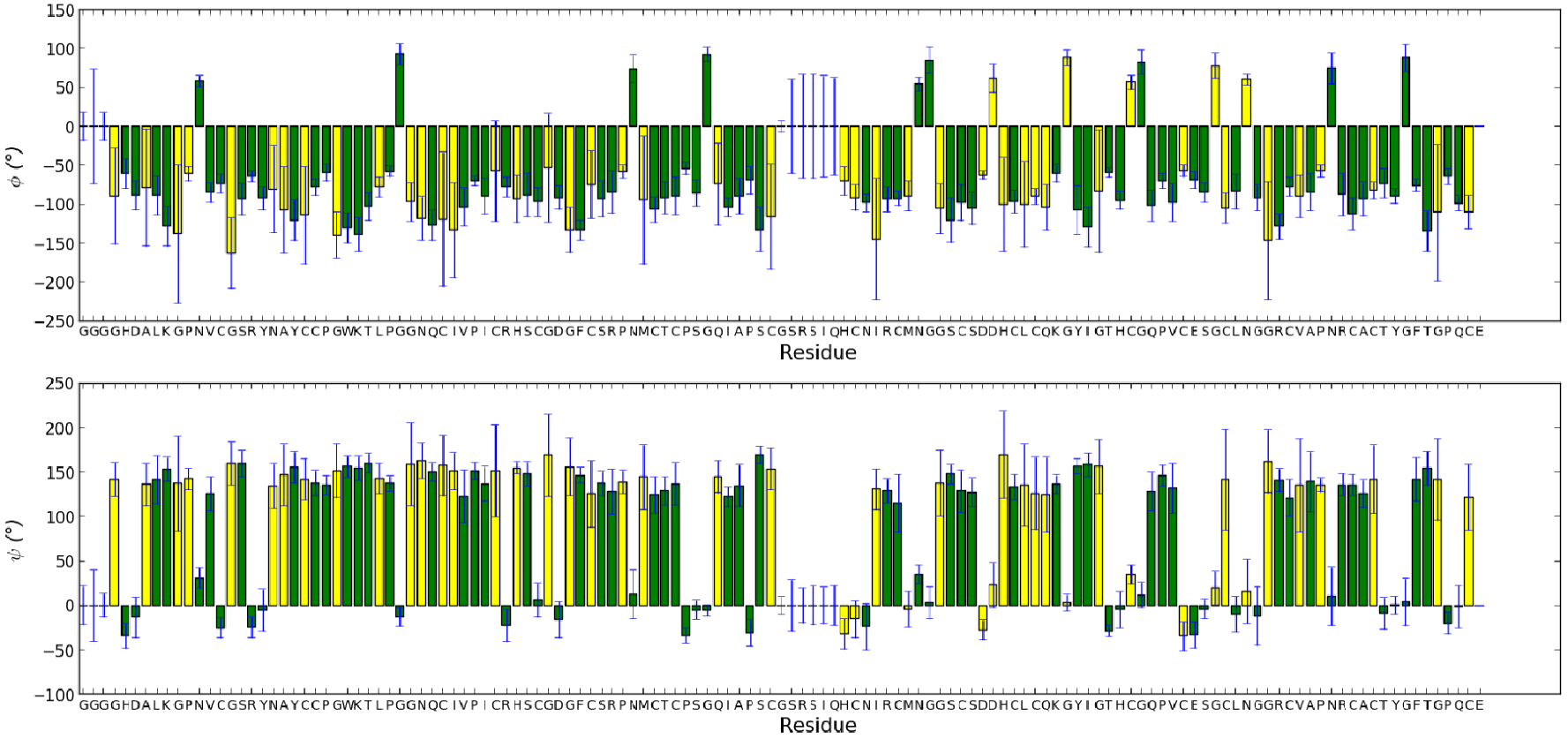


Figure 4.10. TALOS+ predictions for FUN-EGF3. Chemical shift values were used to predict *φ and ψ* angles for the FUN-EGF3 fragment. Predicted angles are plotted as bars for each residue in sequence order. Top: *φ* angle; bottom: *ψ* angles. Green and yellow bars show predictions marked as 'good' and 'bad' by TALOS+ respectively. Blue error bars show the prediction error.

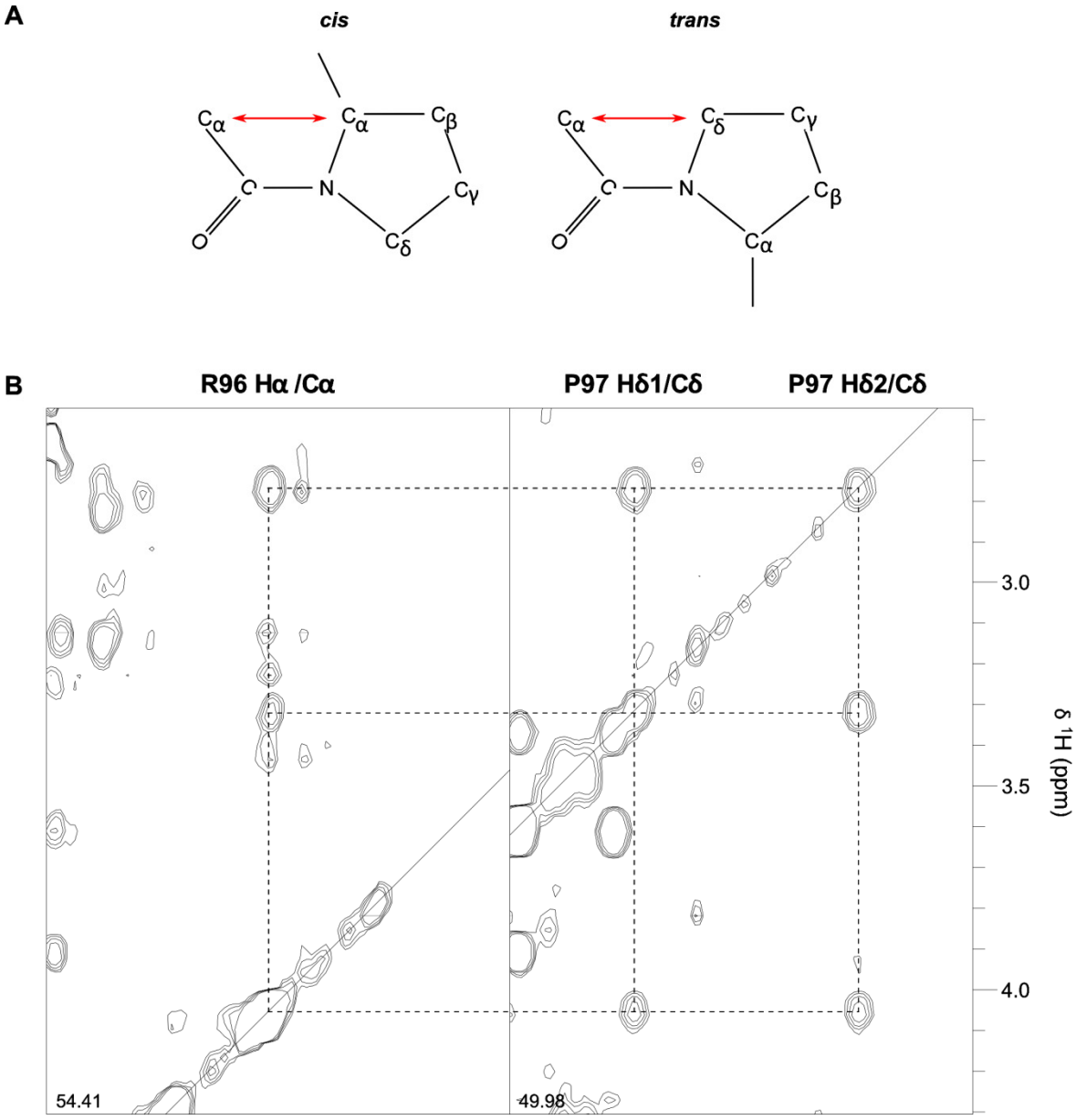


Figure 4.11 Distinguishing between *cis* and *trans* prolines. A) Illustration of *cis* and *trans* peptide bonds preceding a proline residue. The C^{α} nucleus of the residue before the proline is closest to C^{α} of the proline in the *cis* configuration, but closest to C^{δ} of the proline in the *trans* case (indicated by red arrows). B) Strip plot of 3D ^{13}C NOESY-HSQC (spectrum also in Figure 4.8B) showing connections between $^{13}C^{\alpha}$ of R96 and $^{13}C^{\delta}$ of P97 in the EGF1 domain. $^{1}H^{\delta1}$ and $^{1}H^{\delta2}$ labels shown here are arbitrary, not stereospecific. Intense cross-peaks indicated that the peptide bond between R96 and P97 was in the *trans* configuration.

4.7 Initial structure calculation by simulated annealing

NOE-derived distance and torsion angle restraints were used to generate structures of FUN-EGF3 by simulated annealing. Simulated annealing is a computational method for finding the maximum or minimum value of a function with many independent variables (Kirkpatrick *et al.*, 1983). It is therefore well suited to NMR structure calculation by energy minimisation. In a structure calculation, the total energy of a target function, E_{total}, is minimised:

$$E_{total} = E_{bond} + E_{angle} + E_{improper} + E_{vdW} + E_{NOE} + E_{dihedral} + \ldots + E_n \qquad 4.2$$

where E_{bond}, E_{angle}, $E_{improper}$, E_{VdW}, E_{NOE} and $E_{dihedral}$ are the bond, angle, improper, van der Waals, NOE and dihedral energy terms respectively (Nilges *et al.*, 1988). E_n represents additional energy terms representing other restraints that could be added. The bond, angle and improper terms maintain correct covalent geometry, while the van der Waals term ensures that the distance between a pair of non-bonded atoms cannot be less than the sum of their van der Waals radii. Deviation from a target value increases the energy of one term and therefore also the total energy of the system. The contribution of each term to E_{total} is determined by its force constant.

In initial structure calculations starting with an extended template, the simulated annealing process begins with the system at high temperature. Atoms move around according to their energies, following Newton's laws of motion. At a given temperature, atoms are allowed to move until the energy reaches a minimum, the temperature is lowered and the process is repeated. Force constants are initially set at low values to allow atoms to move freely, but are increased as the system cools, in order to restrict motion. When the system has been cooled to the lowest temperature of the simulation, it should have reached its energy minimum. Multiple calculations are run with different, randomised starting velocities, as the system may become stuck in local

energy minima. An ensemble of the lowest energy structures should represent the best agreement of the resulting models with the input restraints.

Here, initial structures were calculated by simulated annealing from an extended template in Xplor-NIH (Section 2.7.1) (Schwieters *et al.*, 2003). This well-established software package was chosen because it has significant functionality and annealing scripts are readily customisable. Distance restraints based on manual NOE assignments were used in the calculations. Disulphide bonds were incorporated initially as ambiguous distance restraints between the S^{γ} atoms of cysteine residues. Although assigning NOESY cross-peaks manually can introduce human bias, automated assignment programs do not necessarily perform better. Trial calculations were run with automated, iterative NOE assignment in ARIA-CNS (Rieping *et al.*, 2007). The structures produced were very similar to those generated by Xplor-NIH, but arbitrary correction factors for the peak intensity-distance calibration were needed to prevent exclusion of valid distance restraints and some cross-peaks were assigned incorrectly (not shown). This approach was therefore not taken further.

4.7.1 Iterative NOE assignment

Ensembles of 200 structures were generated and analysed using Xplor-NIH. Violations of distance restraints greater than 0.5 Å occurring in more than 50 of the 100 lowest-energy structures were identified. Incorrect cross-peak assignments were modified on the basis of inter-^{1}H distances in calculated structures. Peak assignments that were previously ambiguous were refined on the basis of the initial structures. Moreover, intensity thresholds for upper distance limit bins were modified to reduce the number of violations. Structures were also analysed to compare restraints to distances in an ensemble of structures. The final set of distance bins was 2.8, 3.5, 4.0, 5.0, 6.0 and 6.5 Å.

4.7.2 Stereospecific assignment using prochiral swapping

Chemical shift assignments for FUN-EGF3 were not stereospecific. Chemically identical groups, such as two ^{1}H nuclei in a CH_2 group, can give rise to resonances with different chemical shifts in NMR, if they are in different magnetic environments. These groups, in so-called prochiral pairs, cannot be readily distinguished using standard assignment experiments. In the absence of this information, separate distance restraints cannot be employed for prochiral pairs. Instead, a single ambiguous restraint is used, which is satisfied if the internuclear distances for both groups is shorter than the upper distance bound, or two restraints with the same upper distance bound are used (Figure 4.12A-B). In either case, information from the NOESY spectra is lost and the smaller number of restraints can potentially reduce the quality of the resulting structures. Some stereospecific assignments can be obtained by experimental methods, but a common approach is to use floating prochirality in structure calculations (Folmer *et al.*, 1997). Prochiral pairs are swapped during simulated annealing and the swapped configuration is retained if it results in a lower overall energy. Improper violations resulting from swaps do not contribute to the energy, but indicate the presence of successful swaps.

The prochiral swapping approach was used to obtain stereospecific assignments for FUN-EGF3 (Section 2.7.2). Initial calculations were carried out first using Xplor-NIH with ambiguous restraints to eliminate NOE violations. For swapping calculations, Xplor 3.8 was used instead of Xplor-NIH, as the more recent version lacks the required functionality. The results of the swapping calculations were analysed in an automated fashion (Figure 4.12C). 50 prochiral pairs, out of a total of 197 in FUN-EGF3, were stereospecifically assigned. Distance restraints were modified accordingly and were used in subsequent calculations without floating chirality.

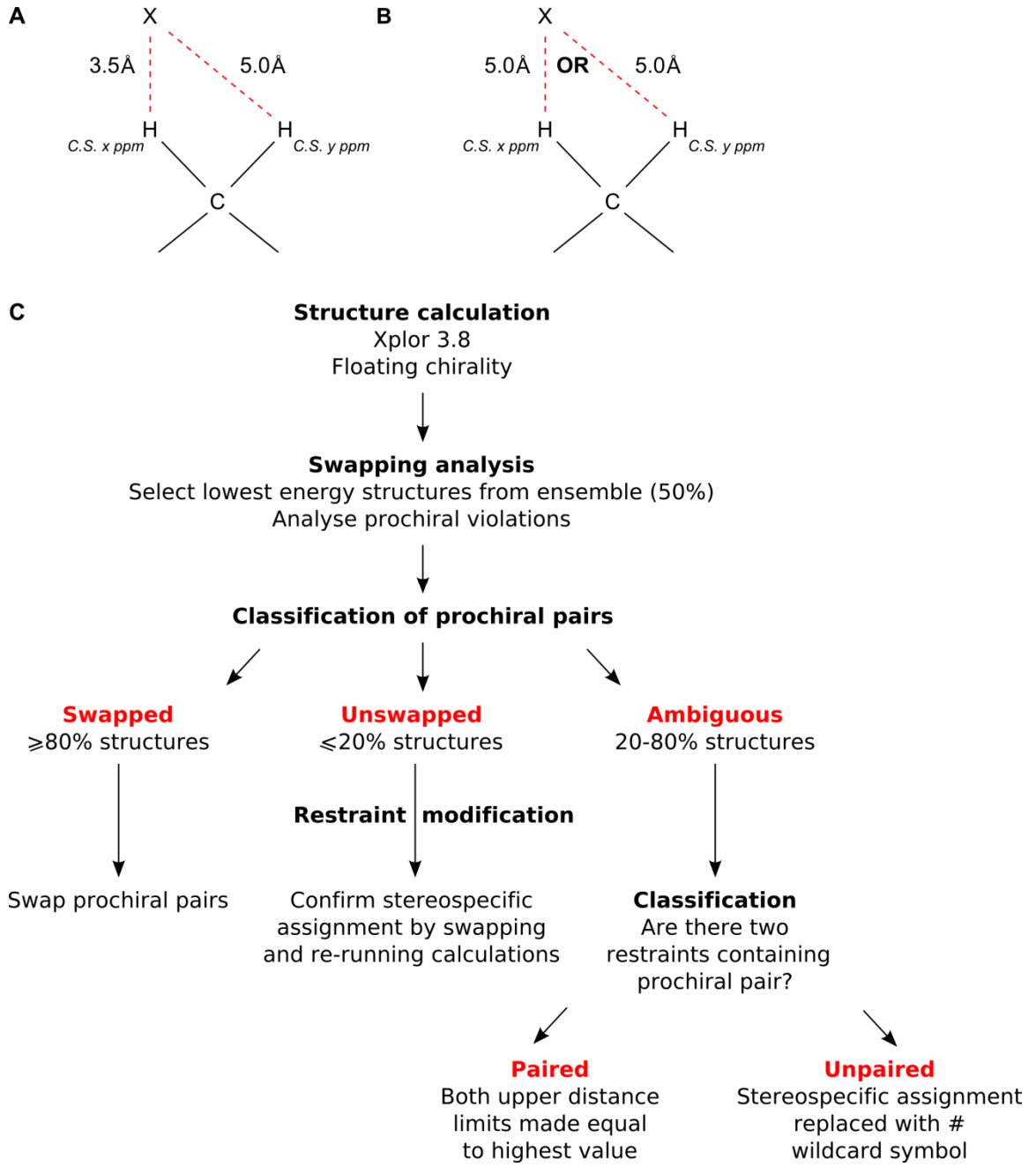


Figure 4.12 Stereospecific assignment using floating chirality. A) Illustration of the prochirality problem in NMR structure calculation. Two ^{1}H nuclei in a methylene group with different chemical shifts (C.S), *x* and *y* ppm, are separated by different distances from nucleus X. B) Without stereospecific assignments of the two ^{1}H nuclei, they cannot be distinguished in the distance restraints. They are therefore incorporated as an ambiguous restraint or two restraints with the same upper distance bound. C) Procedure for stereospecific assignment using floating chirality. The proportion of structures in which a prochiral pair was swapped was analysed and then used to classify restraints. Distance restraints were modified according to these classifications. Analysis was automated using a Python script (Section 2.7.2).

4.8 Refinement

Initial structures of FUN-EGF3 had no consistent distance (>0.3 Å) or dihedral angle (>5°) restraint violations. Moreover, the lack of other violations indicated that covalent geometry was not distorted by the experimentally derived restraints (not shown). Structures were then subjected to refinement. Refinement protocols operate on the same principles as the initial simulated annealing calculations, but only allow smaller changes in structure. The cooling process starts with the coordinates of a low-energy initial structure calculated by simulated annealing (Section 2.7.3). As the relative orientation of the FUN-EGF1 and EGF2-EGF3 regions was not well defined, the domain pairs were refined separately first. This was particularly important for the implementation of RDC restraints (Section 4.8.5). Stages in FUN-EGF3 refinement are illustrated in Figure 4.13.

4.8.1 Ramachandran torsion angle database potential

Structures produced by simulated annealing were first refined with a conformational database (Rama) potential. This energy term ensures that torsion angles are restricted to values that are known to occur from previous X-ray crystallography and NMR structural studies. It therefore improves the stereochemical quality of NMR structures (Kuszewski *et al.*, 1996). The potential was not used in initial structure calculations as it was found to introduce large angle and improper violations. However, it was successfully employed in refinement calculations.

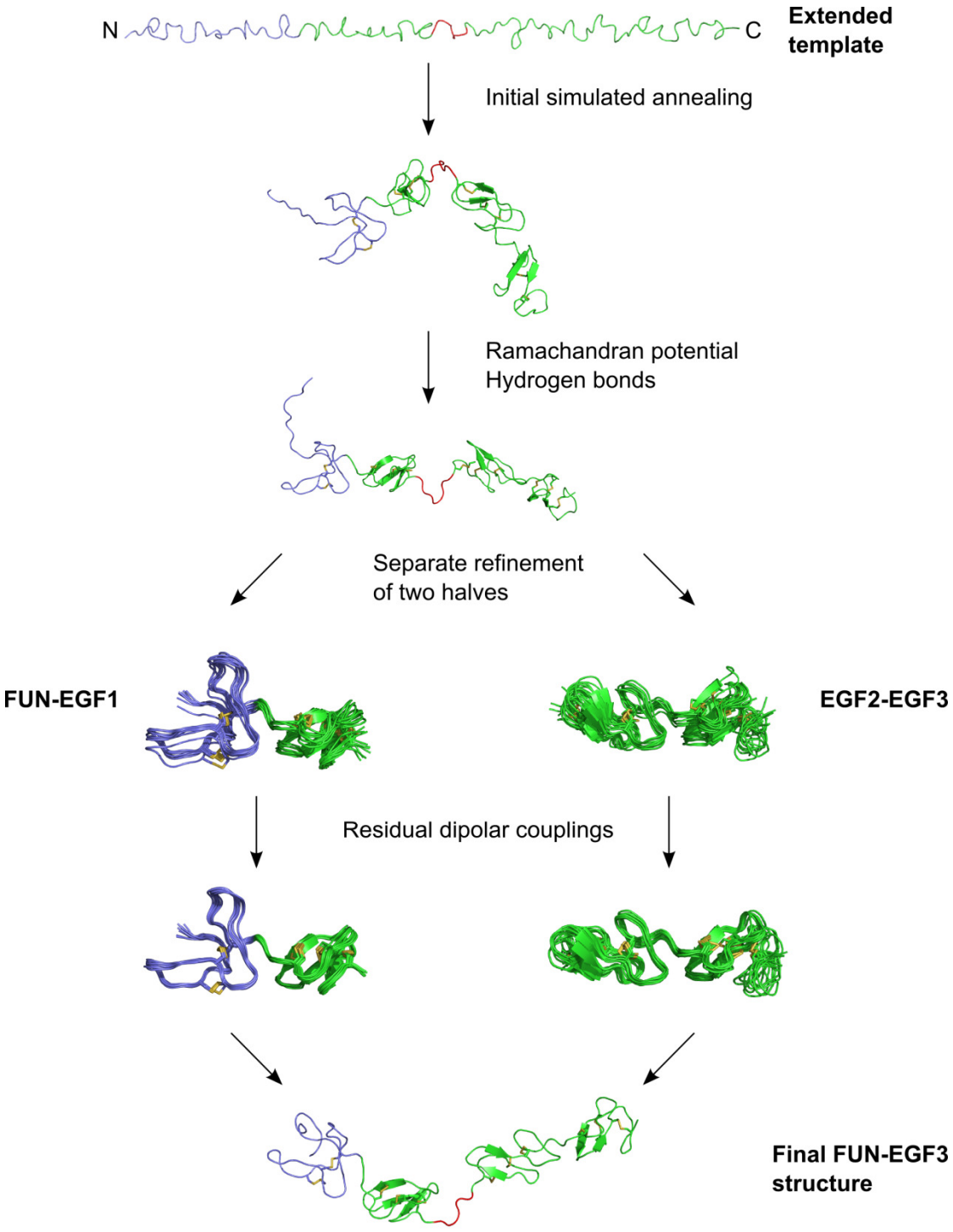


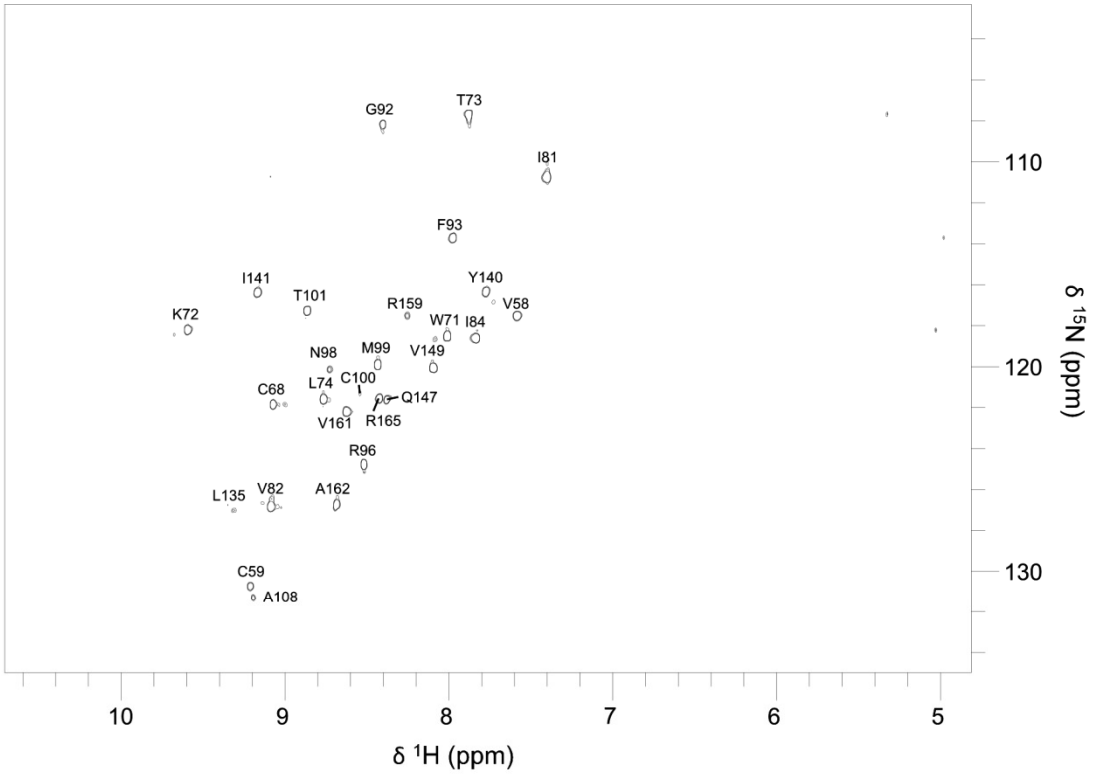
Figure 4.13 Refinement of FUN-EGF3 structures. Initial structures calculated by simulated annealing from an extended template were refined to produce the final FUN-EGF3 structure. The two halves of the structure, FUN-EGF1 and EGF2-EGF3, were first refined separately to optimise the procedure. This ensured that energy terms were not biased towards one half of the structure during optimisation. After determination of parameters for refinement using RDCs and hydrogen bonds, the restraints for the two halves were combined for final refinement of full-length FUN-EGF3. Single, low-energy structures are shown of full-length FUN-EGF3 for clarity.

4.8.2 Hydrogen bonds

Hydrogen bonds are important interactions that influence protein structure. They are also useful restraints in NMR structure calculations. The presence of hydrogen bonds can be detected indirectly using NMR. Slowly exchanging $^{1}H^{N}$ nuclei were identified in FUN-EGF3 by dissolving freeze-dried ^{15}N-labelled protein in 100% D_2O and acquiring a series of ^{1}H-^{15}N HSQC spectra (Section 2.6.8). 26 peaks were still visible in the 10th spectrum, acquired ~10 hrs after dissolving FUN-EGF3 in 100% D_2O (Figure 4.14A). The majority of peaks disappeared due to exchange of $^{1}H^{N}$ for $^{2}H^{N}$ in the solvent. Slow exchange indicates that the N-H group is a hydrogen bond donor or is buried in the hydrophobic core of the protein. Since EGF-like domains do not have a discernible hydrophobic core, the latter was unlikely in the FUN-EGF3 fragment.

These results were not sufficient to generate hydrogen bond restraints for use in structure calculations, as the acceptor groups were unknown. It is theoretically possible to directly correlate $^{1}H^{N}$ resonances with ^{13}C' resonances using the weak J coupling arising from hydrogen bonds (Cordier and Grzesiek, 1999). This experiment did not yield any cross-peaks in a spectrum of ^{13}C/^{15}N-labelled FUN-EGF3 (not shown). An alternative approach is to predict hydrogen bonds using initial calculated structures. However, this can potentially lead to bias, as hydrogen-bond restraints are fixed in subsequent calculations. A better method is to use a potential of mean force that promotes the optimal geometry for N-H and C-O groups in possible hydrogen bond donor-acceptor pairs (Grishaev and Bax, 2004). This has been shown to improve structure quality. In Xplor-NIH, this is implemented as the hydrogen bond database (HBDB) potential. FUN-EGF1 and EGF2-EGF3 structures were refined using the HBDB potential. The resulting structures were analysed to find consistently occuring N-H/C-O pairs. These agreed well with the D_2O exchange data, although there were some differences (Figure 4.14B).

A



B

FUN	*Ex.*	R G G G G H D A L K G P N V C G S R Y N A Y C C P G W K T L P G G N Q C I V P I
	HBDB	R G G G G H D A L K G P N V C G S R Y N A Y C C P G W K T L P G G N Q C I V P I
EGF1	*Ex.*	C R H S C G D G F C S R P N M C T C P S G Q I A P S C G S R S I Q H
	HBDB	C R H S C G D G F C S R P N M C T C P S G Q I A P S C G S R S I Q H
EGF2	*Ex.*	C N I R C M N G G S C S D D H C L C Q K G Y I G T H C G Q P V
	HBDB	C N I R C M N G G S C S D D H C L C Q K G Y I G T H C G Q P V
EGF3	*Ex.*	C E S G C L N G G R C V A P N R C A C T Y G F T G P Q C E
	HBDB	C E S G C L N G G R C V A P N R C A C T Y G F T G P Q C E

Figure 4.14 Hydrogen bonds in FUN-EGF3. A) ^{1}H-^{15}N HSQC of ^{15}N-labelled FUN-EGF3 in 100% D_2O at pH 3.8. This was the 10th spectrum acquired after dissolving freeze-dried protein in D_2O. A lower pH was used than for assignment experiments, as the rate of exchange of $^{1}H^{N}$ nuclei with the solvent was slower. Peaks are labelled with residue numbers. B) Potential hydrogen bond donor residues in FUN-EGF3, highlighted in blue on the primary sequence. The top line (Ex.) shows residues identified by the $^{1}H^{N}$-$^{2}H^{N}$ exchange experiment. The bottom line (HBDB) shows residues identified in the structures refined with the HBDB potential in Xplor-NIH. Highlighted residues had N-H groups that were hydrogen bond acceptors in at least 16 of the 20 lowest-energy structures from ensembles of 200 calculated structures.

4.8.3 Residual dipolar couplings: concepts

Residual dipolar couplings (RDCs) are powerful restraints for structure determination and refinement. While inter-^{1}H distance and torsion angle restraints provide local information about structure, RDCs give insights into the relative orientations of bond vectors irrespective of their separation distance (Lipsitz and Tjandra, 2004). The magnitude of the dipolar coupling *D* between two nuclei is directly related to their orientation relative to a static magnetic field, which is applied in NMR spectroscopy:

$$D = \frac{\kappa}{r^3}(3cos^2\theta - 1) \qquad 4.3$$

where *r* is the distance between the coupled nuclei, ϑ is the angle of the internuclear vector relative to the magnetic field and the pre-factor κ is as follows:

$$\kappa = -\frac{1}{16\pi^2}\gamma_A\gamma_B\,\mu_0 h \qquad 4.4$$

where γ_A and γ_B are the gyromagnetic ratios of nuclei A and B respectively, μ_0 is the permeability of a vacuum and *h* is Planck's constant (Kramer *et al.*, 2004).

Dipolar couplings are not normally evident in NMR spectra of molecules in solution. Rapid tumbling means that the time-averaged values of the couplings are equal to 0. In order to observe dipolar couplings, a degree of alignment needs to be introduced into the sample, so that the protein molecules tumble anisotropically. The result is that time-averaged dipolar couplings are no longer equal to 0, but are still much smaller than would be observed in a solid. The measured values are residual dipolar couplings. If molecules align too strongly, RDCs cause significant peak broadening and therefore reduce the quality of spectra.

4.8.4 RDC measurement

The measurement of RDCs requires the addition of a suitable alignment medium to the molecule of interest. Several types have been developed for proteins (Lipsitz and Tjandra, 2004). Bicelles were found to be suitable for aligning the FUN-EGF3 fragment. They can be produced by mixing a short-chain lipid with a longer-chain lipid in an appropriate ratio. Incubation at a temperature greater than the phase transition temperature causes the lipids to form disc-like structures containing a bilayer in the liquid crystalline phase (Figure 4.15A) (Sanders and Schwonek, 1992). Here, bicelles were made using D13OPC and DHOPC in a ratio of 3.5:1 (mol:mol) (Section 2.6.9.1). Ether-linked lipids were used as ester-linked lipids hydrolyse spontaneously at pH < 6.0 (Ottiger and Bax, 1999). CTAB was also added, as it prevents phase separation and modifies the surface charge of the bicelles to prevent electrostatic interactions with the protein (Lipsitz and Tjandra, 2004; Ottiger and Bax, 1999). A similar mixture was employed successfully in a previous study of lysozyme (Higman *et al.*, 2004).

The lipid mixture was added to $^{13}C/^{15}N$-labelled FUN-EGF3, such that the lipids and protein were at final concentrations of 5% (w/v) and 0.5 mM respectively (Section 2.7.3). Measurements were performed at 35°C, above the phase transition temperature for this lipid mixture. Stable alignment of bicelles over time was monitored by measuring the splitting of the ^{2}H signal arising from D_2O (Section 2.6.9.2; Figure 4.15B). Peak splitting arises from incomplete averaging of the quadrupolar coupling, which is normally averaged to 0 in solution (Ottiger and Bax, 1998). Stable peak splitting over time indicated that the bicelles were stable and the magnitude of splitting indicated the extent of alignment.

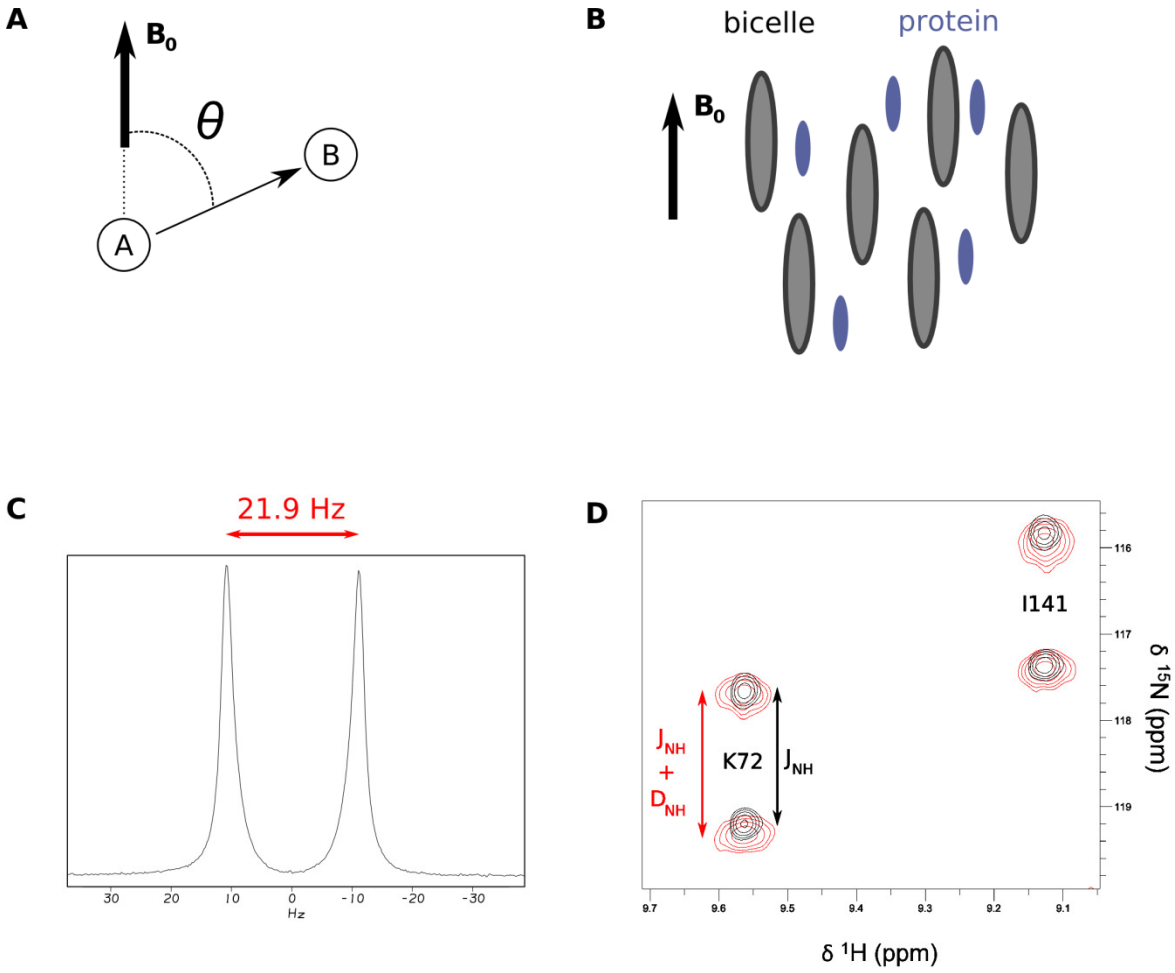


Figure 4.15 Measurement of residual dipolar couplings for the FUN-EGF3 fragment. A) The RDC between nuclei A and B depends on the angle of the internuclear vector from the magnetic field B_0. B) Bicelles align with the magnetic field and cause weak alignment of the protein molecules. C) 1D ^{2}H spectrum showing the split peak resulting from bicelle-induced alignment. D) ^{1}H-^{15}N IPAP-HSQC spectra under isotropic (black) and bicelle-aligned (red) conditions, showing peaks for two residues at 35°C. The contributions of J couplings (J_{NH}) and dipolar couplings (D_{NH}) to peak splitting are indicated.

^{15}N-^{1}H^{N} RDCs were measured using the ^{1}H-^{15}N IPAP-HSQC experiment (Figure 4.15D; Section 2.7.3). The IPAP-HSQC pulse sequence does not eliminate one-bond ^{15}N-^{1}H couplings, so peaks in the resulting spectra were split in the ^{15}N dimension. The experiment reduces spectral crowding by collecting in-phase and anti-phase spectra, which can be added and subtracted to eliminate one of the peaks in the doublet (Ottiger *et al.*, 1998). In the presence of an alignment medium, both scalar and dipolar couplings contribute to the peak splitting, while only J coupling has an influence under isotropic conditions. RDCs were therefore calculated by subtracting the total coupling values measured in these spectra from the values measured from spectra acquired in the absence of the alignment medium.

4.8.5 Use of RDCs in structure refinement

A reference frame has to be defined before using measured RDCs in structural analysis. This is achieved using an alignment tensor ***A*** (Tjandra and Bax, 1997), which defines the relationship of dipolar couplings to the internuclear vector of interest and the Cartesian coordinates of the reference frame (Clore *et al.*, 1998):

$$\boldsymbol{A} = \begin{bmatrix} A_{xx} & 0 & 0 \\ 0 & A_{yy} & 0 \\ 0 & 0 & A_{zz} \end{bmatrix} \qquad 4.4$$

where A_{xx}, A_{yy} and A_{zz} are the principal components of the tensor. Their sum is equal to 0 and $|A_{zz}| > |A_{yy}| > |A_{xx}|$. The following relationship is defined between D^{AB}, the RDC between nuclei A and B and the angles ϑ and φ:

$$D^{AB} = D_a^{AB}(3cos^2\theta - 1) + \frac{3}{2}A_r(sin^2\theta cos2\varphi) \qquad 4.5$$

where $D_a^{AB} = \frac{\kappa}{r^3}(S)(A_a)$ *4.6*

ϑ is the angle between A-B and the z-axis, and φ is the angle between the projection of A-B on the x-y plane and the x-axis. A_a and A_r are the axial and rhombic components of ***A*** respectively. *S* is the order parameter. The rhombicity *R* of ***A*** is defined:

$$R = \frac{A_r}{A_a} \qquad 4.7$$

A_a and *R* are the only values needed to define ***A***, as

$$A_{zz} = 2A_a \qquad 4.8$$

$$A_{yy} = -A_a\,(1 + \frac{3}{2}R) \qquad 4.9$$

$$A_{xx} = -A_a\,(1 - \frac{3}{2}R)\,. \qquad 4.10$$

5 parameters are required to use RDCs in structure calculations: D_a^{AB}, *R* and three angles that define the orientation of ***A*** relative to the coordinate axes of the structures. A common way to implement RDCs is to allow the orientation of ***A*** to vary during refinement. ***A*** rotates in order to find the angles that result in the minimum deviation of calculated RDCs from the measured values (Schwieters *et al.*, 2006). D_a^{AB} and *R* are fixed during the calculation and so must be determined first.

4.8.6 Refinement of FUN-EGF3 with RDCs

The presence of a flexible linker (Section 4.4) and the lack of an observable interface between the EGF1 and EGF2 domains (Section 4.5) suggested that the FUN-EGF1 and EGF2-EGF3 regions behaved independently in solution. Therefore, RDCs for the two domain pairs were treated separately as the alignment tensors for their respective RDCs are likely to be different. Moreover,

it was important to exclude dynamic residues from the analysis. As each RDC depends on the order parameter *S* (Equation 4.6), dynamic residues (with lower *S* values) have smaller RDCs than those in more structured regions of the protein. Although the order parameters were not known, the ^{15}N-{^{1}H} heteronuclear NOE gave an indication of which residues were dynamic (Section 4.4). Residues with ratios greater than 0.60 should have minimal internal dynamics (Clore and Garrett, 1999). RDCs from residues with heteronuclear NOE ratios less than 0.60 were therefore excluded from structure calculations.

There are several possible methods for estimating D_a^{AB} and *R.* The distribution of RDC values can provide valuable information about the alignment tensor (Clore *et al.*, 1998). Unfortunately, the relatively small number of RDCs for each of the two regions of FUN-EGF3 meant it was not possible to calculate the alignment tensor parameters accurately by this method. However, estimates for D_a^{NH} could be obtained by looking at the extreme RDC values − the value with the greatest magnitude arises when a bond vector is parallel to the z-axis (Equations 4.5 and 4.8). For FUN-EGF1, the largest RDC was 22.7 Hz, suggesting a D_a^{NH} of at least 11.4 Hz. For EGF2-EGF3, the value of the greatest magnitude was −15.9 Hz, suggesting a D_a^{NH} value of less than or equal to −8.0 Hz.

D_a^{AB} and *R* can also be estimated using the atomic coordinates of known structures. For particular values of D_a^{AB}, *R* and the 3 angles defining the orientation of the alignment tensor (Section 4.8.5), it is possible to calculate an RDC value for a given residue using equation 4.5. The equation is fitted to the data to find a set of 5 parameters that minimises the overall deviation of the calculated RDCs from the measured RDCs. Fitting was performed for N-H RDCs with partially refined structures of FUN-EGF3 (Section 2.7.4), using separate alignment tensors for FUN-EGF1 and EGF2-EGF3. For FUN-EGF1, structures with the best fits gave D_a^{NH} and R values of 14.0 Hz and 0.32 respectively. For EGF2-EGF3, structures with the best fits gave D_a^{NH} and R values of -8.3 Hz

and 0.48 respectively. Trial calculations using these tensor parameters produced structures with minimal violation of RDC and other restraints (data not shown).

4.9 Final structure ensembles

Final ensembles of refined structures were separately generated for FUN-EGF1, EGF2-EGF3 and full-length FUN-EGF3 (Section 2.7.3). For each one, the 10 lowest-energy structures from an ensemble of 200 partially refined structures were used as starting structures in refinement calculations. 20 structures were generated per template, resulting in a total of 200 structures. The 20 lowest-energy structures from this ensemble were chosen as the final structure ensemble. As expected, the orientation of the FUN-EGF1 region relative to EGF2-EGF3 was highly variable in the FUN-EGF3 structures (Figure 4.16A). The lowest-energy structure of FUN-EGF3 is shown in Figure 4.16B. Final structure ensembles of FUN-EGF1 (Figure 4.16C) and EGF2-EGF3 (Figure 4.16D) showed that the relative orientations of the two domains were well defined in both domain pairs. Energy-minimised average structures were also calculated for FUN-EGF1 (Figure 4.16E) and EGF2-EGF3 (Figure 4.17F) using Xplor-NIH (Section 2.7.5).

4.9.1 Quality statistics

The final structures showed good agreement with the experimental restraints (Table 4.4). There were no consistent distance or dihedral restraint violations, or violations of bond angle, bond length or improper terms. In some structures, calculated RDC values exceeded the violation threshold of 1.5 Hz. However, the average RMSD from the expected values was low and likely to be within measurement error.

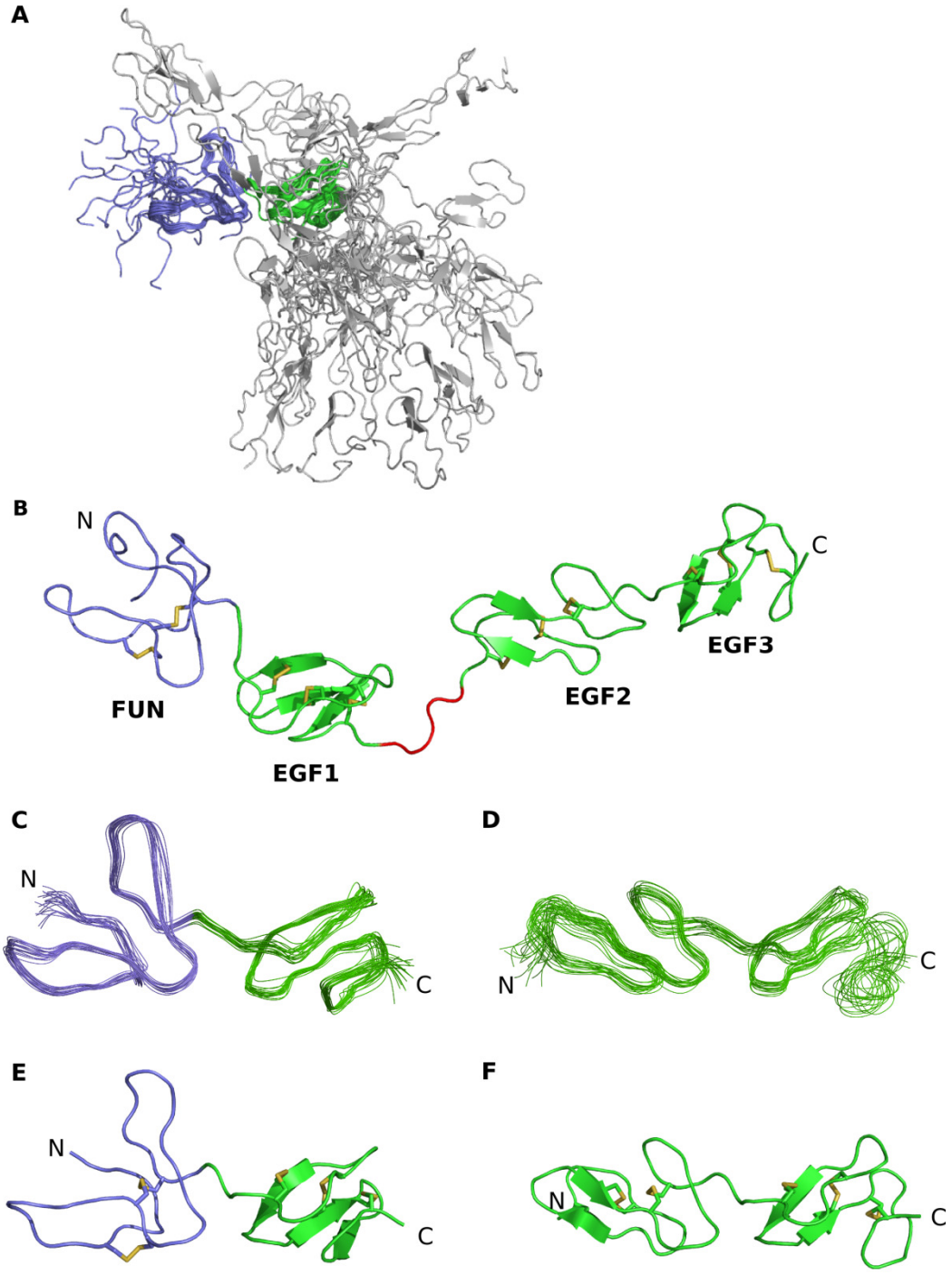


Figure 4.16 Final structures of the FUN-EGF3 fragment. A) 20-structure ensemble of FUN-EGF3, with structures aligned to the FUN-EGF1 region, rendered in Pymol. The FUN and EGF1 domains are coloured lilac and green respectively; the EGF2-EGF3 region is coloured gray. B) Lowest-energy structure of FUN-EGF3. C) 20-structure ensemble of FUN-EGF1, showing residues 52-112, rendered in ICM Molsoft Browser (http://www.molsoft.com/icm_browser.html). D) 20-structure ensemble of EGF2-EGF3, showing residues 118-178. E) Energy-minimised average structure of FUN-EGF1. F) Energy-minimised average structure of EGF2-EGF3.

Table 4.4 Statistics for the final FUN-EGF3, FUN-EGF1 and EGF2-EGF3 structure ensembles. Structures were calculated separately for full-length FUN-EGF3, FUN-EGF1 and EGF2-EGF3. Quality statistics were calculated using Xplor-NIH. Restraint violations and RMSD values are given as mean values per structure ± the standard deviation.

	FUN-EGF3 (R45-Q178)	**FUN-EGF1** (A52-G112)	**EGF2-EGF3** (H118-E178)
NOE-derived distance restraints			
Total	2599	1559	1040
Intra-residue	890	516	374
Inter-residue	1434	893	541
Sequential (| i-j |= 1)	571	340	231
Short-range (| i-j |< 5)	231	147	84
Long-range (| i-j |≥ 5)	632	406	226
Ambiguous	275	150	125
Hydrogen bond restraints (HBDB)	33	19	14
Dihedral angle restraints			
Total	135	67	68
ϕ	72	38	34
ψ	63	29	34
Residual dipolar couplings (RDCs)			
$^{1}D_{NH}$	54	28	26
Total number of restraints	2821	1673	1148
Restraint violations			
Distance restraint violations > 0.3 Å	0.5 ± 0.7	0.3 ± 0.5	0.0 ± 0.0
Dihedral angle violations > 5°	0.1 ± 0.2	0.1 ± 0.3	0.0 ± 0.0
RDC violations > 1.5 Hz	2.8 ± 1.4	0.8 ± 1.0	1.4 ± 0.8
RMSD from experimental restraints			
Distance restraints (Å)	0.022 ± 0.001	0.023 ± 0.001	0.020 ± 0.002
Dihedral angle restraints (°)	0.248 ± 0.127	0.175 ± 0.226	0.326 ± 0.135
RDC restraints (Hz)	0.717 ± 0.067	0.615 ± 0.090	0.723 ± 0.084
RMSD from idealised geometry			
Bonds (Å)	0.003 ± 0.000	0.004 ± 0.000	0.003 ± 0.000
Angles (°)	0.492 ± 0.009	0.512 ± 0.010	0.462 ± 0.010
Impropers (°)	0.320 ± 0.007	0.348 ± 0.013	0.277 ± 0.010

4.9.2 Structure precision

Structure precision was examined by calculating RMSD values for atomic coordinates relative to the energy-minimised average structures of the FUN-EGF1 and EGF2-EGF3 regions (Section 2.7.5). Plotting the backbone average RMSD values for each residue against amino acid number indicated how well particular regions of the structure were defined (Figure 4.17). As expected, the N- and C-terminal residues of the FUN-EGF1 region (A52 and G112 respectively) had higher average RMSD values (Figure 4.17A). With the exception of G90, all other residues had backbone RMSD values less than 1 Å. Regions with better precision generally coincided with regions of defined secondary structure. However, the RMSD values and heteronuclear NOE ratios were not always well correlated. For example, residues L53-K54 and S61-A65 had heteronuclear NOE ratios less than 0.60 (Section 4.4), but did not have significantly higher RMSD values than adjacent residues. Inter-residue NOEs to these residues in these regions are likely to have improved the definition of their structures.

The EGF2-EGF3 region showed greater variability in the average backbone RMSD (Figure 4.17B). The N- (H118-R122) and C-terminal (T169-Q178) regions were not well defined. In addition residues in the turn between strands of the β-hairpin of the EGF2 domain, S130-H133 were less well defined. Residues in the loop before the first strand of the β-hairpin in the EGF3 domain, S152-G157, also had higher RMSD values. All of these regions were defined by fewer NOE and RDC restraints than the regions in proximity to the interdomain interface. Further statistics are given in Table 4.5, which summarises RMSD values for the two regions of FUN-EGF3 and individual domains.

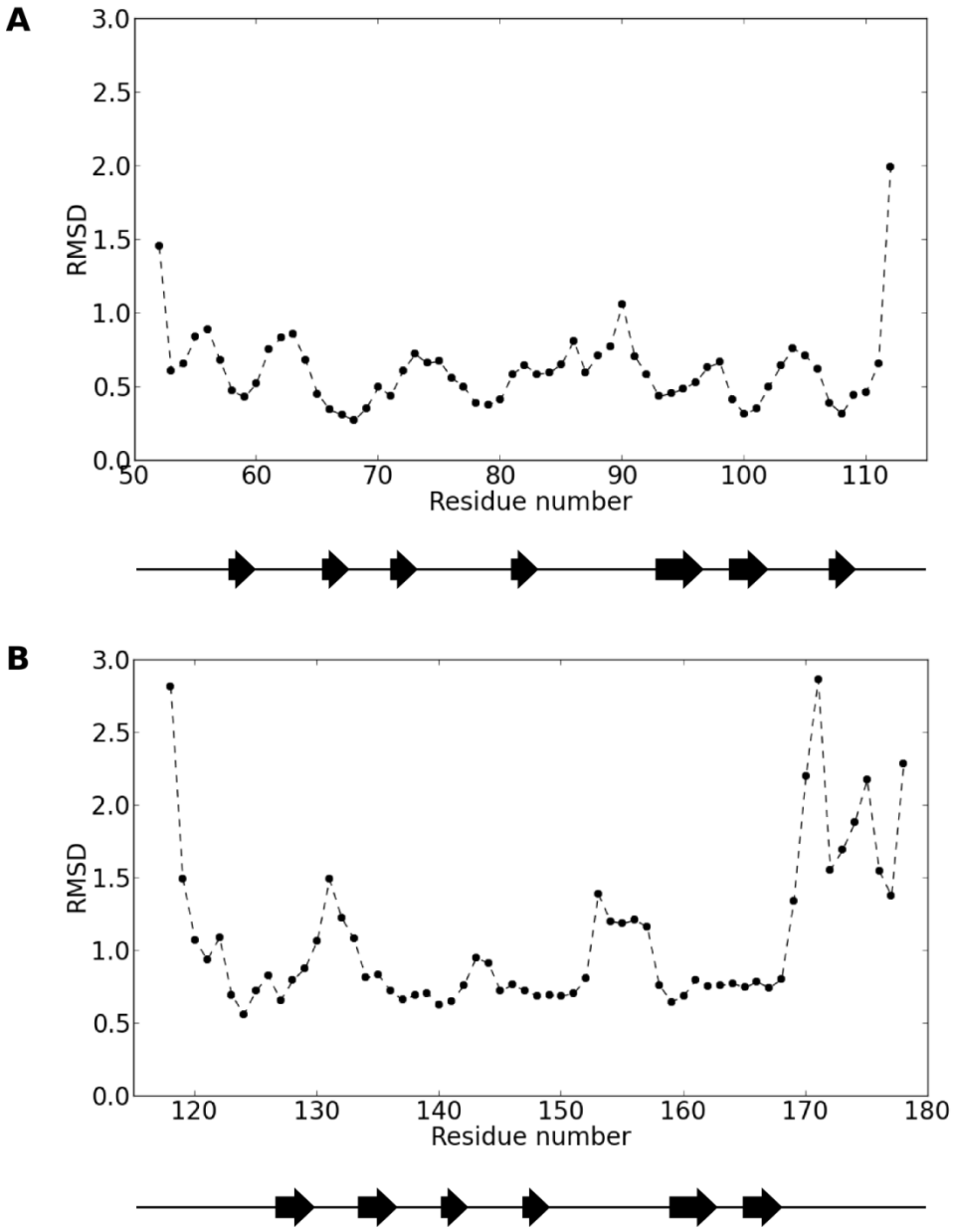


Figure 4.17 Structure precision across the FUN-EGF3 fragment A) Backbone RMSD values from the energy-minimised average structure of FUN-EGF1 are plotted as a function of sequence. Regions of β-secondary structure are illustrated below the graph. B) Backbone RMSD versus sequence for EGF2-EGF3. RMSD values were calculated using a Python script and regions of defined secondary structure were identified using Procheck (Laskowski *et al.*, 1996).

Table 4.5 Structure precision statistics. Structures were aligned to the energy-minimised average structures of FUN-EGF1 or EGF2-EGF3 before performing calculations. Average RMSD values are given relative to the energy-minimised average structures. 'Core' refers to the core structured regions, excluding the N-terminal sequence (residues 3-51) and the EGF1-EGF2 linker sequence (residues S113-Q117). 'Ordered' values only include residues with heteronuclear NOE ratios > 0.60.

	Backbone heavy atoms (N, C, Cα) (Å)	**All heavy atoms (Å)**
FUN-EGF1 (A52-G112)		
Core	0.653 ± 0.131	0.984 ± 0.149
Ordered	0.559 ± 0.126	0.956 ± 0.174
EGF2-EGF3 (H118-E178)		
Core	1.161 ± 0.227	1.851 ± 0.248
Ordered	0.844 ± 0.187	1.417 ± 0.182
FUN domain (A52-I81)		
Core	0.627± 0.115	0.935± 0.158
Ordered	0.496± 0.107	0.757± 0.120
EGF1 domain (V82-G112)		
Core	0.674± 0.164	1.020± 0.213
Ordered	0.599± 0.149	1.074± 0.250
EGF2 domain (H118-G146)		
Core	1.019± 0.144	1.652± 0.183
Ordered	0.838± 0.141	1.408± 0.170
EGF3 domain (Q147-E178)		
Core	1.263± 0.341	1.999± 0.395
Ordered	0.825± 0.315	1.408± 0.313

4.9.3 Structure validation

Structures were validated with Procheck-NMR (Laskowski *et al.*, 1996) and Molprobity software (Lovell *et al.*, 2003) using the Protein Structure Validation Software suite (http://psvs-1_4-dev.nesg.org/). These programs analyse torsion angles and inter-atom clashes. They showed that the majority of residues were in allowed regions of the Ramachandran plot (Table 4.6). The FUN-EGF3 statistics were worse than the combined statistics of the FUN-EGF1 and EGF2-EGF3, as the full-length structure included residues in unstructured regions. These residues adopted disallowed conformations in some of the structures. Closer inspection of these residues highlighted a number of residues that adopted disallowed conformations more frequently. P75 in the FUN domain lies just outside the allowed region and residues C145-G146 in the EGF2 domain were in disallowed regions in some structures. Therefore, restraints involving these residues need to be re-examined. Furthermore, the Molprobity clash scores were relatively high. A possible explanation was that scale factor for the van der Waals restraints was too low. The scale factor was ramped between 0.9 and 0.8 in the refinement calculations (Section 2.7.3). The effect of this parameter on the clash score of the final structures needs to be investigated further.

Table 4.6 Validation statistics. Statistics are given for analysis by Procheck and Molprobity. The clash score indicates the number of bad clashes per 1000 atoms.

	FUN-EGF3 (R45-Q178)	**FUN-EGF1** (A52-G112)	**EGF2-EGF3** (H118-Q178)
Procheck			
Most favoured regions	73.1%	79.3%	83.1%
Additionally allowed regions	21.9%	20.2%	14.3%
Generously allowed regions	3.1%	0.2%	1.6%
Disallowed regions	1.9%	0.2%	1.0%
Molprobity			
Most favoured regions	85.8%	90.5%	91.1%
Additionally allowed regions	9.5%	7.9%	5.9%
Disallowed regions	4.7%	1.6%	3.0%
Clash score	19.72	18.72	22.70

4.10 Domain structures in the FUN-EGF3 fragment

The structure of the FUN-EGF3 fragment showed the fold of the FUN domain and interdomain packing interactions in this region for the first time. Like cbEGF and EGF domains, the FUN domain has little regular secondary structure (Figure 4.18A). It comprises two loops linked by two disulphide bonds and an N-terminal sequence that packs in between. The disulphide bond pairs are C59-C68 and C67-C80 − a C1-C3, C2-C4 arrangement. The EGF-like domains all have the same topology and pattern of disulphide bonds: C1-C3, C2-C4 and C5-C6, as expected (Figure 4.18B-D). However, there are differences in their sizes and shapes. All three have the canonical β-hairpin, but the EGF1 domain has an additional β-strand, forming a small three-stranded β-sheet (Figure 4.18B). The C-terminal loop in the EGF3 domain, between C5 and C6, is not well defined, consistent with the lower heteronuclear NOE values for the residues in this region (Figure 4.7). In full-length fibrillin-1, this loop is likely to be involved in packing with the hyb1 domain. In terms of sequence, EGF1 is distinct from EGF2 and EGF3 (Figure 4.18E). It lacks the asparagine residue between C2 and C3 and the aromatic residue between C5 and C6. However, EGF1 and EGF3 both have the P-N sequence between C3 and C4, which is missing in EGF2.

There are some similarities between the structures of the FUN and EGF-like domains, despite little obvious sequence similarity. Superimposing the FUN domain structure on the EGF2 domain shows that their C-terminal segments are similar (Figure 4.18F-G). The β-hairpin in the EGF2 domain is replaced by a longer, flexible loop in the FUN domain. In EGF-like domains, the β-hairpin is linked to the N-terminal segment by disulphide bonds. This is not the case in the FUN domain, which may explain why these regions are more flexible, as demonstrated by the lower heteronuclear NOE ratios values for these residues. The loop between the penultimate and final cysteines is also much longer in the FUN domain than in the EGF-like domains.

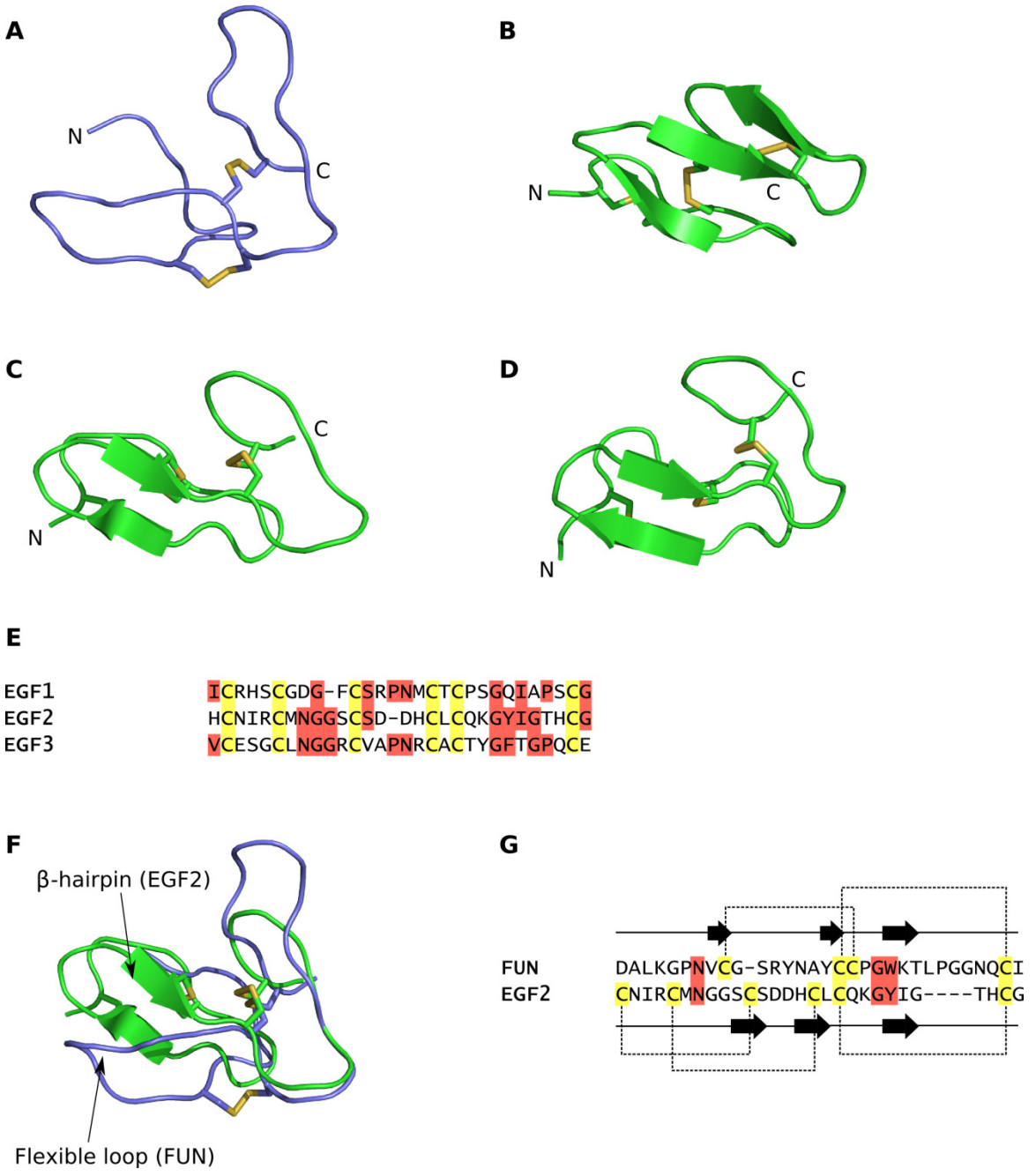


Figure 4.18 Structures of domains in the FUN-EGF3 fragment. Structures are displayed using the cartoon representation in Pymol. A) FUN domain. B) EGF1 domain. C) EGF2 domain. D) EGF3 domain. E) Sequence alignment of EGF domains 1-3. Cysteine residues are coloured yellow and other conserved residues are highlighted in red. Sequences were aligned using the ClustalW2 server (http://www.ebi.ac.uk/Tools/msa/clustalw2/). F) EGF2 domain (green) superimposed on the FUN domain (lilac). The β-hairpin in EGF2 and the flexible loop in the FUN domain are indicated. G) Comparison of the FUN and EGF2 domain sequences. Sequences were aligned manually on the basis of structural homology. Arrows represent stretches of β-sheet-like secondary structure (based on torsion angles) and dotted lines indicate disulphide bonds.

However, in this case the loop in the FUN domain appears to be relatively rigid, probably stabilised by hydrophobic interactions with other parts of the domain and with the EGF1 domain.

4.11 Interdomain packing

Interdomain packing influences the stability of interdomain interfaces and the relative orientation of domains. It is clear from the structure and primary sequence of FUN-EGF3 that the FUN-EGF1 and EGF2-EGF3 interdomain interfaces are similar, while there is no observable interface between the EGF1 and EGF2 domains. The FUN-EGF1 interface is formed by packing interactions between residues G70, W71, K72, L73 and I81 in the FUN domain, and residues V82, P83, I84, and P97 in the EGF1 domain (Figure 4.19A). Domain boundaries were defined according to the exon boundaries of the *FBN1* gene (Pereira *et al.*, 1993). Packing residues were identified by their proximity to residues in the opposing domain, using a cut-off distance of 4 Å between atoms (Section 2.7.6), as described previously for the cbEGF9-hyb2-cbEGF10 fragment (Jensen *et al.*, 2009). Analogous packing residues are found at the EGF2-EGF3 interface: G139, W140 and I141 in the EGF2 domain, and residues Q147, P148, V149 and P163 in the EGF3 domain (Figure 4.19B). However, there are no analogues of L73 and I81 in the EGF2 domain. Sequence alignments of the two domain pairs confirm that the packing residues occur at equivalent positions (Figure 4.19C). Interestingly, these residues are missing from the EGF1-EGF2 domain pair, explaining the lack of an observed interface (Figures 4.18E and 4.19C).

Other interactions, in addition to hydrophobic and van der Waals contacts, are likely to stabilise the interdomain interfaces. At the FUN-EGF1 interface, there are hydrogen bonds between the G-W-K motif in the FUN domain and V-P-I motif in the EGF1 domain. There are equivalent interactions at the EGF-EGF3 interface between the G-Y-I and Q-P-V motifs. Residue N57 in the FUN domain, and its equivalent N125 in the EGF2 domain, may also contribute to the stability of

the interface. Side-chain NH_2 peaks assigned to these two residues have unusual chemical shifts (Figure 4.6). The $^{1}H^{\delta 2}$ resonances of N57 are upfield-shifted, with chemical shifts of 5.05 and 6.09 ppm, while the $^{15}N^{\delta}$ resonance of N125 is downfield-shifted, with a chemical shift of 118.47 ppm. These chemical shifts are likely to result from the proximity of the side chains to aromatic rings.

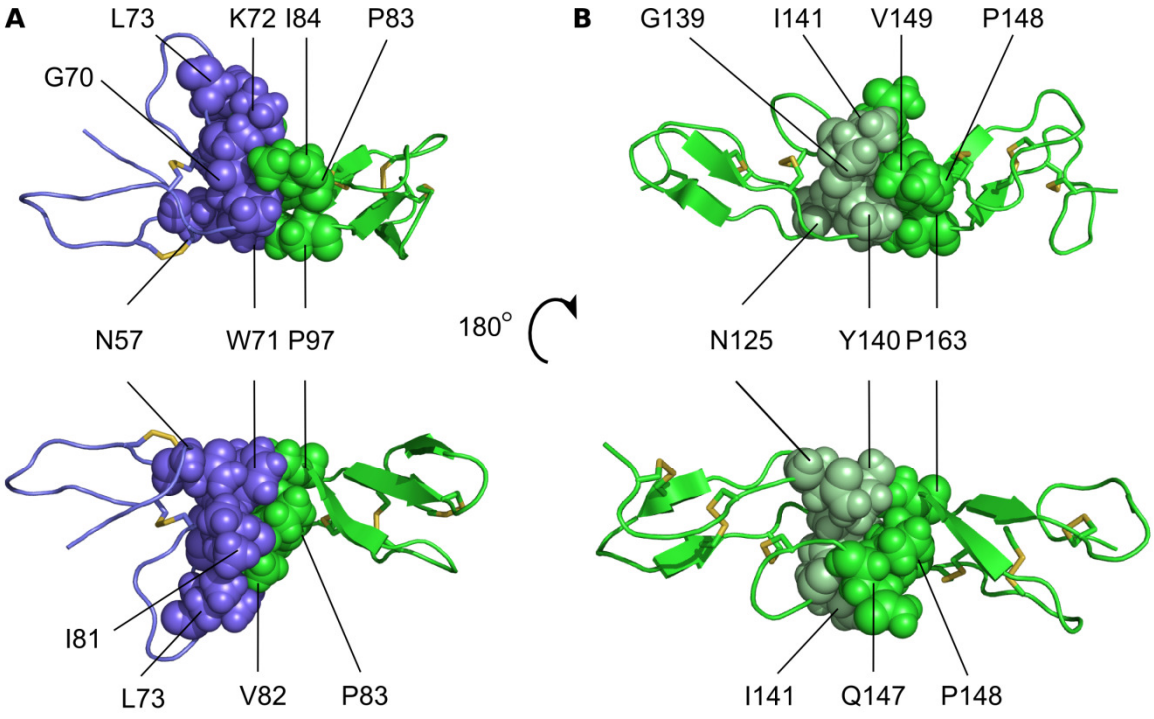


Figure 4.19 Interdomain packing interactions in the FUN-EGF3 fragment. A) Packing interactions between the FUN and EGF1 domains, showing two views of opposite faces. Residues involved in interdomain packing are shown as spheres and are coloured according to domain. B) Packing interactions between the EGF2 and EGF3 domains. Residues from EGF2 and EGF3 involved in packing are coloured light green and green respectively for clarity. C) Sequence alignment of the FUN-EGF1 and EGF2-EGF3 regions. Cysteines and packing residues are coloured yellow and red respectively. The dotted line delimits the domain boundaries.

Consistent with these observations, the NH_2 groups of N57 and N125 lie above or close to the planes of the W71 and Y98 side-chains respectively, although their precise orientations are not the same in all structures of the ensemble. It is likely that there is a favourable interaction between the partial positive charge on the $H^{\delta 2}$ atoms and the π electrons of the aromatic ring. Similar interactions between positively charged side-chain groups and aromatic rings are commonly found at interfaces of protein-protein and protein-DNA complexes (Biot *et al.*, 2003). Moreover, they often occur at the surface of proteins. In this case, the asparagine-aromatic interaction may serve to shield the aromatic ring from the solvent.

Additional observations confirmed the importance of the packing residues. Fibrillin N-terminal sequences from a range of animal species were aligned (Figure 4.20). As well as the positions of the cysteine residues, several residues are absolutely conserved: the G-Y/W sequence in the N-terminal domain of the pair, as well as P83/P148 and P97/P163 in the C-terminal domain. Others found in human fibrillin-1 are replaced by similar residues in other species, but some are very different. It is not clear whether other residue types could perform the same role as the packing residues in human fibrillin-1. However, the variable length and sequence of the EGF1-EGF2 linker sequence is consistent with the lack of an interface between these domains.

While N125 in EGF2 appears to be completely conserved, N57 is replaced by a histidine in some species. To further investigate the potential structural role of these residues, the dynamics of their side-chains were examined. $\{^1H\}$-^{15}N heteronuclear NOE ratios were measured for NH_2 side-chain groups in an analogous fashion to the backbone N-H groups (Section 2.6.4; Table 4.7) (Buck *et al.*, 1995). The NH_2 groups of N57 and N125 had much higher heteronuclear NOE ratios than those of other asparagine and glutamine side chains. This suggested that their side chains had limited mobility, rather than rotating rapidly about the C^{β}-N^{δ} (asparagine) or C^{γ}-N^{ε} (glutamine) bond. This was consistent with their proposed role in packing with aromatic side chains.

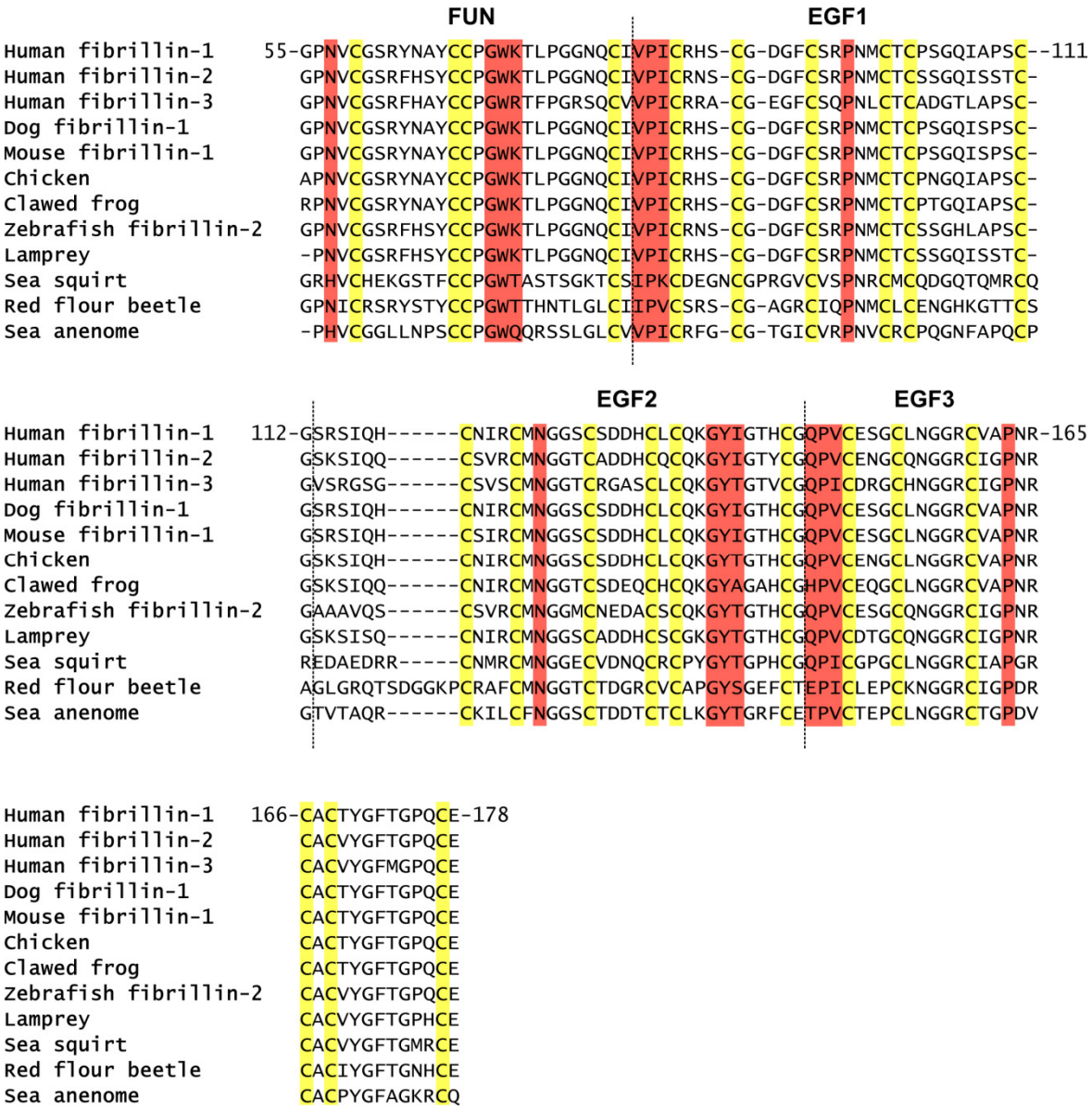


Figure 4.20 Conservation of interdomain packing residues in the FUN-EGF3 region. Sequences of the FUN-EGF3 region from a range of species across the phylogenetic tree were aligned. They were obtained from NCBI and Ensembl databases (references in Appendix F) and aligned using ClustalW2. The colour scheme is the same as in Figure 4.19. Packing motifs are present in sequences from all species examined, but not all individual residues are absolutely conserved. In particular, L73 in the FUN domain is absent from fibrillin-3 and sea squirt, red flour beetle and sea anenome fibrillins.

Table 4.7 Heteronuclear NOE ratios for NH_2 groups in glutamine and asparagine side chains. Values were measured in a separate experiment from the one used to calculate the backbone heteronuclear NOE values. The delay period between the 90° ^{15}N pulse and the 180° ^{1}H pulse in the INEPT portion of the pulse sequence was set to 1.375 ms instead of 2.75 ms (used for backbone N-H groups) to maximise the efficiency of magnetisation transfer from ^{15}N to ^{1}H in NH_2 groups (Section 2.6.4). Heteronuclear values were calculated as for the backbone N-H groups. A value was not calculated for the second ^{1}H of the N98 NH_2 group due to peak overlap. For N120, there were no discernible peaks in the NOE spectrum. Therefore, the heteronuclear NOE ratios were likely to be close to 0.

Residue	Peak 1		Peak 2	
	NOE	Error	NOE	Error
N57	0.646	0.072	0.645	0.059
Q64	-0.388	0.01	-0.376	0.01
N78	0.37	0.026	0.358	0.025
Q79	-0.481	0.012	-0.489	0.013
N98	0.17	0.024	–	–
Q106	-1.215	0.14	-1.003	0.01
Q117	-1.399	0.01	-1.435	0.011
N120*	0	0	–	–
N125	0.627	0.128	0.669	0.085
Q137	-0.85	0.016	-0.871	0.015
Q147	-0.78	0.099	-0.725	0.077
N164	0.122	0.022	0.244	0.026
Q176	-0.59	0.016	-0.589	0.014

4.12 Comparisons with other interdomain interfaces in human fibrillin-1

The interdomain interfaces were compared to others in fibrillin. First, the twist and tilt angles of the EGF2-EGF3 pair were calculated using the Mod2 program (Downing *et al.*, 1996) (Section 2.7.6). It is not possible to measure these angles for the FUN-EGF1 pair, as these domains lack homologous regions of the structure that can be used to define planes needed to perform the alignment and rotation operations. Using the β-hairpin structures to define the planes of the EGF2 and EGF3 domains, the average twist and tilt angles for the ensemble are 138 ± 7° and 47 ± 9° respectively. The twist angle is slightly less than for the fibrillin-1 cbEGF32-cbEGF33 and cbEGF12-cbEGF13 domain pairs, which have values of 159 ± 6° and 152 ± 13° respectively. By contrast, the tilt angle for EGF2-EGF3 is larger than for the cbEGF-cbEGF pairs: 18 ± 6° for cbEGF32-cbEGF33 and 30 ± 15° for cbEGF12-cbEGF13.

Second, the surface area buried by the FUN-EGF1 and EGF2-EGF3 interdomain interfaces was calculated from the average structures using the Areaimol program in the CCP4 software package (Collaborative Computational Project, 1994) (Section 2.7.6). The FUN-EGF1 interface buries a surface area of 290.0 $Å^2$, averaged across the two surfaces, while the buried surface area of the EGF2-EGF3 pair is 239.0 $Å^2$. L73 and I81 in the FUN domain are likely to account for the slightly larger interaction surface in FUN-EGF1, are there are no equivalent residues in the EGF2 domain. Both interfaces are slightly more extensive than in cbEGF-cbEGF pairs, but smaller than in TB-cbEGF and hyb-cbEGF pairs. The cbEGF32-cbEGF33 interface has a buried surface area of 183.6 $Å^2$, while the TB4-cbEGF23 and hyb2-cbEGF10 interfaces have buried areas of 552.3 $Å^2$ and 670.0 $Å^2$ respectively (Jensen *et al.*, 2009; Lee *et al.*, 2004).

4.13 Comparisons with EGF-EGF pairs in other extracellular proteins

The FUN-EGF1 and EGF2-EGF3 domain pairs are unique in fibrillin – the only other non-calcium binding EGF-like domain in human fibrillin-1, EGF4, is flanked by the Pro-rich region and the cbEGF3 domain. A search of known protein structures using the Dali alignment server (Holm and Rosenström, 2010) did not find proteins with structures similar to the FUN-EGF1 region. However, there are some other examples of EGF-EGF domain pairs with calcium-independent interfaces for which structures are known. In addition to the Dali search, these were identified by a basic local alignment search (BLAST) of proteins with known structures (Altschul *et al.*, 1990) and the Protein Modal Portal (Arnold *et al.*, 2009), using the primary sequence of the EGF2-EGF3 pair. Human Jagged-1 contains several non-calcium binding EGF domains. The structure of the Jagged-1 DSL-EGF3 fragment (Cordle *et al.*, 2008), containing three EGF-like domains, shows that the first two (EGF1 and EGF2) are most similar in size and shape to the fibrillin-1 EGF2-EGF3 pair, with a backbone RMSD of 2.9 Å. The EGF1-EGF2 domain pair at the C-terminus of human hedgehog interacting protein (HHIP) ectodomain (Bishop *et al.*, 2009) and the EGF2-EGF3 domain pair from Wnt inhibitory factor-1 (WIF-1) (Malinauskas *et al.*, 2011) are also similar to the fibrillin domains , with backbone RMSD values of 2.1 Å and 3.1 Å compared to fibrillin-1 EGF2-EGF3 respectively.

There are similarities and differences between the structures of the domain pairs (Figure 4.21A-D). All four domain pairs contain a core G-F/Y/W sequence in the N-terminal domain and the proline residue in the C-terminal domain (Figure 4.21E). The HHIP EGF1-EGF2 pair is the most similar to the fibrillin EGF2-EGF3 pair, with the lowest backbone RMSD. However, the four residues between C6 of the N-terminal EGF and C1 of the C-terminal EGF are different. In particular, the conserved proline residue (P148) in the fibrillin EGF2-EGF3 pair is replaced by an alanine (A637) in HHIP. Nonetheless, in both cases, the packing residues are arranged in a similar fashion (Figures 4.21A-B). The buried surface area of the HHIP domain pair is 259.1 $Å^2$, slightly

larger than for the fibrillin pair (Section 4.12). While the twist angle is slightly larger than for fibrillin EGF2-EGF3 (156°), the twist angle is very similar (47°).

The interdomain interface between the EGF2 and EGF3 domains of WIF-1 appears different, even though it contains the consensus packing sequences found in fibrillin-1 and HHIP. The proline residue (P259) in the loop between C3 and C4 of the EGF3 domain does not pack closely with the aromatic residue (F246) in the EGF2 domain (Figure 4.21C). While the surface area buried by the interaction surface (257.3 $Å^2$) is comparable to that of the fibrillin EGF2-EGF3 pair, the twist and tilt angles are different: 108° and 23° respectively − that is, the domains adopt a more linear conformation.

The packing residues of the Jagged-1 EGF1-EGF2 pair are different from the other three EGF-EGF pairs (Figure 4.12E). There are only two residues between C6 of EGF1 and C1 of EGF2, two fewer than in the other three EGF-EGF pairs described. One of these residues, K264, is nonetheless involved in interdomain packing. In addition, a tyrosine residue (Y255) from EGF1 packs against a tryptophan (W280) in EGF2. These two residues are absent from the fibrillin and HHIP and WIF-1 EGF domains. The consequence of these differences is that the Jagged EGF1-EGF3 buries a larger surface area (316.8 $Å^2$). Moreover, although the twist angle, 138°, is very similar to that of the fibrillin pair, the tilt angle of 23° is more similar to the WIF pair. In addition, the asparagine residue in fibrillin-1 (N125) and WIF-1 (N220) that packs against an aromatic side chain is replaced by histidine (H242) in Jagged-1 and arginine (R615) in HHIP. This is consistent with the variation in the residue type at this position between species in fibrillin (Section 4.11). Moreover, it suggests that multiple residue types could fulfil the same role.

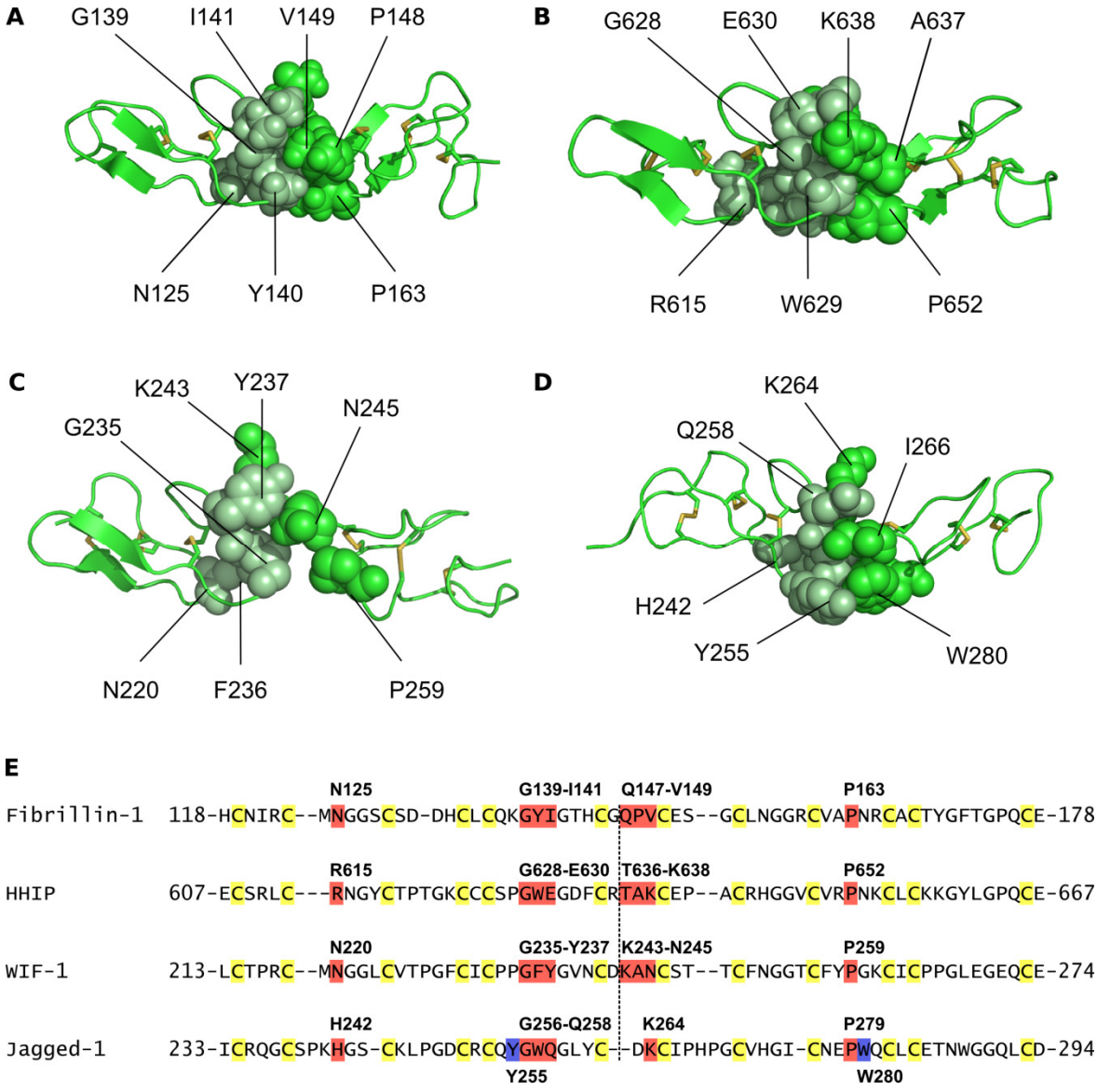


Figure 4.21 Comparison of interdomain packing interactions in other EGF-EGF domain pairs. A) Human fibrillin-1 EGF2-EGF3 domain pair showing residues involved in interdomain packing. B) EGF1-EGF2 domain pair from the human hedgehog interacting protein (HHIP) structure (PDB ID 3H0B). C) EGF2-EGF3 domain pair from the human Wnt inhibitory factor-1 (WIF-1) structure (PDB ID 1IVO). D) EGF1-EGF2 domain pair from the human Jagged-1 DSL-EGF3 structure (PDB ID 2V2J). E) Alignment of sequences from fibrillin-1, HHIP, WIF-1 and Jagged EGF-EGF domain pairs. The colour scheme is the same as in Figure 4.19 and packing residues are labelled above the sequences. Additional packing residues in the Jagged-1 domain pair are highlighted in blue and are labelled below the sequence. Packing residues were identified by the same method described for the fibrillin domain pairs (Section 4.11). The HHIP EGF-EGF pair has an additional cysteine residue between C3 and C4 that forms a disulphide bond with another portion of the HHIP structure (not shown).

4.14 Discussion

NMR spectroscopy was used here to determine the structure of the FUN-EGF3 fragment, for which there was previously no high-resolution structural information. NOE-derived distance restraints and torsion angle restraints were used to generate initial structures, which were then refined using hydrogen bond and RDC restraints. Further work is required to improve the quality of the structures. Torsion angles of some residues (Table 4.6) lie outside the allowed regions of the Ramachandran plot and so some of the restraints need to be checked. Additional RDC restraints, measured from protein in a different alignment medium may also be useful for structure validation, in order to confirm that structures were correctly refined with the RDCs used here. Structures could also be refined further using a more realistic force-field involving electrostatics, a complete Lennard-Jones potential function and a shell of water molecules. So-called water refinement has been shown previously to improve overall structure quality, including side-chain packing (Linge *et al.*, 2003).

The structure of the FUN-EGF3 fragment adds to the already large body of information about fibrillin structure and interdomain interactions. This new information is particularly important for the fibrillin N-terminal domains because they are distinct from the rest of the molecule, which comprises mostly arrays of cbEGF domains interrupted by TB and hyb domains. The FUN domain, which is unique to the fibrillins and LTBPs, is distinct from the other domain types found in these proteins. It appears to have a novel fold, as a Dali search did not identify any structurally homologous domains in other proteins with known structures (Section 4.13). The domain has two disulphide bonds in a C1-C3 C2-C4 arrangement, analogous to the arrangement of the first two disulphides in EGF-like domains. However, the spacing of cysteine residues is different and the C-terminal C5-C6 disulphide in the EGF-like domain is structurally homologous to the C2-C4 disulphide bond in the FUN domain. The FUN domain is also a more dynamic structure than the

EGF-like domain, containing longer loops that are not linked to the core of the structure by disulphide bonds. At present, the functional significance of these differences is unclear.

In addition, the interdomain packing interactions in the FUN-EGF3 structure are distinct from those found elsewhere in fibrillin and in other extracellular, EGF domain-containing proteins. The homologous FUN-EGF1 and EGF2-EGF3 interfaces are formed by the packing of non-polar side-chain groups contributed by both domains. The orientation of the key aromatic residue in the G-F/Y/W motif in the N-terminal domain is likely to be stabilised by interactions with an asparagine, histidine or arginine side chain. Packing interactions appear to maintain a rod-like arrangement of the FUN-EGF1 and EGF2-EGF3 pairs, as in cbEGF-cbEGF pairs. Structural information about these interdomain interfaces is critical for modelling the overall conformation of fibrillin and therefore also its arrangement in microfibrils.

Many of the packing residues are conserved between fibrillin isoforms and species. Analogous domain pairs in other proteins contain elements of these packing residues, but so far, no identical interfaces have been identified outside fibrillin. The HHIP EGF1-EGF2 pair is the most similar pair for which a structure is known. Although the WIF EGF2-EGF3 pair contains similar packing motifs to the fibrillin EGF2-EGF3 pair, its twist and tilt angles are different. Therefore, more high-resolution structures are required before predictions about the interdomain interfaces can be made from protein sequences. Nonetheless, the similarity of the arrangement of packing residues between the various EGF-EGF pairs suggests that the fold of the FUN-EGF3 fragment is native-like, rather than being an artefact of *in vitro* refolding.

The key packing residues at these interfaces are absent from the EGF1-EGF2 domain pair, which explains the lack of cross-peaks in NOESY spectra between peaks assigned to residues in these domains. Moreover, the interdomain linker between the EGF1 and EGF2 domains appears to be

unstructured. This is significant as it is only the second flexible region to be conclusively identified in fibrillin. The only other known dynamic interface occurs between the TB6 and cbEGF32 domains (Yuan *et al.*, 2002). This has implications for the organisation of fibrillin in microfibrils, as the orientation of the FUN-EGF1 relative to the rest of the molecule may not be fixed. On the basis of the available evidence, it seems likely that the FUN-EGF1 and EGF2-EGF3 pairs behave independently, but more detailed analysis of the dynamics is required.

Most importantly, the structure lays the foundation for the characterisation of the interaction of the fibrillin N-terminus with the fibrillin C-terminus and other ECM components. It can potentially allow the mapping of binding sites, important for understanding how elastic fibres are assembled. As resonance assignments are now available for the FUN-EGF3 fragment, NMR-based interaction studies are possible. They will also allow the assessment of residue substitutions on the structure of the protein. This is necessary to validate the results of the interaction studies using modified proteins. NMR assignments also make it possible to assess effects the structural effects of MFS mutations without having to determine the structure of modified proteins. These potential applications are investigated further in Chapters 5 and 6.

Chapter 5

Dissection of the fibrillin-1
N-C terminal interaction

5.1 Introduction

Although the minimal domains required for the fibrillin N-C terminal interaction have been identified, there is no detailed molecular information about how the two regions of the protein interact. It was demonstrated previously that prokaryotically-produced and refolded N- and C-terminal fragments interact *in vitro* (Chapter 3). Unlike in other studies using larger fibrillin fragments (Lin *et al.*, 2002; Marson *et al.*, 2005), the interaction did not appear to be dependent on the presence of calcium. NMR spectroscopy was used to determine the structure of the N-terminal FUN-EGF3 fragment and to provide preliminary insights into its dynamics (Chapter 4). The structure is an important prerequisite for understanding how FUN-EGF3 interacts with the C-terminal cbEGF41-43 fragment in detail.

Experiments were designed to probe the N-C interaction using the FUN-EGF3 structure and other data. Several techniques were employed in attempts to map the binding site of cbEGF41-43 on the FUN-EGF3 fragment (illustrated in Figure 5.1). They were divided into two separate approaches. The first was an investigation of the minimal domains within FUN-EGF3 needed for binding. Smaller fragments were designed and produced, based on the FUN-EGF3 structure and exon boundaries of the *FBN1* gene.

In the second approach, residues in FUN-EGF3 potentially involved in the interaction were identified and targeted by site-directed mutagenesis. Key residues involved in the interaction between human Notch-1 and Jagged-1 were previously identified on the basis of NMR titration data and sequence conservation. Single residue substitutions, such as F207A in Jagged-1, were sufficient to abrogate binding of fragments in a pull-down assay (Cordle *et al.*, 2008a). A similar method was followed here for the fibrillin N-C interaction. The availability of NMR resonance assignments for the FUN-EGF3 fragment (Chapter 4) meant that NMR titrations could potentially

be analysed in terms of individual residues. Moreover, the fold of FUN-EGF3 proteins with single residue substitutions could be characterised using 1D and 2D ^{1}H NMR experiments.

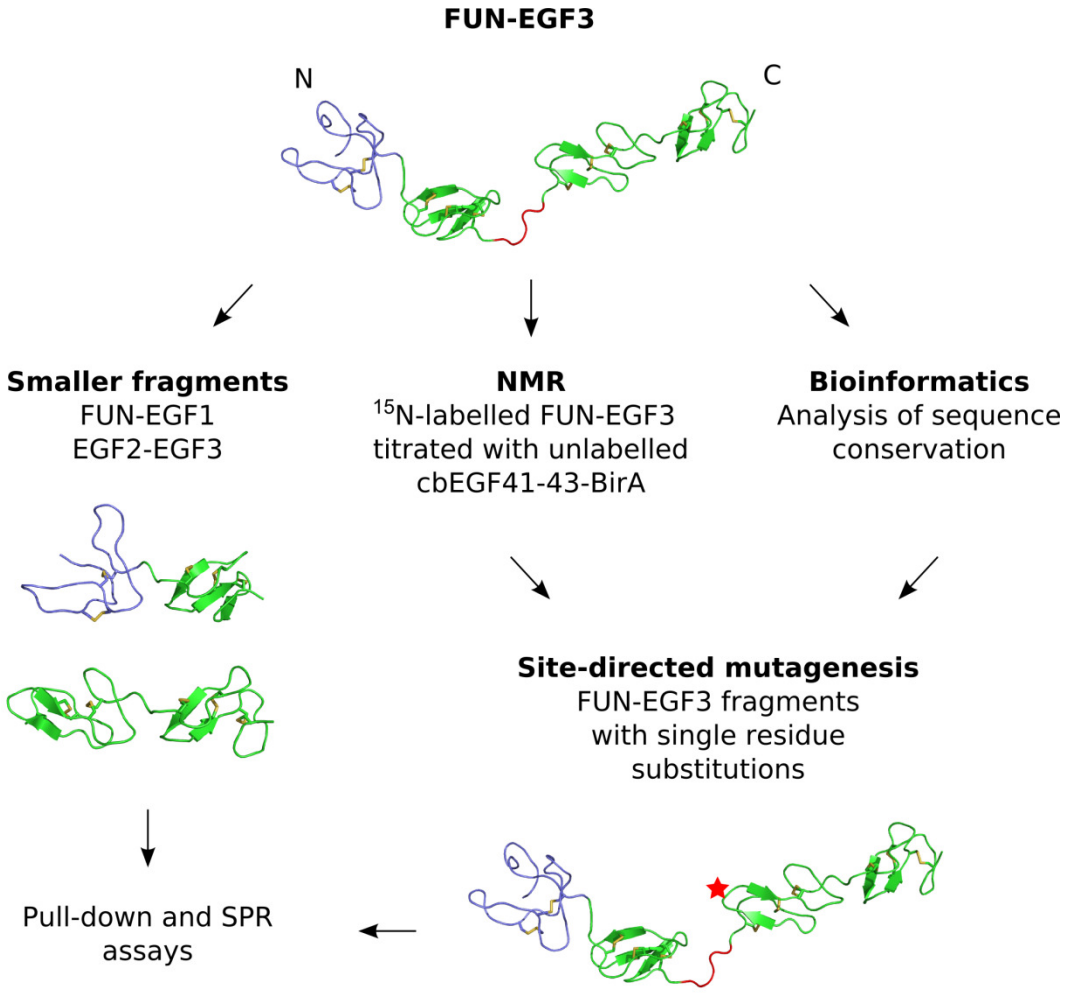


Figure 5.1 Approach to dissecting the binding site of cbEGF41-43 on the FUN-EGF3 fragment. Smaller N-terminal fragments, FUN-EGF1 and EGF2-EGF3, were produced, characterised and tested in interaction assays. NMR titrations were performed, with the aim of identifying residues in the FUN-EGF3 fragment involved in binding to cbEGF41-43. Sequence alignments identified conserved residues that could be important for the interaction. Using this information, FUN-EGF3 fragments with single alanine substitutions were produced and tested in interaction assays (single residue change indicated by red star).

5.2 Production of FUN-EGF1 and EGF2-EGF3 fragments

Two smaller N-terminal fragments were made using the prokaryotic expression, purification and *in vitro* refolding system also used to produce the FUN-EGF3 and cbEGF41-43-BirA fragments (Section 2.2; Section 3.4). As the FUN-EGF1 and EGF2-EGF3 regions appear to form independent structural units in FUN-EGF3 (Section 4.11), it was predicted that fragments comprising these regions would refold independently *in vitro*. FUN-EGF1 and EGF2-EGF3 were defined as residues R45-G112 and S113-E178 respectively, based on exon boundaries in the *FBN1* gene (Pereira *et al.*, 1993). A pQE-30 plasmid encoding the FUN-EGF1 fragment was kindly provided by Ian Robertson.

5.2.1 Cloning of a construct encoding EGF2-EGF3

The DNA sequence encoding the EGF2-EGF3 domains was amplified from human fibrillin-1 cDNA by PCR (Section 2.1.1; primer sequences in Appendix B). For cloning of the construct encoding the non-BirA tagged form, *Bam*HI and *Hind*III restriction sites were introduced by the forward and reverse primers respectively. This allowed ligation of the digested fragment into the pQE-30 vector (Sections 2.1.6 and 3.4.1). DNA sequencing confirmed that the construct contained the expected nucleotide sequence (Section 2.1.11).

5.2.2 Expression, purification and refolding of FUN-EGF1

The FUN-EGF1 fragment was expressed, purified and refolded, as described previously for FUN-EGF3 (Section 3.4.2.1). As for FUN-EGF3, glycerol was required for efficient refolding (data not shown). Representative HPLC and FPLC traces from the purification are shown in Figure 5.2. His-tagged FUN-EGF1 (His-FUN-EGF1) and non-His-tagged FUN-EGF1 proteins were produced for use in the pull-down and SPR assays respectively, as for FUN-EGF3 (Section 3.4.2.1).

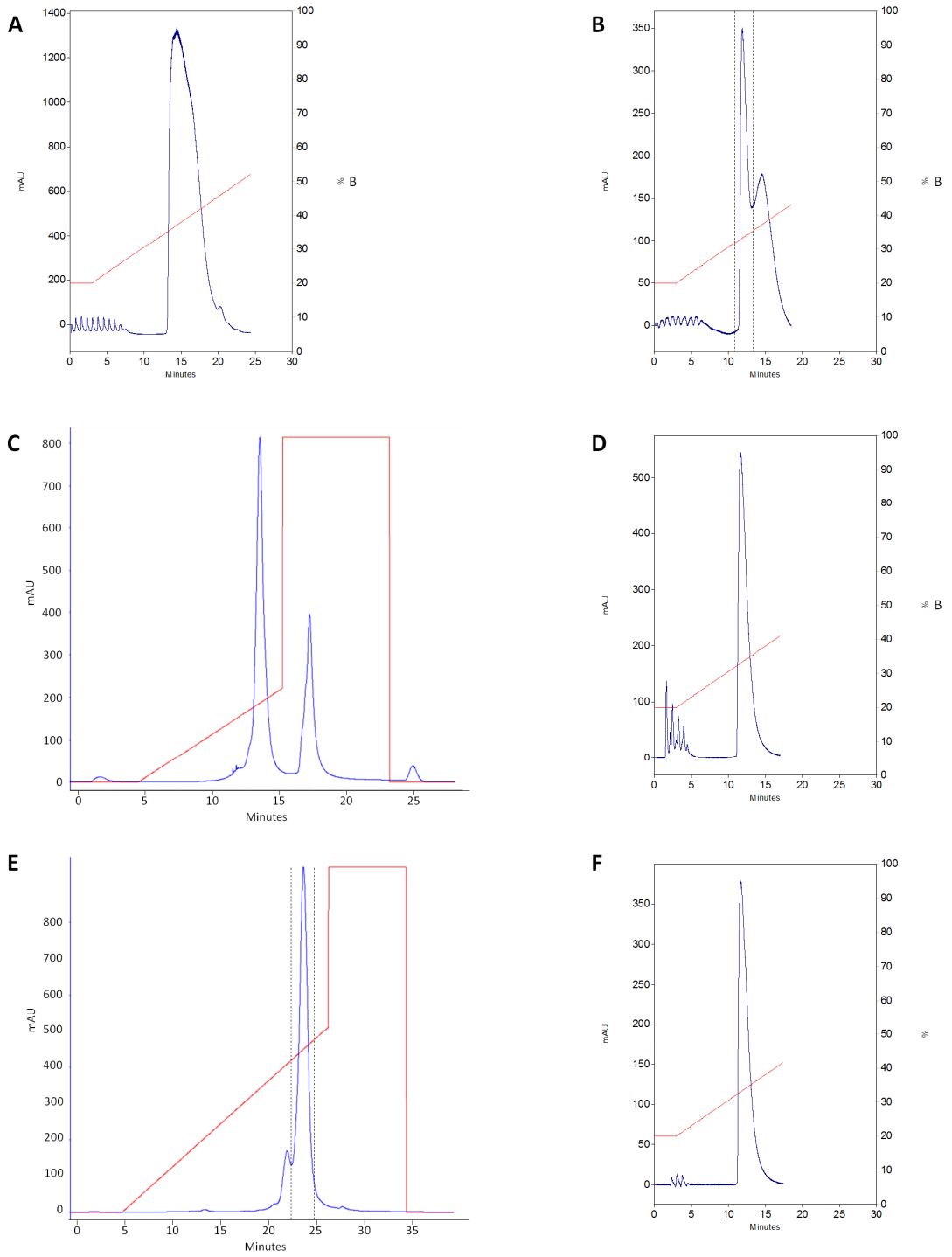


Figure 5.2 HPLC and FPLC chromatograms of FUN-EGF1. A) HPLC chromatogram of reduced FUN-EGF1. B) HPLC chromatogram of refolded protein. Dashed lines indicate the fraction boundaries of the putative natively-folded material. C) FPLC chromatogram of refolded FUN-EGF1 treated with factor Xa. D) Final HPLC chromatogram of FUN-EGF1 after FPLC purification. E) FPLC chromatogram of refolded His-FUN-EGF1. E) Final HPLC chromatogram of His-FUN-EGF1 after FPLC purification. Buffer conditions were the same as for FUN-EGF3 (Figure 3.5, Section 3.4.2.1).

5.2.3 SDS-PAGE and mass spectrometry analysis of FUN-EGF1 and His-FUN-EGF1

FUN-EGF1 and His-FUN-EGF1 final products were analysed by SDS-PAGE (Section 2.2.10) and mass spectrometry (Section 2.2.11). Proteins ran as diffuse bands in the SDS-PAGE gel when prepared under non-reducing sample conditions, but ran as sharper bands under reducing conditions (Figure 5.3). By contrast, FUN-EGF3 proteins ran as sharp bands under both conditions (Figure 3.6, Section 3.4.2.2). It was not clear why FUN-EGF3 and FUN-EGF1 behaved differently. Despite this, mass spectrometry showed that the masses of FUN-EGF1 and His-FUN-EGF1 were as expected (Table 5.1).

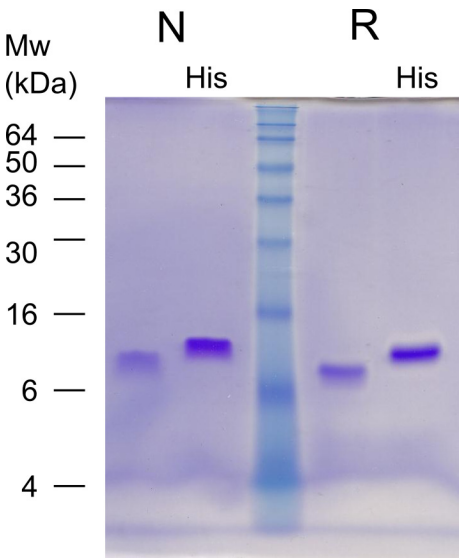


Figure 5.3 16% SDS-PAGE gel of FUN-EGF1 and His-FUN-EGF1. Non-reducing and reducing sample conditions (absence and presence of β-mercaptoethanol are denoted by 'N' and 'R' respectively). His-FUN-EGF1 is labelled 'His'.

Table 5.1 Mass spectrometry results for FUN-EGF1 and His-FUN-EGF1

Protein	Expected molecular weight (Da)	Measured molecular weight (Da)
FUN-EGF1	7123	7124
His-FUN-EGF1	8977	8978

5.2.4 Expression, purification and refolding of EGF2-EGF3: effect of a C-terminal BirA tag

The EGF2-EGF3 fragment was also expressed, purified and refolded. However, HPLC chromatograms of refolded EGF2-EGF3 suggested that its refolding efficiency was low (Figure 5.4A). The elution profile showed two major species, neither of which was present in significant quantities; two samples from independent refolding attempts had very similar profiles. In an attempt to improve the refolding efficiency of the EGF2-EGF3 region, a second construct encoding EGF2-EGF3 with a C-terminal BirA recognition peptide (EGF2-EGF3-BirA) was cloned. The addition of a C-terminal BirA tag to human Notch-1 cbEGF11-13 fragment was shown previously to increase the efficiency of refolding of the cbEGF13 domain (Cordle *et al.*, 2008a). Cloning of the pQE-30 EGF2-EGF3-BirA construct was carried out as described for cbEGF41-43-BirA and cbEGF32-34-BirA (Section 3.4.1). *Bam*HI and *Sal*I restrictions sites were introduced by the forward and reverse primers respectively (primer sequences in Appendix B).

In contrast to the non-BirA-tagged form, the HPLC elution profile of refolded EGF2-EGF3-BirA fragment was similar to that of FUN-EGF3 (Figure 5.4B; Figure 3.5, Section 3.4.2.1). Therefore, EGF2-EGF3-BirA was used in subsequent interaction work instead of EGF2-EGF3. Despite the previously observed effect of the C-terminal BirA tag on the fold of the Notch cbEGF11-13 fragment, it was surprising in this case, as FUN-EGF3 was successfully refolded. Refolded EGF2-EGF3-BirA was purified further by cation exchange FPLC and a final HPLC step (Figure 5.4C-D). His-tagged protein (His-EGF2-EGF3-BirA) was also produced for use in the pull-down assay.

.

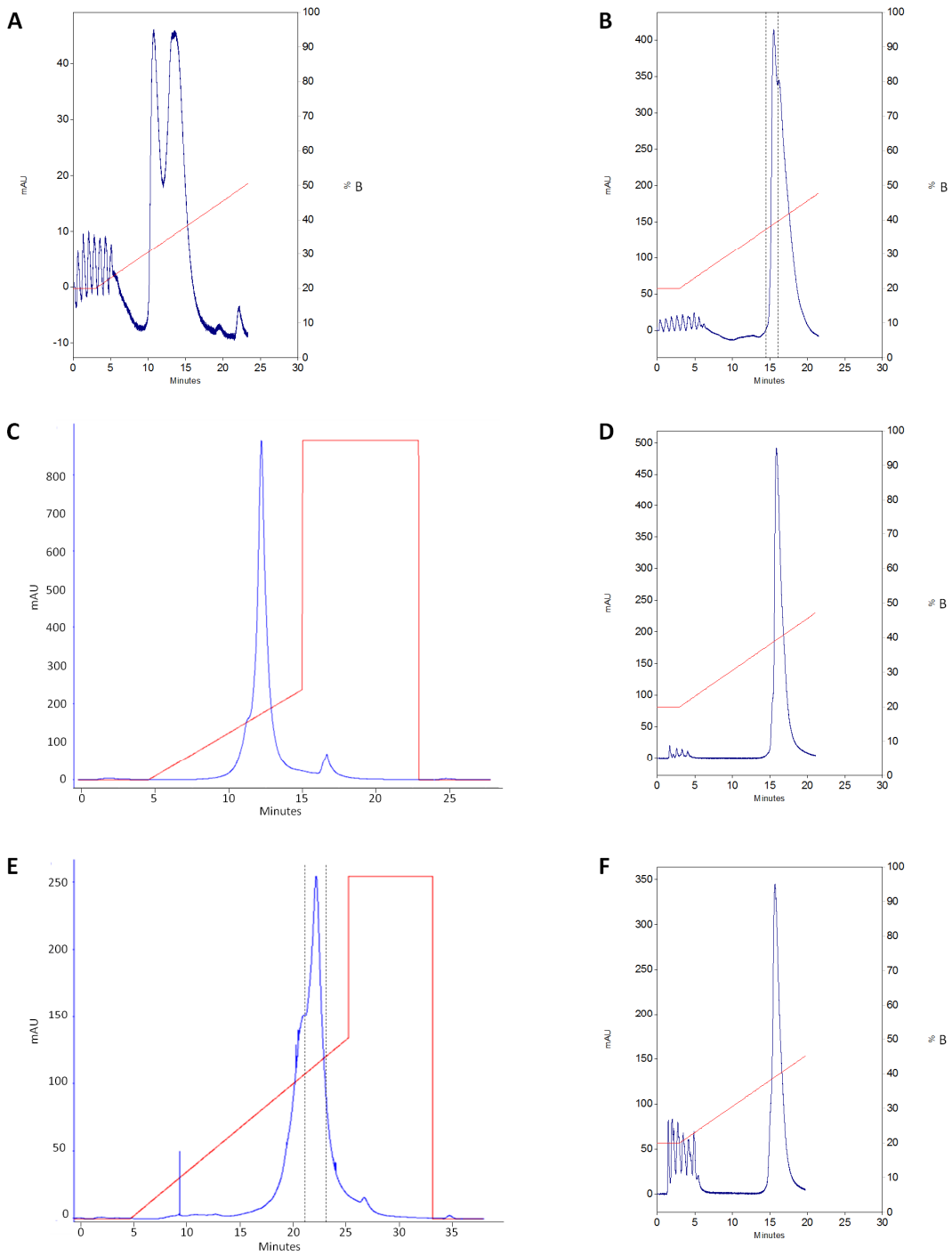


Figure 5.4 HPLC and FPLC chromatograms of EGF2-EGF3 and EGF2-EGF3-BirA. A) HPLC chromatogram of refolded EGF2-EGF3, showing a heterogeneous elution profile and low yield (judged by the low absorbance of light with wavelength 280 nm). B) HPLC chromatogram of refolded EGF2-EGF3-BirA. Dashed lines indicate the fraction boundaries of the putative natively-folded material. C) FPLC chromatogram of refolded EGF2-EGF3-BirA treated with factor Xa. D) Final HPLC chromatogram of EGF2-EGF3-BirA. E) FPLC chromatogram of His-EGF2-EGF3-BirA. F) Final HPLC chromatogram of His-EGF2-EGF3-BirA. Buffer conditions were the same as for FUN-EGF3 (Figure 3.5, Section 3.4.2.1).

5.2.5 SDS-PAGE and mass spectrometry analysis of EGF2-EGF3-BirA and His-EGF2-EGF3-BirA

EGF2-EGF3-BirA and His-EGF2-EGF3-BirA final products were also analysed by SDS-PAGE and mass spectrometry. A minor species with higher mobility was observed in the EGF2-EGF3-BirA sample under non-reducing conditions on the SDS-PAGE gel. The origin of this apparent heterogeneity was not clear. Mass spectrometry confirmed the expected sizes of the fragments (Table 5.2).

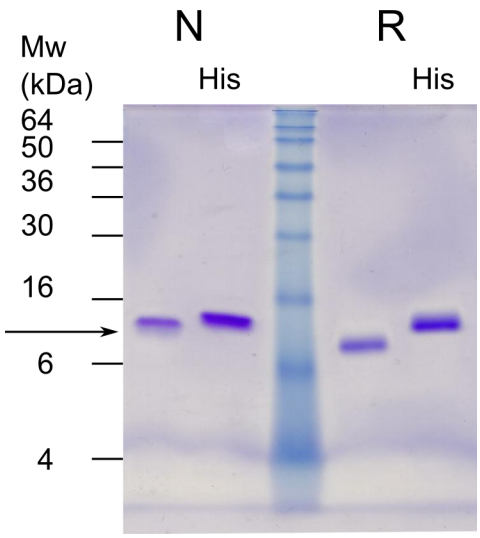


Figure 5.5 16% SDS-PAGE gel of EGF2-EGF3-BirA and His-EGF2-EGF-BirA. Non-reducing and reducing sample conditions (absence and presence of β-mercaptoethanol are denoted by 'N' and 'R' respectively). His-EGF2-EGF3-BirA is labelled 'His'. The arrow indicates the minor species in the non-reducing EGF2-EGF3-BirA sample.

Table 5.2 Mass spectrometry results for EGF2-EGF3-BirA and His-EGF2-EGF3-BirA

Protein	Expected molecular weight (Da)	Measured molecular weight (Da)
EGF2-EGF3-BirA	9229	9226
His-EGF2-EGF3-BirA	11083	11083

5.3 Characterisation of FUN-EGF1 and EGF2-EGF3-BirA by NMR spectroscopy

The FUN-EGF1 and EGF2-EGF3-BirA fragments were compared to FUN-EGF3 using NMR, in order to confirm that they adopted the same fold. Both fragments were labelled with ^{15}N, as described for FUN-EGF3 (Sections 2.2.2 and 3.5). Mass spectrometry confirmed labelling (Table 5.3).

Table 5.3 Mass spectrometry results for ^{15}N-labelled FUN-EGF1 and EGF2-EGF3-BirA

Protein	Expected molecular weight (Da)	Measured molecular weight (Da)
FUN-EGF1	7127	7128
EGF2-EGF3-BirA	9354	9354

^{1}H-^{15}N HSQC spectra were acquired for both labelled proteins under the same conditions as for FUN-EGF3. Comparing the spectra of FUN-EGF1 and FUN-EGF3 indicated that the FUN-EGF1 fragment was correctly folded (Figure 5.6A). Peaks assigned to the FUN domain in FUN-EGF3 coincided with peaks in the FUN-EGF1 fragment (Figure 5.6B). However, peaks assigned to residues in proximity to G112, the C-terminal residue of the EGF1 domain, were different. This is illustrated on the FUN-EGF1 structure (Figure 5.7). The C111 and G112 $^{1}H^{N}$-^{15}N peaks in the FUN-EGF3 spectrum were both absent from the FUN-EGF1 spectrum, but there were two additional peaks at different locations. Moreover, several backbone $^{1}H^{N}$-^{15}N peaks were slightly shifted in the FUN-EGF1 spectrum compared with FUN-EGF3 − for example, the H87 $^{1}H^{N}$-^{15}N peak (Figure 5.6B). These discrepancies can mostly likely be attributed to the absence of EGF1-EGF2 linker sequence at the C-terminus of FUN-EGF1. However, an interaction with the EGF2 domain cannot be ruled out. The HSQC spectrum of a fragment comprising the FUN-EGF1 region and the EGF1-EGF2 linker sequence would be needed to resolve the origin of these differences.

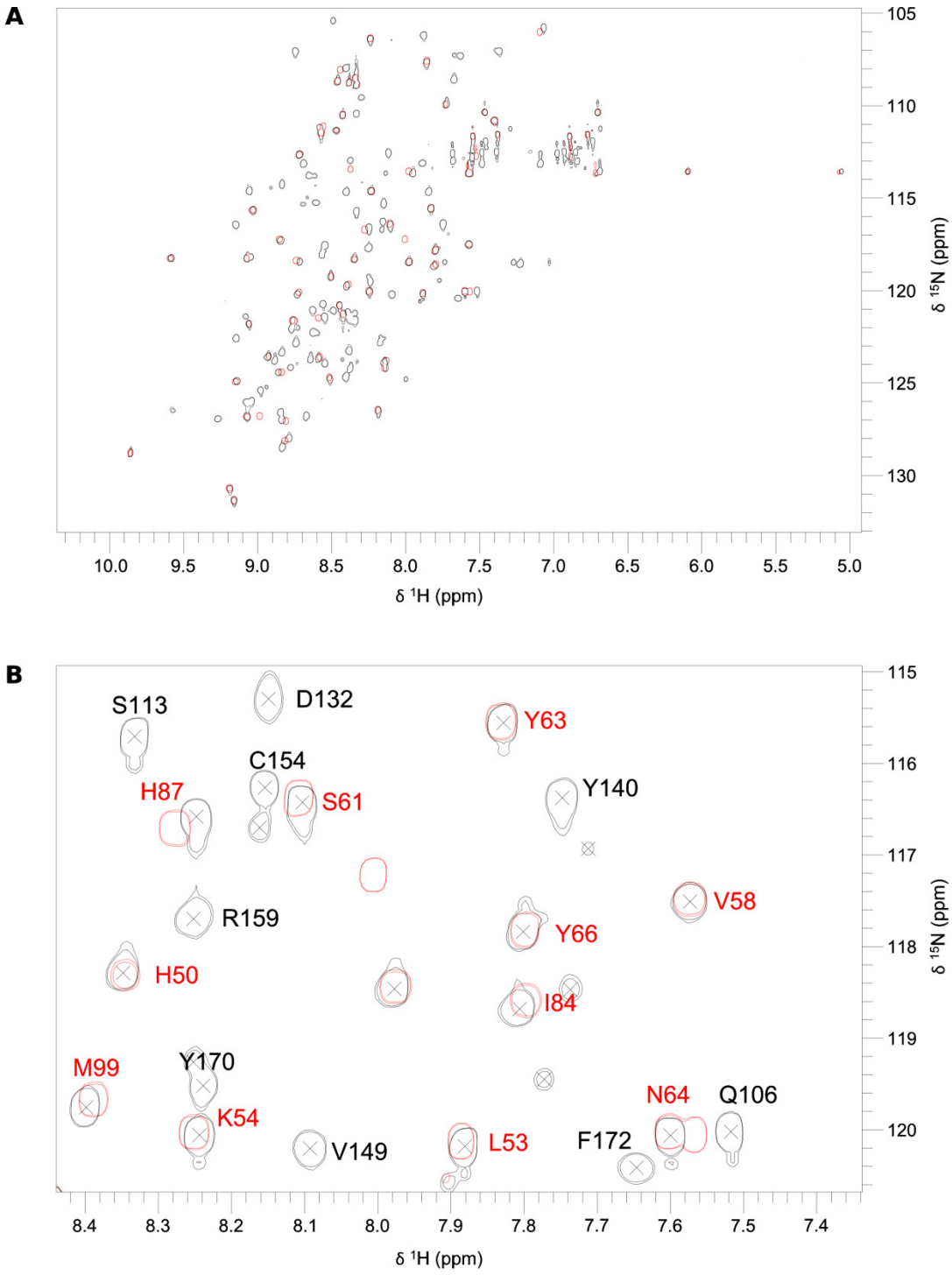


Figure 5.6 ^{1}H-^{15}N HSQC spectra of FUN-EGF1 and FUN-EGF3. A) Contours of ^{15}N-labelled FUN-EGF1 spectrum (red) plotted on the same axes as the spectrum of ^{15}N-labelled FUN-EGF3 (black). The FUN-EGF1 spectrum was acquired with 0.3 mM protein in 5% D_2O/95% H_2O (v/v) at pH 5.4 and 25°C. The FUN-EGF3 spectrum is the same as in Figure 3.9B, Section 3.5. B) Zoomed-in region of the HSQC spectra. Residue assignments are given for backbone $^{1}H^{N}$-^{15}N peaks, where assignments are unambiguous. Labels for peaks for residues in the FUN-EGF1 region of FUN-EGF3 are shown in red.

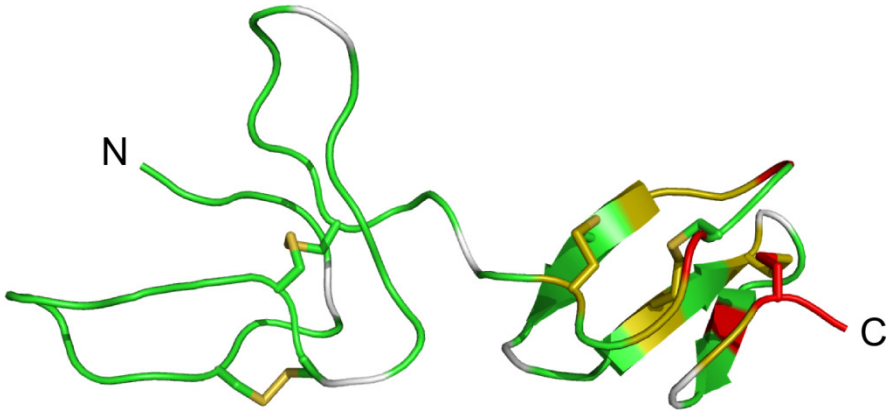


Figure 5.7 Differences in $^{1}H^{N}$ and ^{15}N chemical shifts between FUN-EGF1 and the FUN-EGF1 region of FUN-EGF3. The backbone FUN-EGF1 pair structure (Chapter 4) is coloured according to the combined chemical shift difference (Δ_{ppm}) between equivalent peaks in the FUN-EGF3 and FUN-EGF1 HSQC spectra. $\Delta_{ppm}=\left((HN\Delta_{ppm})^2+(0.17\times N\Delta_{ppm})^2\right)^{1/2}$, where $HN\Delta_{ppm}$ and $N\Delta_{ppm}$ are the chemical shift differences in the ^{1}H and ^{15}N dimensions respectively (Farmer *et al.*, 1996). Green indicates that Δ_{ppm} between the two spectra was less than 0.03 ppm. Yellow indicates that Δppm was between 0.03 and 0.05 ppm, while red indicates that Δ_{ppm} was greater than 0.05 pm or that the peaks were absent. Proline residues are coloured white.

It was less straightforward to compare the HSQC spectra of EGF2-EGF3-BirA and FUN-EGF3, as EGF2-EGF-BirA had a C-terminal BirA tag. Many of the peaks in the EGF2-EGF3-BirA spectrum coincided with those in the FUN-EGF3 spectrum (Figure 5.8A). These included peaks assigned to residues in the EGF2 domain, for example S128, D132 and H133, as well as residues at the EGF2-EGF3 interdomain interface, such as G139 and Y140 (Figure 5.8B). This comparison indicated that the fold of the EGF2 domain and interdomain interface was very similar to the FUN-EGF3 fragment. However, peaks assigned to residues in the EGF3 domain, for example S152, R159 and Q176, appeared to be shifted or absent from the EGF2-EGF3-BirA spectrum compared with FUN-EGF3. These residues are likely to be close to the BirA tag. Correct folding of the EGF3 domain was confirmed by the 2D ^{1}H homonuclear NOESY spectrum of EGF2-EGF3-BirA. C160-C166 and N164-C150 $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peaks, which are markers for folding of the EGF3 domain in the FUN-EGF3 NOESY, were also present in the EGF2-EGF3-BirA NOESY spectrum (data not shown).

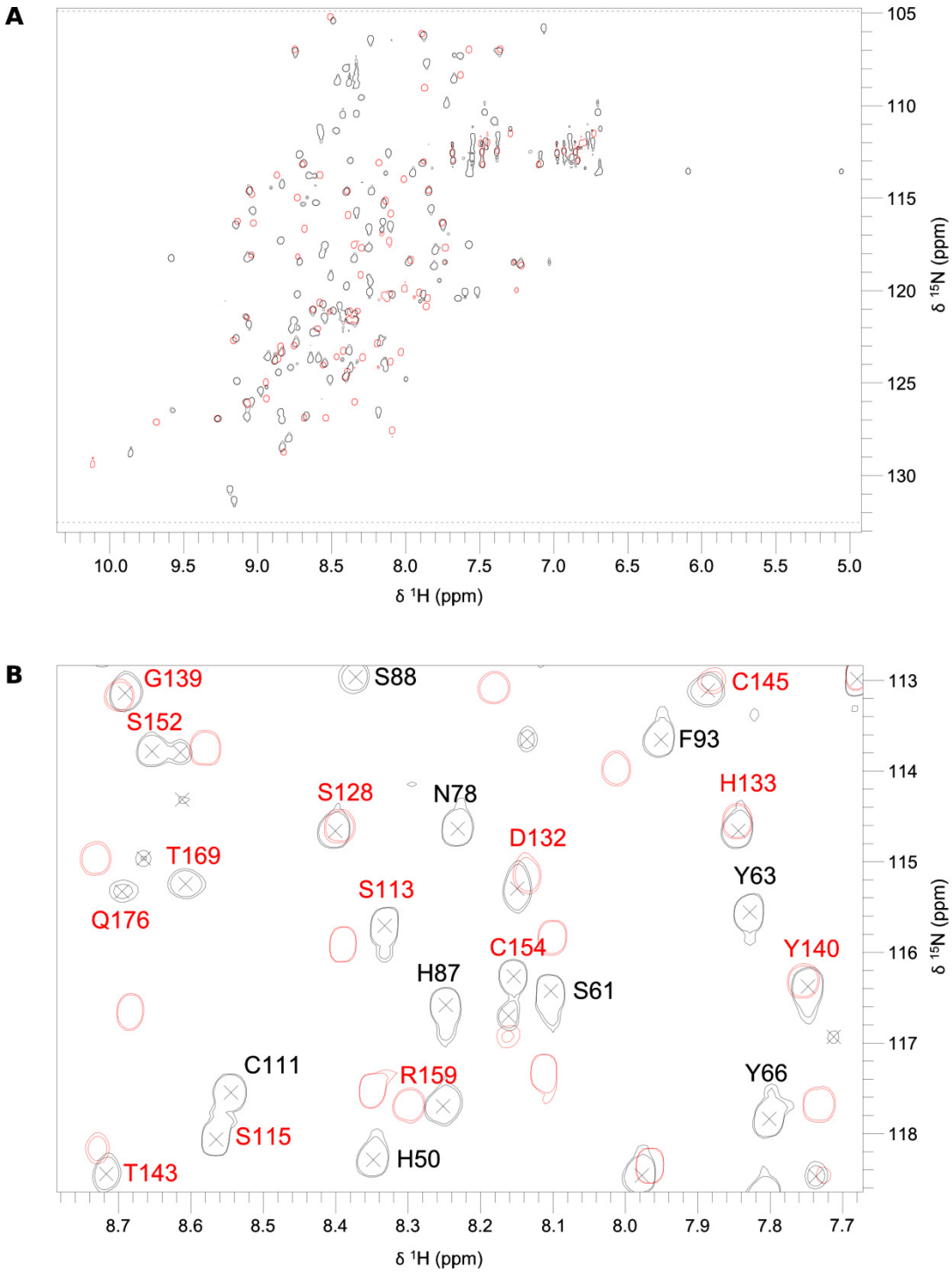


Figure 5.8 ^{1}H-^{15}N HSQC spectra of EGF2-EGF-BirA and FUN-EGF3. A) Spectra of ^{15}N-labelled EGF2-EGF3-BirA (red) and ^{15}N-labelled FUN-EGF3 (black) plotted on the same axes. The EGF2-EGF3-BirA spectrum was acquired with 0.3 mM protein in 5% D_2O/95% H_2O (v/v) at pH 5.4 at 25°C. The FUN-EGF3 spectrum is the same as in Figure 3.9B, Section 3.5. B) Zoomed-in region of the HSQC spectra. Residue assignments are given for backbone $^{1}H^{N}$-^{15}N cross-peaks, where assignments are unambiguous. Labels for peaks for residues in the EGF2-EGF3 region of FUN-EGF3 are shown in red.

5.4 Interaction of FUN-EGF1 and EGF2-EGF3-BirA fragments with cbEGF41-43-BirA

Interactions of the FUN-EGF1 and EGF2-EGF3-BirA fragments with cbEGF41-43-BirA were tested in pull-down and SPR assays. These assays had been previously established for the interaction between the FUN-EGF3 and cbEGF41-43-BirA fragments (Section 3.7).

5.4.1 Pull-down assay

In initial pull-down experiments, 15 μg of the His-tagged protein was incubated with streptavidin-coated magnetic beads. As expected, the His-FUN-EGF3 fragment was pulled down by beads containing cbEGF41-43-BirA-bio, but not by the beads alone, or by beads carrying immobilised negative control human Notch-1 cbEGF11-13-BirA-bio fragment, kindly provided by Paul Taylor (Figure 5.9A). However, bead-immobilised cbEGF41-43-BirA-bio appeared not to pull down His-FUN-EGF1 or His-EGF2-EGF3-BirA. The His-FUN-EGF1 fragment was readily detected by the anti-His-tag antibody, but more His-EGF2-EGF3-BirA had to be loaded in order to produce an equivalent signal. Pull-downs were then performed with larger quantities of His-tagged protein (Figure 5.9B). In this experiment, His-FUN-EGF1 and His-EGF2-EGF3-BirA bound to bead-immobilised cbEGF41-43-BirA-bio, but not to the beads alone. These two results suggested that both the FUN-EGF1 and EGF2-EGF3 regions in FUN-EGF3 might be required for binding of FUN-EGF3 to cbEGF41-43-BirA.

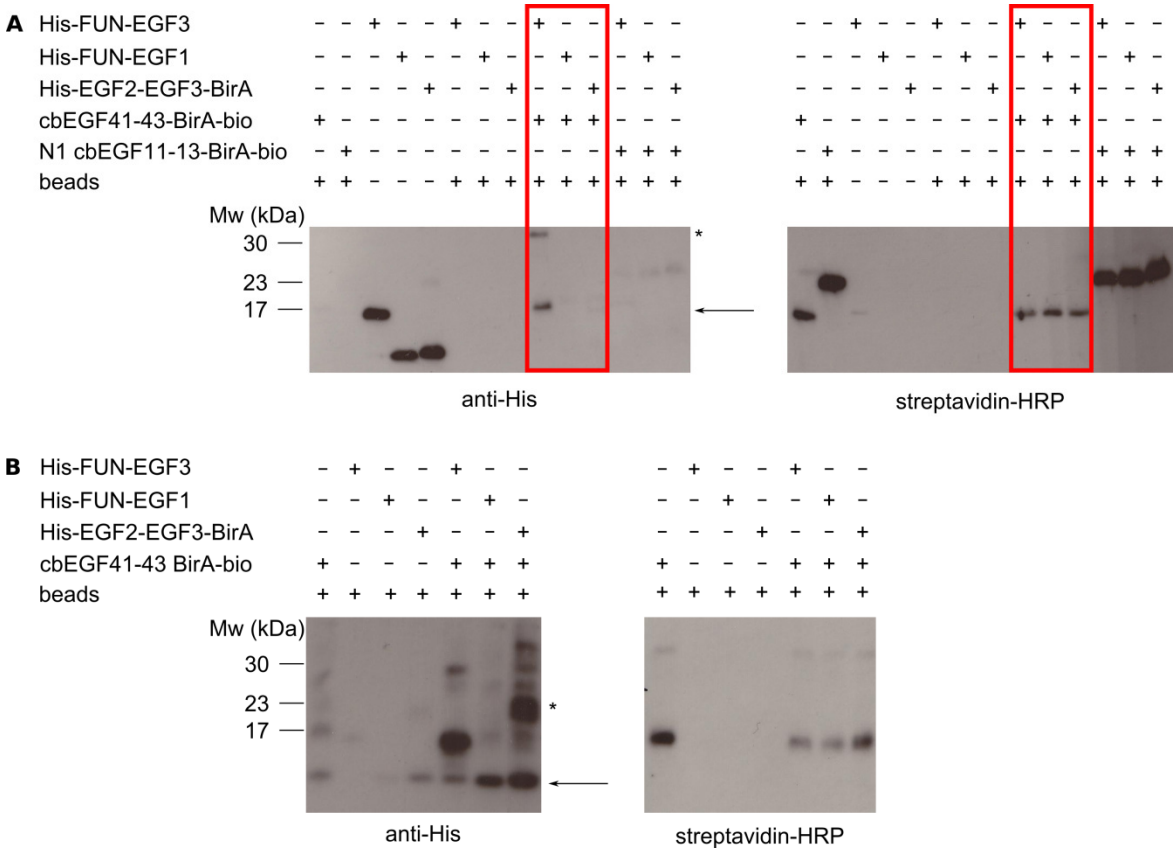


Figure 5.9 Pull-down assay using immobilised cbEGF41-43-BirA-bio and three N-terminal fragments in solution. A) Western blots from the pull-downs of His-FUN-EGF3, His-FUN-EGF1 and His-EGF2-EGF3-BirA. Conditions were the same as for the pull-down with His-FUN-EGF3 and immobilised cbEGF41-43-BirA-bio (Section 3.7.1). All three His-tagged N-terminal fragments were also loaded separately to demonstrate that they could be detected by the anti-His antibody. More His-EGF2-EGF3-BirA was loaded as the detection efficiency was lower: 500 ng compared with 150 ng of FUN-EGF3 and FUN-EGF1. None of the fragments bound non-specifically to the streptavidin-coated beads or an immobilised Notch-1 (N1) cbEGF11-13-BirA-bio negative control fragment. Red boxes indicate lanes containing samples from the pull-downs with cbEGF41-43-BirA-bio. The arrow and * indicate the monomeric FUN-EGF3 band and higher molecular weight species respectively. B) Results from the pull-down using a larger quantity of His-tagged protein (40 µg instead of 15 µg). The arrow and * indicate the monomeric EGF2-EGF3-BirA band and higher molecular weight species respectively, as for FUN-EGF3.

5.4.2 SPR assay

The FUN-EGF1 and EGF2-EGF3-BirA fragments behaved differently from FUN-EGF3 in the SPR assay. As the response level in SPR is directly proportional to the molecular weight of the interacting ligands, FUN-EGF1 and EGF2-EGF3-BirA could not be compared directly with FUN-EGF3. Nonetheless, even accounting for the larger size of FUN-EGF3 (approximately double the molecular weight of both FUN-EGF1 and EGF2-EGF3-BirA), the response levels were much lower for the smaller fragments (Figure 5.10A-B). Moreover, the sensogram profiles were different – there were sudden increases and decreases in response level at the start and end of the analyte injections. Comparisons of the flow cells containing coupled cbEGF41-43-BirA and control protein cbEGF32-34-BirA showed little difference in response upon injection of FUN-EGF1 at a series of concentrations (Figure 5.10C-D). There was a slightly higher response from the cbEGF41-43-BirA flow cell upon injection of EGF2-EGF3-BirA compared with the cbEGF32-34-BirA flow cell (Figure 5.10E-F). It was not clear whether these results were representative of specific binding to cbEGF41-43-BirA. Therefore, it was not clear whether both halves of FUN-EGF3 bound to cbEGF41-43-BirA.

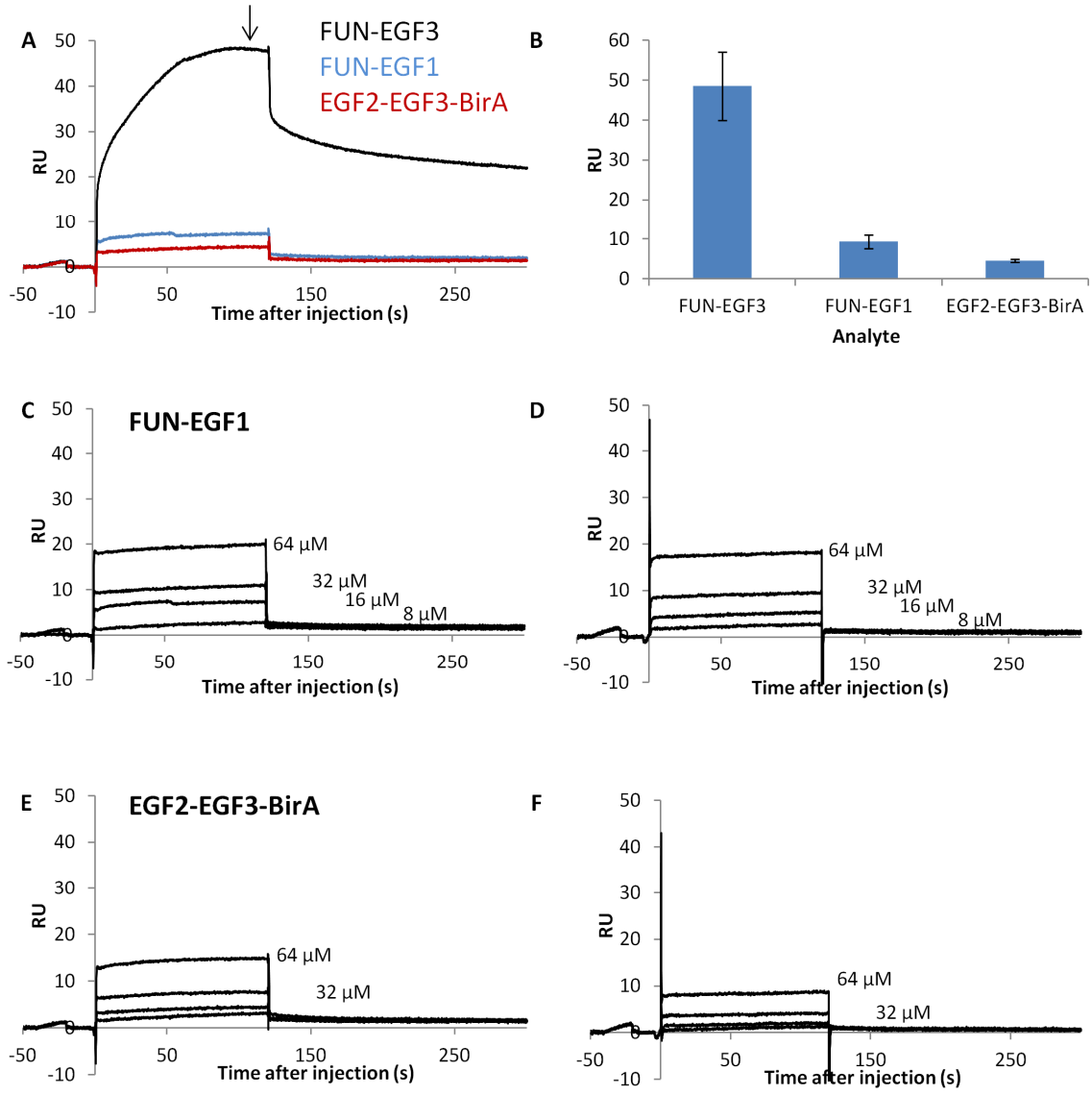


Figure 5.10 Results of the SPR interaction assay for surface-coupled cbEGF41-43-BirA with N-terminal fragments in solution. A) Representative blank-subtracted sensorgrams for FUN-EGF3, FUN-EGF1 and EGF2-EGF3-BirA for the flow cell with amine-coupled cbEGF41-43-BirA. The analyte concentration was 16 µM in each case. The conditions were as described for the FUN-EGF3 (Section 3.7.2). Results shown were from the same Biacore chip, with data acquired on the same day. B) Response levels 110 s after the start of 16 µM analyte injection (time point in sensorgram indicated by arrow in Part A). n= 4, with two replicates for two different preparations of each protein. Error bars show the standard deviation. C) Sensorgrams for the flow cell with immobilised cbEGF41-43-BirA and FUN-EGF1 as the analyte. D) FUN-EGF1 as the analyte and control protein cbEGF32-34-BirA as the ligand. E) EGF2-EGF3-BirA as the analyte and cbEGF41-43-BirA as the ligand. F) EGF2-EGF3-BirA as the analyte and control protein cbEGF32-34-BirA as the ligand. Multiple concentrations were tested for both analytes: 0, 8, 16, 32 and 64 µM.

5.5 NMR titration: ^{15}N FUN-EGF3 with unlabelled cbEGF41-43-BirA

In an attempt to map the binding surface of the fibrillin C-terminal domains on FUN-EGF3, ^{15}N-labelled FUN-EGF3 was titrated with unlabelled cbEGF41-43-BirA (Section 2.6.11). The low solubility of the cbEGF41-43-BirA fragment limited the conditions that could be used in the experiment. Given the apparent calcium-independence of the interaction between FUN-EGF3 and cbEGF41-43-BirA (Section 3.7.1), $CaCl_2$ was not added to the protein solution during the titration. This allowed the titration to be performed at pH 7.5 without significant protein precipitation. A ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled FUN-EGF3 was first acquired at pH 7.5 (Figure 5.11A). Cross-peak assignments were obtained by comparing the positions of peaks in spectra of FUN-EGF3 at pH 5.4, pH 6.0, pH 6.5 and pH 7.5 (data not shown), as assignments were already available for the spectrum acquired at pH 5.4.

^{15}N-labelled FUN-EGF3 was then titrated with lyophilised cbEGF41-43-BirA. A spectrum was acquired after each addition of cbEGF41-43-BirA. Very few peaks shifted (Figure 5.11A) and the intensities of all peaks decreased with every addition of cbEGF41-43-BirA. Once the concentration of cbEGF41-43-BirA had reached 0.25 mM, few peaks were visible in the HSQC spectrum (Figure 5.11B). These peaks had the largest intensities in the spectrum of FUN-EGF3 without added cbEGF41-43-BirA. The titration therefore did not yield any useful information about the interaction between FUN-EGF3 and cbEGF41-43-BirA. Titrations using ^{15}N-labelled FUN-EGF1 and EGF2-EGF3-BirA yielded similar results (data not shown). It was concluded that this was not a viable approach for mapping the binding site.

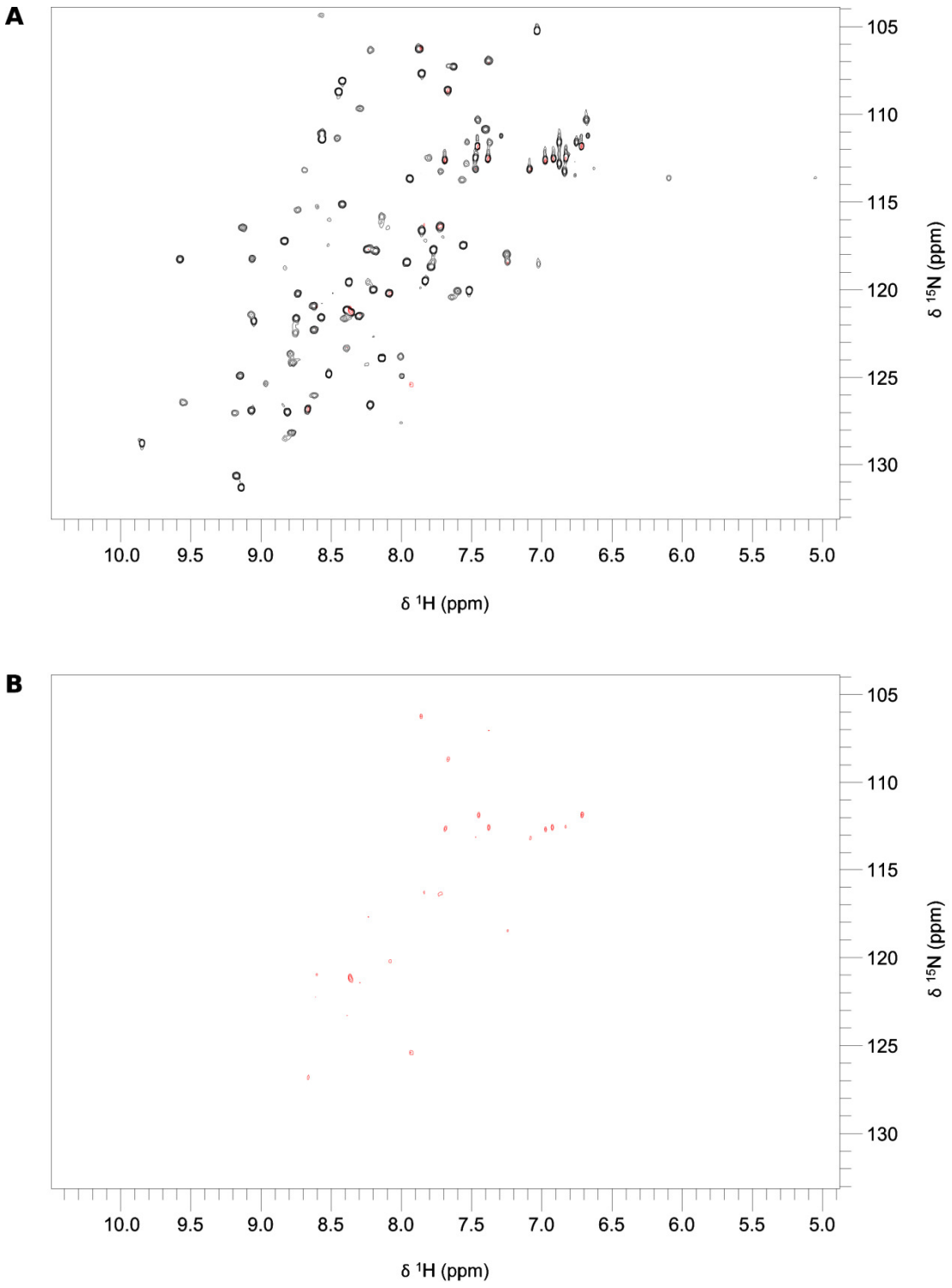


Figure 5.11 ^{1}H-^{15}N HSQC spectra from NMR titration of ^{15}N-labelled FUN-EGF3 with unlabelled cbEGF41-43-BirA. A) Spectra of 0.3 mM ^{15}N-labelled FUN-EGF3 in 5 mM Tris pH 7.5 5% D_2O/95% H_2O (v/v) with (red) and without (black) cbEGF41-43-BirA added to a concentration of 0.25 mM. B) Spectrum of FUN-EGF3 with added cbEGF41-43-BirA in part A, showing the remaining visible cross-peaks.

5.6 Probing the N-C interaction by site-directed mutagenesis of FUN-EGF3

5.6.1 Insights from sequence conservation of FUN-EGF3

Sequence conservation could help to identify residues in FUN-EGF3 that form an interaction surface with cbEGF41-43. Alignments of FUN-EGF3 sequences in fibrillins from various species showed that many residues are conserved (Figure 5.12). It is very likely that a large proportion of the conserved residues are important for the structure of the FUN-EGF3 region. These include cysteines, residues found at β-turns, such as glycines and prolines, and residues involved in interdomain packing (Section 4.11). Although some of these may also form part of the interaction interface with cbEGF41-43, they may not be good targets for site-directed mutagenesis. Replacing one of these residues with another amino acid could disrupt the structure and therefore perturb *in vitro* refolding of the FUN-EGF3 fragment. However, there are other residues in the FUN-EGF3 region of human fibrillin-1 that are at least partially conserved in other species.

It is necessary to consider that several elastic fibre components bind the fibrillin N-terminal domains (Section 1.8) when looking at sequence conservation, although few interaction sites have been well defined. These components include the LTBPs, BMPs, fibulins, MAGPs and ADAMTS/ADAMTSL proteins. Conserved residues could contribute to the binding sites for a number of different molecules. The fibulins, LTBPs and MAGPs evolved later than the fibrillins and so are absent from some fibrillin-containing species, including the sea squirt *Ciona intestinalis*, the red flour beetle *Tribolium castaneum* and sea anenome *Nematostella vectensis* (Robertson *et al.*, 2011). It is therefore likely that fibrillins in these species lack the residues that form the binding sites for fibulins, LTBPs and MAGPs. Sequence alignments showed that some residues in fibrillin-1 are conserved across many species, or are replaced by amino acids with similar properties (Figure

5.11). These are candidates for residues involved in binding to the cbEGF41-43 region, although other roles cannot be excluded.

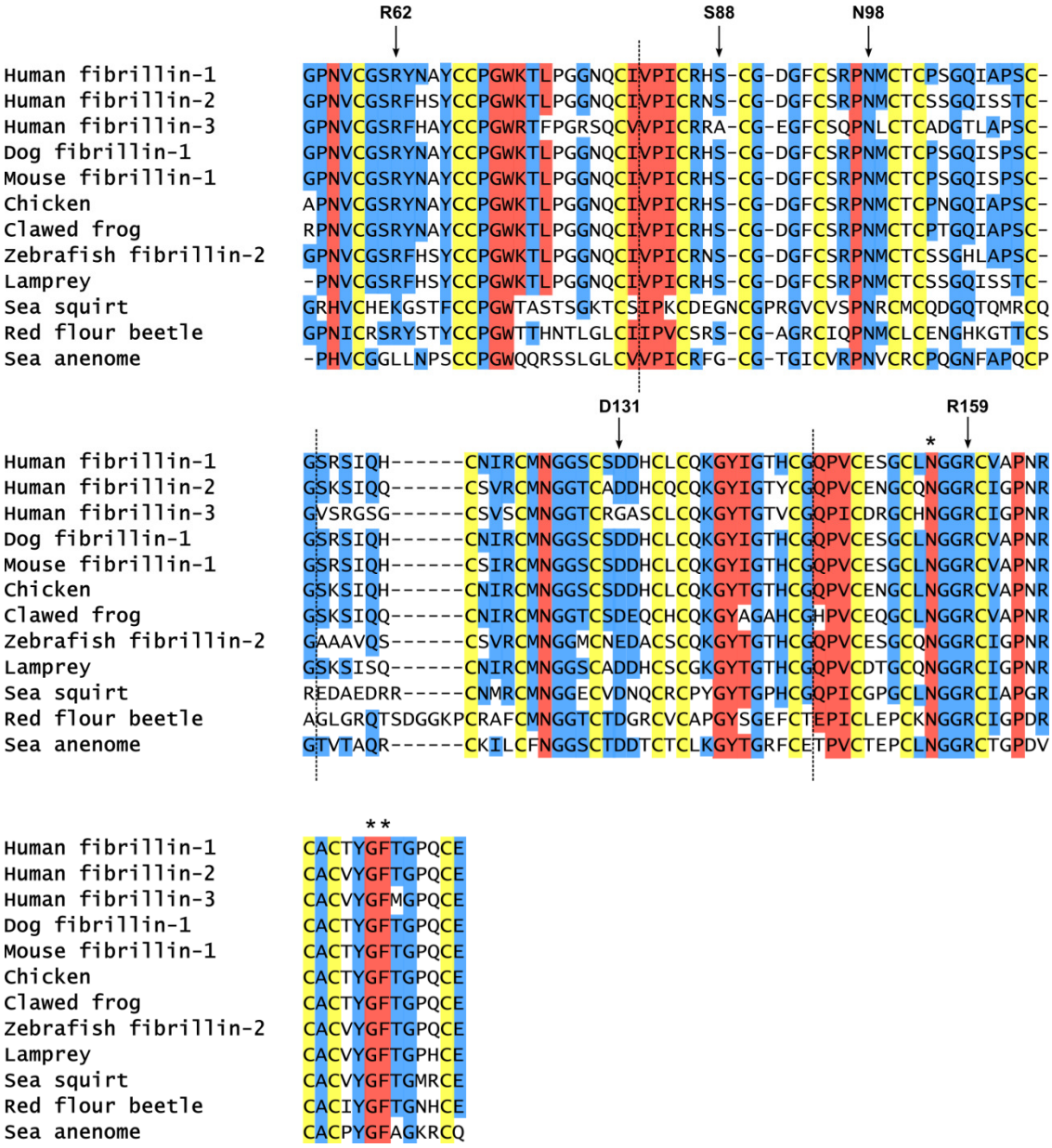


Figure 5.12 Conserved residues in the FUN-EGF3 region of fibrillins. Cysteine residues are highlighted in yellow and residues involved in interdomain packing are highlighted in red (as in Figure 4.20, Section 4.11). Residues in the EGF3 domain that pack against the hyb1 domain have not yet been well defined, but N156 and the G171-F172 motif (indicated by *) are also highlighted in red, according to their similarity to packing residues at the FUN-EGF1 and EGF2-EGF3 interdomain interfaces. Other conserved residues are highlighted in blue. Highlighted conserved residues are present in at least one of the three species lacking LTBPs, fibulins and MAGPs (sea squirt, red flour beetle and sea anenome). Arrows indicate residues targeted by site-directed mutagenesis in this study.

5.6.2 Design of FUN-EGF3 fragments with alanine substitutions

5 residues in FUN-EGF3 were selected for alanine substitution by site-directed mutagenesis (Figure 5.13). They were chosen as they are surface-exposed, unlikely to be structurally important and located at different sites in the fragment. R62 is conserved in all organisms except the sea anenome and is located on the flexible loop between C1 and C2 of the FUN domain (Figure 5.12). S88 lies between C1 and C2 of the EGF1 domain and is not fully conserved, but is present in the red flour beetle. The residue is replaced by alanine in human fibrillin-3, although this isoform lacks several of the residues that are present in many of the fibrillins.

N98 is located between C3 and C4 of the EGF1 domain and is absolutely conserved. N164, the residue at the equivalent position in the EGF3 domain, is not conserved in the three organisms lacking the LTBPs, consistent with its role in binding LTBP-4 in humans (Ono *et al.*, 2009). D131 lies between C3 and C4 of the EGF2 domain is conserved, with the exception of human fibrillin-3. Finally, R159, which is absolutely conserved, is the residue before C5 in the EGF3 domain. Constructs encoding FUN-EGF3 with single alanine substitutions (R62A, S88A, N98A, D131A and R159A) were cloned by Paul Evans (Paul Evans, Part II Project, 2012).

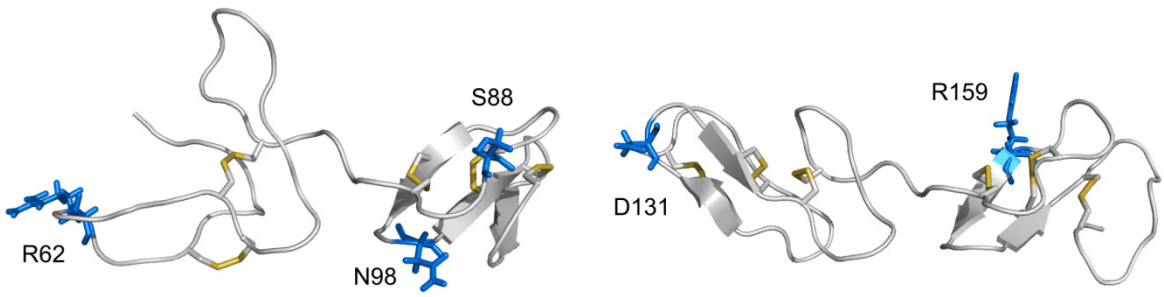


Figure 5.13 Residues in the FUN-EGF3 fragment targeted by site-directed mutagenesis. The EGF1-EGF2 linker sequence is omitted for clarity.

5.6.3 Production of FUN-EGF3 fragments with single alanine substitutions

The 5 FUN-EGF3 fragments with single alanine substitutions were produced using the same method as for the wildtype (WT) protein (Section 3.4.2.1). Protein purification and initial characterisation by SDS-PAGE and mass spectrometry were carried out by Paul Evans, confirming the correct sizes and masses of the proteins (Paul Evans, Part II project, 2012).

5.6.4 Use of 2D ^{1}H NMR spectroscopy in characterisation of FUN-EGF3 proteins

It was important to confirm that the substitutions did not significantly perturb the structure of the FUN-EGF3 fragment. Previous studies examining the effects of residue substitutions on the structure of fibrillin or Notch fragments have measured calcium-binding affinities (Cordle *et al.*, 2008b; Whiteman *et al.*, 2007), as calcium binding is dependent on the correct arrangement of disulphide bonds in cbEGF domains. Although the FUN-EGF3 fragment does not bind calcium ions, NMR spectroscopy can be used to compare the structure of the modified proteins to the WT protein. By analogy with calcium-dependent interfaces, the formation of interdomain interfaces in FUN-EGF3 is likely to be dependent on the presentation of packing residues in the correct orientation. Therefore, markers for the residues at the interdomain interfaces in NMR spectra can be used to study the effects of residue substitutions on the fold of FUN-EGF3.

2D ^{1}H homonuclear NOESY spectra of unlabelled protein (Section 2.6.2.2) provided 'fingerprints' of FUN-EGF3 structure. Using the FUN-EGF3 structure and ^{1}H resonance assignments (Chapter 4), several cross-peaks were identified in the spectrum of WT FUN-EGF3 that could be used as markers of correct folding (Figure 5.14). These cross-peaks were assigned to residues spread across the four domains (Figure 5.15). In particular, some cross-peaks were good indicators of secondary structure as they arose from interactions between $^{1}H^{\alpha}$ nuclei in residues on opposite

strands of β-hairpin structures. For example, there was a $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peak between V58 and C67, two residues in the FUN domain (Figures 5.14B and 5.15A). Moreover, a $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peak between packing residues W71 and P83 was a marker for the FUN-EGF1 interface (Figures 5.14B and 5.15B). For the EGF1 domain, there were $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peaks between C85 and N98. There was also a cross-peak between T101 and I107, located in the second and third strands of the EGF1 β-sheet respectively (Figures 5.14B and 5.15C).

There were few clear markers of folding for the EGF2 domain in the NOESY spectrum, most likely resulting from weak cross-peak intensities and peak overlap. However, a cross-peak between the $^{1}H^{\alpha}$ and $^{1}H^{\beta1}$/$^{1}H^{\beta2}$ resonances of residue D132, located in the loop between strands of the β-hairpin, was unambiguously assigned (not shown in spectrum; Figure 5.15D). Moreover, clear markers of the EGF2-EGF3 interface were identified, including a $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peak between Y140 and P148, analogous to the FUN-EGF1 interface marker (Figures 5.14B and 5.15E). Finally, there were two $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peak markers for the EGF3 domain: the first between C150 and N164, and the second between C160 and C166 in the β-hairpin structure (Figures 5.14B and 5.15F).

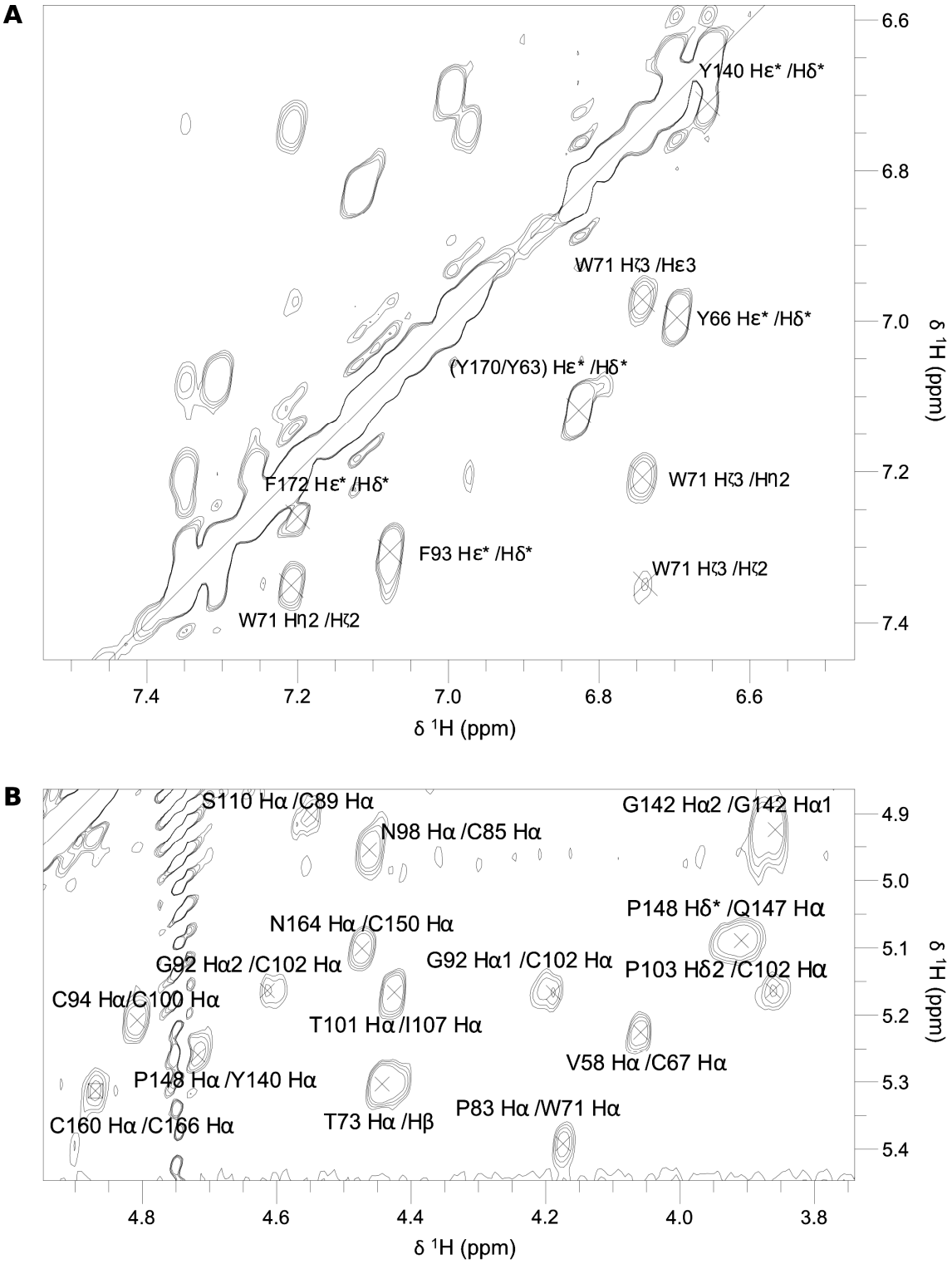


Figure 5.14 Regions of the 2D ^{1}H homonuclear NOESY spectrum of the FUN-EGF3 fragment used to identify correct folding. A) Region of the spectrum showing cross-peaks between aromatic side-chain resonances. B) Region of the spectrum showing predominantly cross-peaks between $^{1}H^{\alpha}$ resonances, including peaks downfield of the H_2O resonance. Downfield-shifted $^{1}H^{\alpha}$ chemical shifts are markers of β-sheet-like secondary structure. The spectrum was acquired using 0.5 mM FUN-EGF3 in 100% D_2O at pH 5.4 and 25°C.

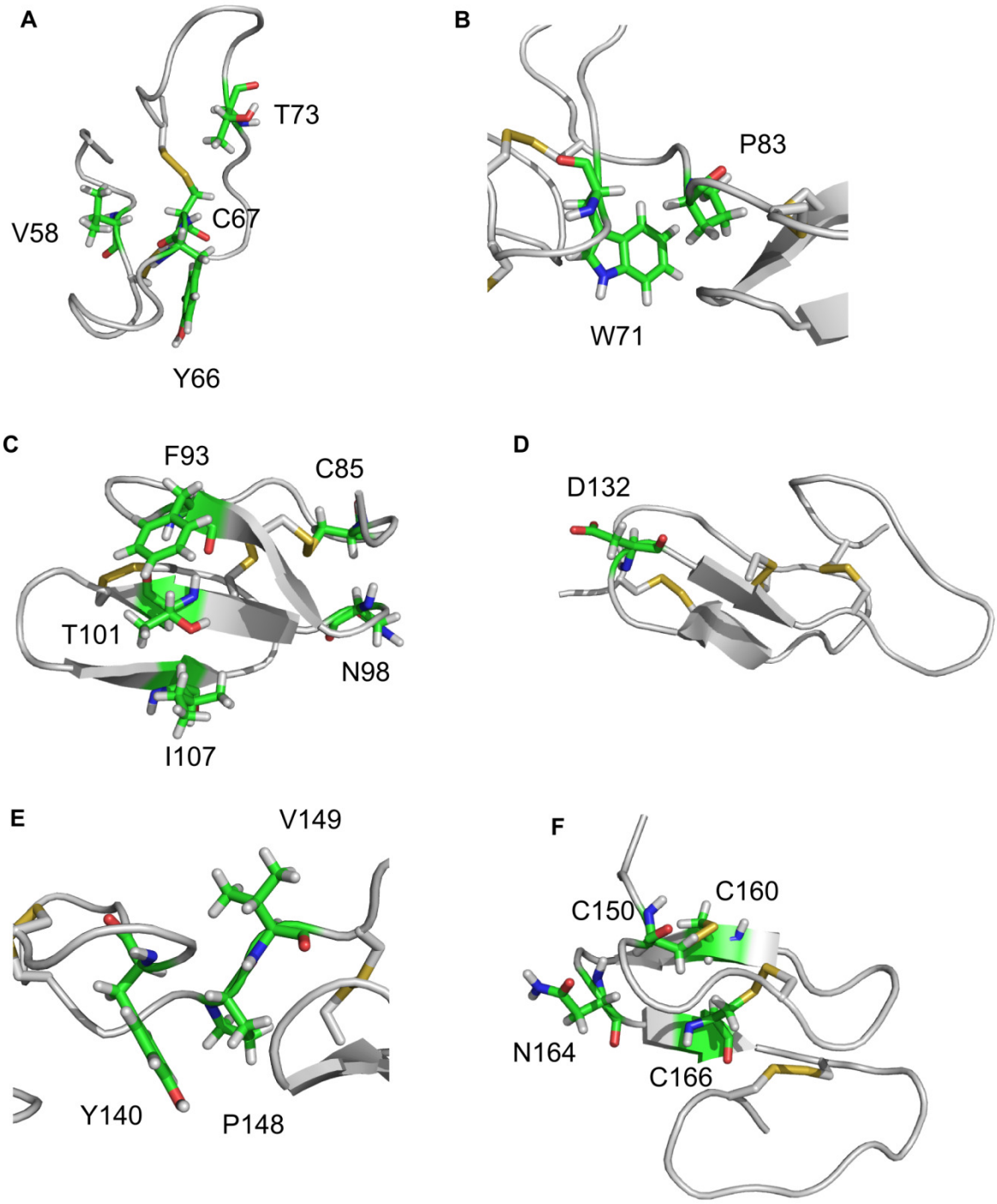


Figure 5.15 Residues in the FUN-EGF3 fragment used as markers of folding. A) Residues in the FUN domain. B) FUN-EGF1 interdomain interface. C) EGF1 domain. D) EGF2 domain. E) EGF2-EGF3 interdomain interface. F) EGF3 domain.

5.6.5 Characterisation of FUN-EGF3 fragments with single alanine substitutions

1D ^{1}H and 2D ^{1}H homonuclear NOESY spectra were acquired of the FUN-EGF3 proteins with single R62A, S88A, N98A, D131A and R159A substitutions. The chemical shifts of the marker cross-peaks in the 2D NOESY spectra were measured and compared to those in the WT spectrum (Table 5.4). It was important to examine multiple folding markers, as the absence of particular cross-peaks may have resulted from a weaker overall signal, rather than structural perturbations. Importantly, peaks arising from interactions between residues at the FUN-EGF1 and EGF2-EGF3 interdomain interfaces were present in the spectra of all modified proteins. This suggested that none of the substitutions had a significant effect on the FUN-EGF3 structure. In the spectrum of R62A FUN-EGF3, the Y66 $H^{\varepsilon*}$-$H^{\delta*}$ cross-peak was shifted compared to WT FUN-EGF3 spectrum, consistent with the proximity of R62 to Y66 in the structure. Likewise, in the S88A FUN-EGF3 spectrum, the N98-C85 and T101-I107 H^{α}-H^{α} cross-peaks were slightly shifted relative to equivalent peaks in the WT spectrum. This was again consistent with the proximity of these residues to the substitution.

There were more differences in the marker peaks in the spectrum of N98A FUN-EGF3. The W71-P83 and T101-I107 H^{α}-H^{α} cross-peaks in the 2D NOESY spectrum appeared to be shifted relative to equivalent peaks in the WT FUN-EGF3 spectrum (Figure 5.15A; Table 5.4). Nonetheless, these shifts were consistent with the proximity of N98 to both pairs of residues (Figure 5.15B). Moreover, the presence of marker peaks for all four domains suggested that the fold of N98A FUN-EGF3 was similar to the WT protein. The two substitutions in the EGF2-EGF3 region also appeared to have limited effects on the NOESY spectra. In the spectrum of D131A FUN-EGF3, all marker peaks were present except D132 $^{1}H^{\alpha}$-$H^{\beta*}$, the residue immediately after the substitution. There were no clear markers for the β-hairpin structure in EGF2, but the presence of the EGF2-EGF3 interface marker peaks suggested that the domain was correctly folded. Finally, the C160-

C166 H^{α}-H^{α} cross-peak was shifted in the spectrum of R159A FUN-EGF3, consistent with the proximity of C160 to R159.

Table 5.4 Chemical shifts of marker cross-peaks in 2D ^{1}H homonuclear NOESY spectra of modified FUN-EGF3 fragments. Cross-peak markers are listed for all four domains and the two interdomain interfaces. The residues involved are shown in Figure 5.14. Red boxes indicate the absence of a peak, while peaks with a combined chemical shift difference for the two dimensions of at least 0.03 ppm are coloured yellow. All modified proteins showed evidence of native folding, with perturbations most evident for cross-peaks assigned to residues in proximity to the substitutions.

	WT	R62A	S88A	N98A	D131A	R159A
FUN domain						
V58 H^{α}-C67 H^{α}	4.06, 5.23	4.05, 5.22	4.07, 5.22	4.07, 5.23	4.07, 5.23	4.07, 5.23
Y66 $H^{\varepsilon*}$-$H^{\delta*}$	6.70, 7.00	6.70, 7.02	6.70, 6.99	6.70, 7.00	6.70, 7.00	6.70, 6.99
T73 H^{β}-H^{α}	4.44, 5.30	4.44, 5.30	4.44, 5.30	4.44, 5.31	4.44, 5.31	4.44, 5.31
FUN-EGF1 interface						
W71 $H^{\zeta3}$-$H^{\eta2}$	6.74, 7.21	6.74, 7.21	6.74, 7.21	6.74, 7.21	6.74, 7.21	6.75, 7.21
W71 H^{α}-P83 H^{α}	4.17, 5.39	4.17, 5.39	4.18, 5.39	4.14, 5.39	4.17, 5.40	4.18, 5.39
EGF1 domain						
F93 $H^{\varepsilon*}$-$H^{\delta*}$	7.08, 7.31	7.08, 7.31	7.08, 7.31	7.07, 7.30	7.08, 7.30	7.08, 7.31
N98 H^{α}-C85 H^{α}	4.46, 4.95	4.45, 4.95	4.46, 4.93		4.45, 4.95	4.45, 4.95
T101 H^{α}-I107 H^{α}	4.43, 5.17	4.43, 5.17	4.43, 5.15	4.45, 5.16	4.42, 5.17	4.42, 5.17
EGF2 domain						
D132 H^{α}-$H^{\beta*}$	2.79, 4.37	2.79, 4.37	2.79, 4.37	2.79, 4.37		2.79, 4.37
EGF2-EGF3 interface						
Y140 $H^{\varepsilon*}$-$H^{\delta*}$	6.66, 6.71	6.66, 6.71	6.66, 6.71	6.66, 6.71	6.66, 6.71	6.66, 6.71
Y140 H^{α}-P148 H^{α}	4.72, 5.26	4.70, 5.26	4.71, 5.26	4.71, 5.26	4.70, 5.25	4.71, 5.25
V149 $H^{\gamma*}$-H^{α}	0.74, 3.94	0.74, 3.94	0.74, 3.94	0.73, 3.94	0.74, 3.94	0.74, 3.95
EGF3 domain						
C160 H^{α}-C166 H^{α}	4.87, 5.31	4.86, 5.31	4.87, 5.31	4.87, 5.31	4.87, 5.32	4.90, 5.36
N164 H^{α}-C150 H^{α}	4.47, 5.10	4.47, 5.10	4.48, 5.10	4.48, 5.10	4.48, 5.10	4.47, 5.09

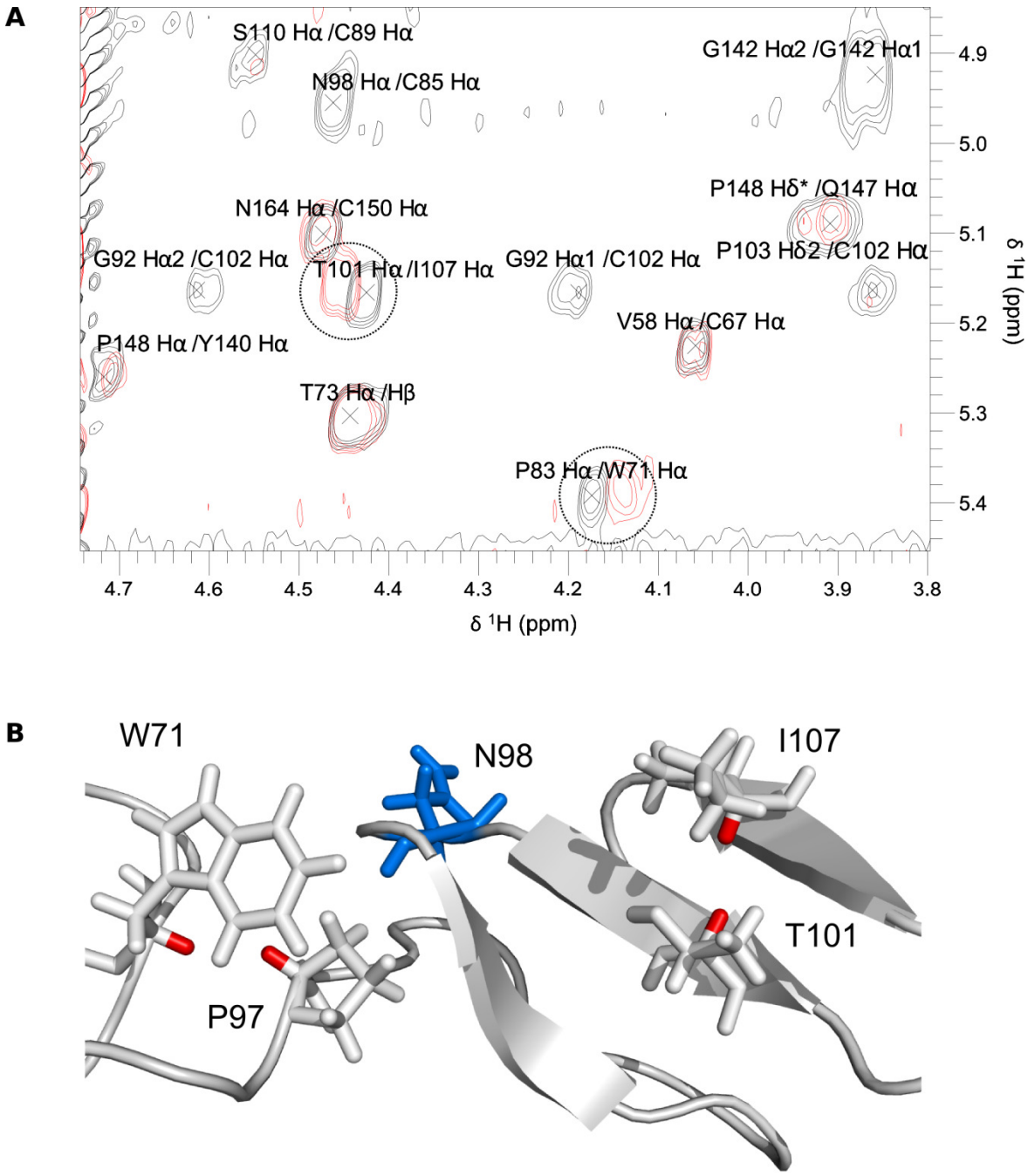


Figure 5.16 Characterisation of the N98A FUN-EGF3 fragment by NMR. A) Region of the 2D ^{1}H NOESY spectra for WT (black) and N98A (red) FUN-EGF3 showing $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peaks, as in Figure 5.14B. Cross-peaks in the spectrum of N98A FUN-EGF3 shifted compared to the WT spectrum are circled. Conditions for data acquisition were as described for the WT protein (Figure 5.14). B) FUN-EGF1 structure showing the substituted residue (N98A) in blue and the residues giving rise to the two shifted cross-peaks. The H^{α} atoms of these residues are coloured red.

5.6.6 Interaction of modified FUN-EGF3 fragments with cbEGF41-43-BirA

5.6.6.1 Pull-down assay

The effects of the alanine substitutions on the interaction of FUN-EGF3 with cbEGF41-43-BirA were tested in the pull-down assay. His-tagged proteins were produced and incubated with bead-immobilised cbEGF41-43-BirA-bio (Paul Evans, Part II project, 2012). All of the modified fragments bound to cbEGF41-43-BirA − there was no difference from WT FUN-EGF3 (Figure 5.17A). In addition, the N98A FUN-EGF3 fragment bound non-specifically to the streptavidin beads alone (Figure 5.17). It is possible that the N98A substitution changed the surface properties of the FUN-EGF3 fragment, although it did not significantly perturb the structure (Section 5.6.5).

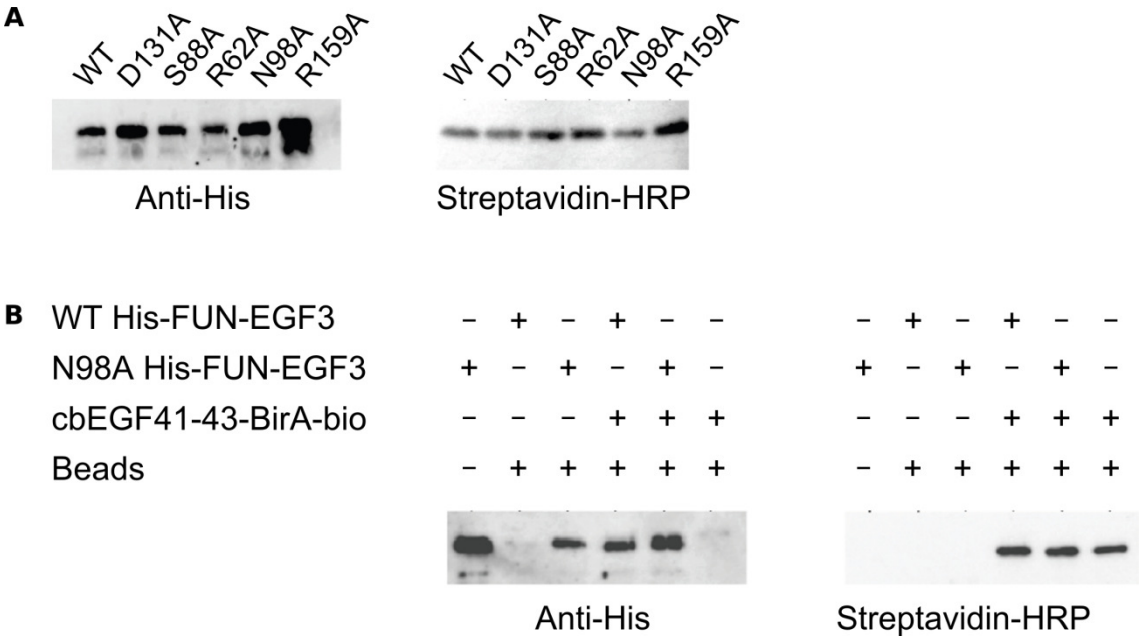


Figure 5.17 Results of the pull-down experiments with His-FUN-EGF3 fragments containing single alanine substitutions. A) Summary Western blots of the supernatant samples from pull-downs with bead-immobilised cbEGF41-43-BirA-bio. B) Western blots from the pull-down experiment with WT and N98A His-FUN-EGF3 fragments, including controls. N98A His-FUN-EGF3 bound non-specifically to the streptavidin-coated beads, but WT His-FUN-EGF3 did not. Data courtesy of Paul Evans (Paul Evans, Part II Project, 2012).

5.6.6.2 SPR assay

The pull-down assay may not be sufficiently sensitive to detect small perturbations to the interaction resulting from single residue changes. Therefore, the modified FUN-EGF3 fragments were tested in the SPR assay, which had the potential to provide quantitative information. The N98A FUN-EGF3 fragment was excluded from the assay because of its unusual behaviour in the pull-down assay. D131A FUN-EGF3 was also excluded as preliminary tests showed that it precipitated when dissolved in the SPR buffer and was therefore unsuitable for quantitative measurements. New batches of the R62A, S88A and R159A FUN-EGF3 proteins were made. SDS-PAGE showed that the protein samples each comprised a single species, although the mobility of R62A FUN-EGF3 was different from the WT under non-reducing conditions (Figure 5.18). Mass spectrometry also confirmed the presence of the substitutions (the mass of WT FUN-EGF3 is 14082 Da).

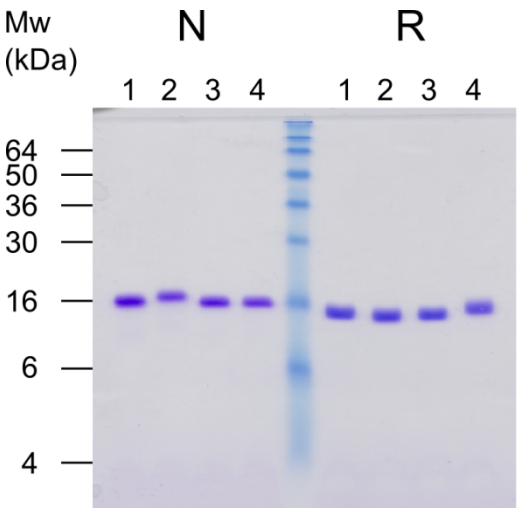


Figure 5.18 16% SDS-PAGE gel of FUN-EGF3 fragments. N and R indicate non-reducing and reducing conditions respectively. 1 = WT FUN-EGF3, 2 = R62 FUN-EGF3, 3 = S88A FUN-EGF3, 4 = R159A FUN-EGF3.

Table 5.5 Mass spectrometry results for FUN-EGF3 fragments with single alanine substitutions.

Protein	Expected molecular weight (Da)	Measured molecular weight (Da)
R62A FUN-EGF3	13997	13998
S88A FUN-EGF3	14066	14070
R159A FUN-EGF3	13997	14000

In the SPR assay, response levels on the flow cell with immobilised cbEGF41-43-BirA were compared when the various FUN-EGF3 fragments were injected onto the chip (Figure 5.19). The level was compared at the same time point for every protein, chosen as the WT interaction appeared to reach equilibrium by this time (Figure 5.10A). R62A and R159 FUN-EGF3 showed consistently lower response levels than WT FUN-EGF3 when injected at a concentration of 16 µM (Figure 5.19A). Response levels for WT FUN-EGF3 were unexpectedly lower than in previous attempts (for example, Figure 5.10), but the comparison was clear. However, the responses on the flow cell containing negative control ligand cbEGF32-34-BirA were reduced by the same factor (Figure 5.19B). It was therefore unlikely that any of the substitutions had specific effects on the binding of FUN-EGF3 to cbEGF41-43-BirA that could be detected in this assay.

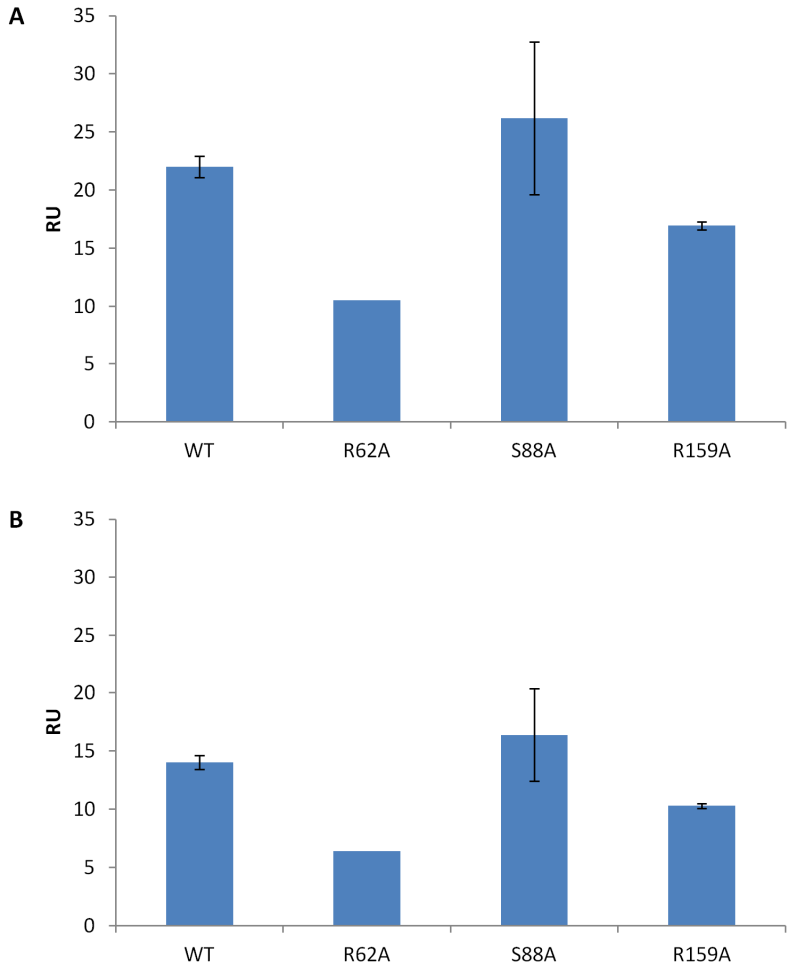


Figure 5.19 Results of SPR-based interaction assay with modified FUN-EGF3 fragments. A) Response units (RU) on cbEGF41-43-BirA-coupled flow cell when proteins were injected at a concentration of 16 µM. The RU value was taken 110 s after injection for every protein, as in Figure 5.9. n = 3 for all proteins except R62A, for which n = 2. Replicates are independent dilutions of the batch of protein. Error bars show the standard deviation on the measurements (with the exception of R62A). B) Response level on cbEGF32-34-BirA (negative control)-coupled flow cell when proteins were injected at a concentration of 16 µM.

5.7 Discussion

Several approaches were taken in attempts to characterise the binding site for cbEGF41-43 on FUN-EGF3. The first was an investigation into the domains of the FUN-EGF3 fragment required for binding to cbEGF41-43. Two smaller fragments were designed, based on the FUN-EGF3 structure, which suggested that the FUN-EGF1 and EGF2-EGF3 regions are independent structural units (Chapter 4). Surprisingly, while the FUN-EGF1 fragment was refolded successfully, the EGF2-EGF3 fragment did not fold efficiently. A C-terminal BirA tag was required to increase the refolding yield of EGF2-EGF3 in the EGF2-EGF3-BirA fragment, as has been previously observed for fragments of human Notch-1 (Cordle *et al.*, 2008a).

It is unclear why the EGF2-EGF3 region refolded efficiently in the FUN-EGF3 fragment, but apparently did not on its own. A possible explanation is that the FUN-EGF1 region promoted *in vitro* refolding of the EGF2 and EGF3 domains in the FUN-EGF3 fragment. In its native context in fibrillin, the EGF3 domain is likely to pack against the hyb1 domain. The EGF2-EGF3 region adopts the same fold as in the FUN-EGF3 protein in another human fibrillin-1 fragment comprising domains EGF2, EGF3, hyb1 and cbEGF1 (Ian Robertson, personal communication). Characterisation of the FUN-EGF1 and EGF2-EGF3-BirA fragments by NMR spectroscopy suggested that they adopted the same fold as the equivalent regions in the FUN-EGF3 fragment. However, not every peak coincided. The most likely explanations for these differences were the absence of the EGF1-EGF2 linker sequence from the FUN-EGF1 fragment and the addition of the C-terminal BirA tag in the EGF2-EGF3 fragment.

The pull-down and SPR assays showed that neither FUN-EGF1 nor EGF2-EGF3-BirA replicated the behaviour of FUN-EGF3 in binding to immobilised cbEGF41-43-BirA. More protein was required in the pull-down assay to observe binding and the SPR sensorgrams for the interaction with

cbEGF41-43-BirA were different for the smaller fragments, also showing a smaller response. The pull-down results appeared to contradict the SPR data. cbEGF41-43-BirA-bio still pulled down the smaller fragments when larger quantities were used, even though a high-affinity interaction was not observed in the SPR experiments. This may be explained by the differences in conditions or multimerisation of cbEGF41-43-BirA as discussed earlier (Section 3.8). Aggregation of FUN-EGF3 may be responsible for the higher response in the SPR assay compared to FUN-EGF1 and EGF2-EGF3-BirA. Therefore, the aggregation properties of FUN-EGF3, FUN-EGF1 and EGF2-EGF3-BirA need to be studied, for example using dynamic light scattering.

Attempts to identify specific residues in FUN-EGF3 important for its interaction with cbEGF41-43-BirA were not successful. Titrations of ^{15}N-labelled N-terminal fragments with unlabelled cbEGF41-43-BirA showed general reductions in cross-peak intensities in HSQC spectra, rather than specific effects. This may have resulted from the formation of a complex between the two fragments, causing slower tumbling of the N-terminal fragments, but equally could have been caused by non-specific interactions and aggregation. Therefore, sequence conservation was used to select residues for substitution by site-directed mutagenesis. 5 FUN-EGF3 fragments with single alanine substitutions (R62A, S88A, N98A, D131A and R159A) were produced and characterised. Pull-down and SPR assays did not show any effects on binding to cbEGF41-43-BirA resulting from these substitutions. There are two possible explanations. First, the substitutions did not significantly disrupt binding of FUN-EGF3 to cbEGF41-43-BirA. Second, the assay was not sufficiently sensitive to detect the effects of single residue changes. Larger-scale changes, such as loop deletions, may be needed to perturb binding and dissect the interaction. Possible alternative strategies are discussed in Section 7.4, Chapter 7.

Despite the lack of success in identifying specific residues involved in binding, it was demonstrated that the fold of FUN-EGF3 fragments could be assessed using NMR spectroscopy.

Specific marker cross-peaks were identified in the 2D ^{1}H homonuclear NOESY spectrum of FUN-EGF3 that could be used for comparison with modified fragments. These were used to demonstrate the correct formation of interdomain interfaces in the FUN-EGF3 fragments with single alanine substitutions. This method has possible applications in the study of the effects disease-linked mutations that result in residue substitutions and, potentially, misfolding. The effects of three such MFS mutations on the FUN-EGF3 fragment are investigated in Chapter 6.

Chapter 6

Effects of MFS mutations on the N- and C-terminal domains of human fibrillin-1

6.1 Introduction

Disease-linked mutations in the human *FBN1* gene affect the structural and functional properties of the fibrillin-1 protein (Section 1.9). In particular, many of the missense mutations linked to MFS have been shown to cause structural defects (Section 1.9.5). They can result in the addition or removal of a cysteine residue, loss of calcium-binding ligands from cbEGF domains, disruption of interdomain packing, or can have more subtle effects on protein folding. The severity of the effects on folding cannot necessarily be predicted from the change in amino acid sequence, and so high-resolution structural information is required. However, there are some examples of mutations that result in defective fibrillin function without obvious structural consequences. A patient with skeletal features of MFS had a mutation resulting in the R2726W substitution, causing the loss of C-terminal processing and the failure of mutant fibrillin molecules to incorporate into the ECM (Section 1.7.1) (Milewicz *et al.*, 1995). In addition, a mutation leading to the N164S substitution, found in a patient with dominant ectopia lentis (Comeglio *et al.*, 2002) causes a greater than 10-fold reduction in the binding of fibrillin-1 binding to LTBP-4 (Ono *et al.*, 2009).

Structural perturbations, leading to protein misfolding, may affect the secretion of fibrillin-1 by cells into the extracellular medium. The effects of some disease-linked mutations on both structure and secretion have been studied in detail. Some mutations in the 'neonatal' region of the *FBN1* gene (Section 1.9.1) perturb calcium binding by cbEGF domains and also result in fibrillin-1 retention in the endoplasmic reticulum (Whiteman *et al.*, 2007). However, this is not the case for all mutations that cause the loss of calcium-binding ligands. Furthermore, some mutations can have long-range effects on the folding of adjacent domains. Mutations in *FBLN5*, the gene encoding fibulin-5, found in patients with age-related macular degeneration (AMD) and cutis laxa (CL) can also cause defective calcium binding by cbEGF domains. One study found that

two such mutations, resulting in the substitutions G267S (AMD) and S227P (CL), led to long-range structural perturbations in fibulin-5 fragments produced by *in vitro* refolding, as well as intracellular retention of protein produced by fibroblast cells (Schneider *et al.*, 2010).

6.2 Insights into disease-linked missense mutations affecting the FUN-EGF3 region of fibrillin-1

32 unique missense mutations in exons 2-5 of the *FBN1* gene are listed in databases of patients with MFS or MFS-like symptoms (Section 1.9.2; Appendix A). These exons encode the FUN-EGF3 region of fibrillin-1. Residues affected by these mutations can now be mapped on to the structure of the FUN-EGF3 fragment, which was determined previously (Chapter 4). The majority of mutations in this region are predicted to result in the gain or loss of a cysteine residue, but 7 do not. Three of these affect the FUN-EGF1 region (Figure 6.1A). The T101A substitution was shown to perturb heparin binding by recombinantly-produced N-terminal fragments (Cain *et al.*, 2005) and is associated with classical MFS (Stheneur *et al.*, 2009). However, the affected residue, located on the second strand of the β-hairpin of the EGF1 domain, is not fully conserved (Figure 5.12, Section 5.6.1). Interestingly, the N57D and W71R substitutions affect residues involved in interdomain packing (Figure 6.1B). They might therefore disrupt the FUN-EGF1 interdomain interface. Whereas the patient with the predicted N57D substitution had classical MFS (Chung *et al.*, 2009), the patient with the W71R substitution only had skeletal and ocular features of MFS (Sakai *et al.*, 2006).

In the EGF2 domain, the G127D substitution (Kan *et al.*, 2010) affects a residue in a turn before the first strand of the β-hairpin of the EGF2 domain (Figure 6.1C). Mutations affecting a glycine residue at the same position in cbEGF domains have previously been identified. A mutation leading to the G1127S substitution in the fibrillin-1 cbEGF13 domain is associated with a predisposition to aortic aneurysms and dissection (Francke *et al.*, 1995). In addition, the G60S

substitution in human factor IX is found in some patients with mild haemophilia B. Both of these substitutions affect protein folding *in vitro* (Whiteman *et al.*, 1998). It is possible that the G127D substitution could also perturb the fold of the EGF2 domain.

The P148S substitution was identified in a patient with classical MFS (Stheneur *et al.*, 2009) and affects a residue involved in interdomain packing between the EGF2 and EGF3 domains (Figure 6.1D). It therefore may disrupt the EGF2-EGF3 interface. In the EGF3 domain there are two further MFS-associated substitutions: N164S, described previously (Section 6.1), and Q176P (Waldmüller *et al.*, 2007), which affects a residue on turn before C6. Without a structure of the EGF3-hyb1 interface, it is hard to predict possible structural effects of the Q176P substitution.

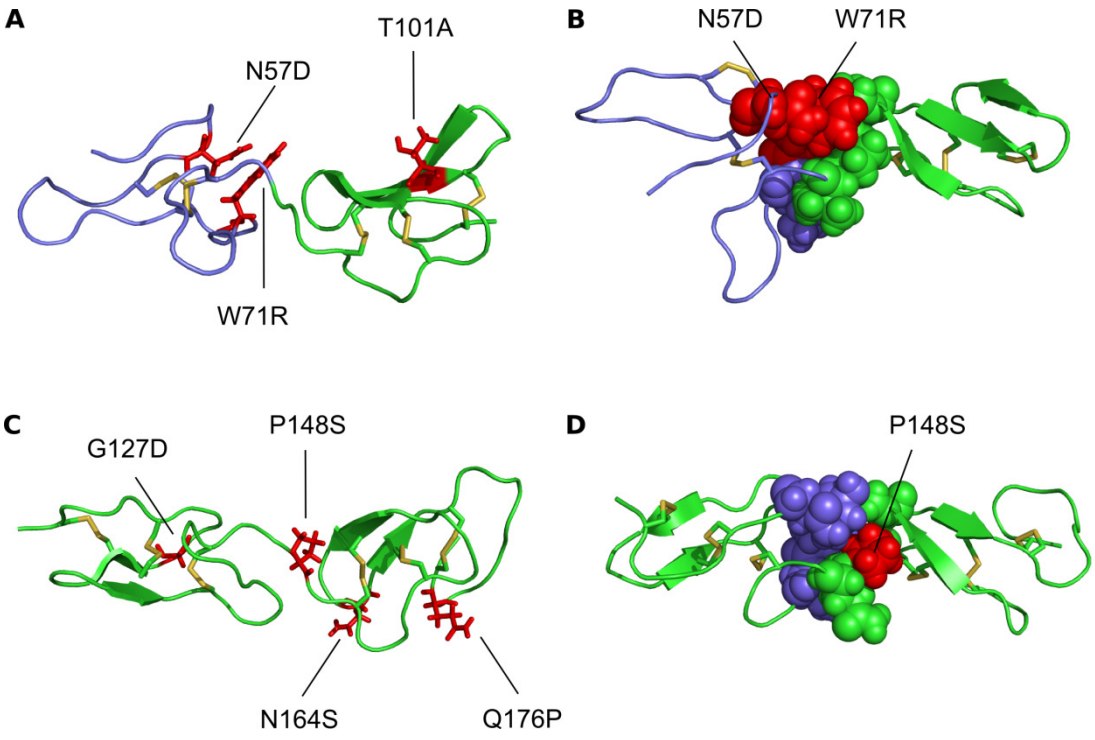


Figure 6.1 Locations of disease-linked residue substitutions in FUN-EGF3. A) Structure of the FUN-EGF1 region showing residues (red) substituted as the result of mutations. Substitutions that result in the addition or removal of cysteine residues are not shown for clarity. B) Residues involved in packing between the FUN and EGF1 domains, highlighting the residues affected by the N57D and W71R substitutions. C) Structure of the EGF2-EGF3 region showing residue substitutions. D) Residues involved in interdomain packing in the EGF2-EGF3 region, highlighting the location of the P148S substitution.

6.3 MFS missense mutations affecting the cbEGF41-43 region

The large amount of information about the structure, calcium-binding properties and interdomain packing of cbEGF domains (Sections 1.3.1 and 1.4) means that predictions can be made about the effects of many missense mutations. Indeed, the majority of MFS missense mutations affecting the cbEGF41-43 region of fibrillin-1 are likely to cause structural perturbations, based on the consequent residue substitutions (Figure 1.11, Section 1.9). However, the effects of two MFS-linked substitutions, Y2639N and S2660R, cannot be easily explained (Section 1.9).

Sequence alignments of the cbEGF41-43 region show that S2660 is not conserved, but Y2639 is present, or is replaced by phenylalanine, in all species examined except the sea squirt (Figure 6.2A). Interestingly, there are several other aromatic residues in the cbEGF41-43 region that are largely conserved, but are not part of the consensus calcium-binding or interdomain packing motifs. A homology model of the human fibrillin-1 cbEGF41-43 region was made using Modeller software (Eswar *et al.*, 2006), with the structure of the fibrillin-1 cbEGF32-33 fragment as a template (Downing *et al.*, 1996). This suggested that these residues might lie on the surface of the molecule (Figure 6.2B) and contribute to its functional properties. It is possible that the Y2639N substitution could perturb the function of this region of fibrillin-1.

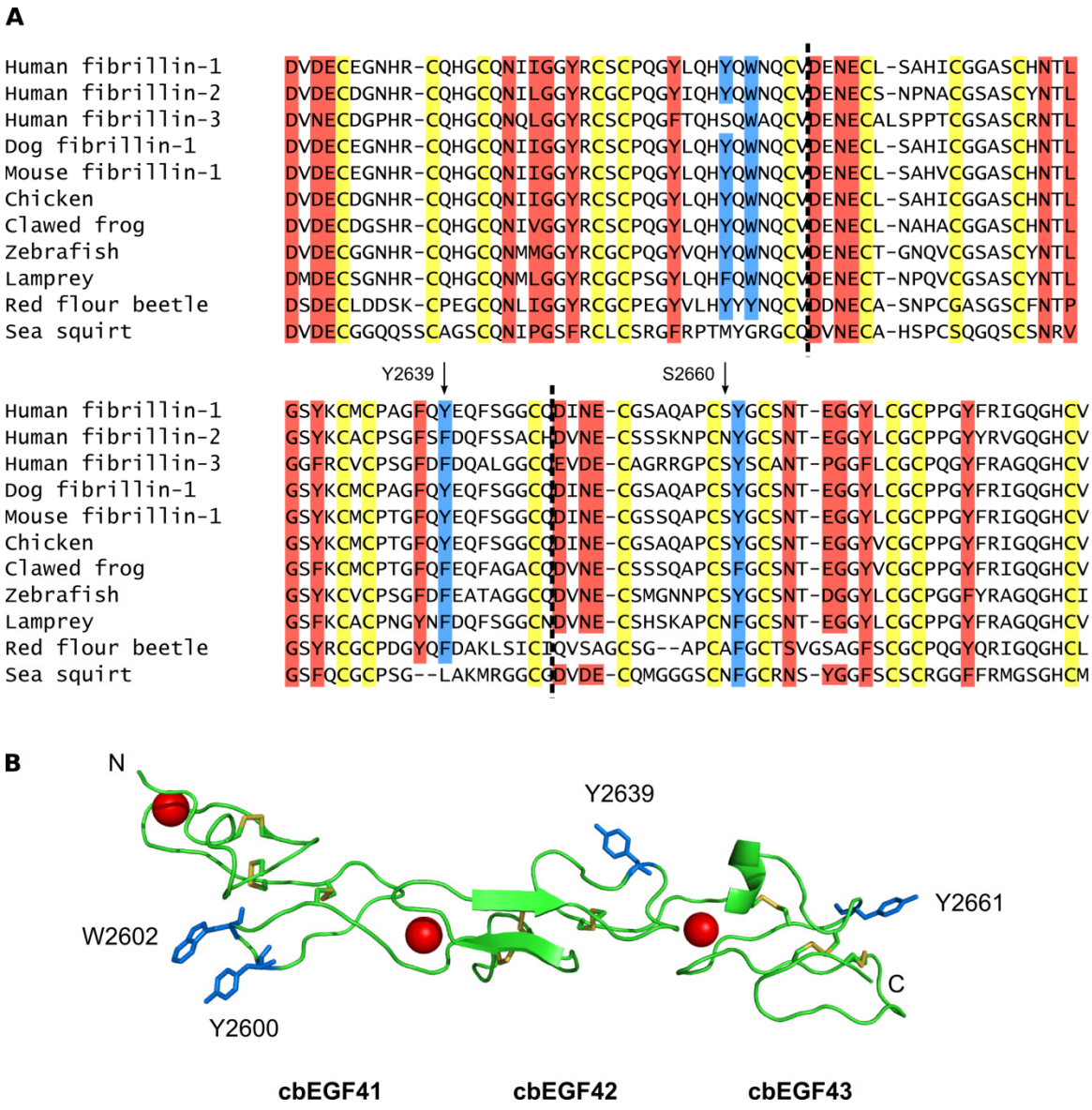


Figure 6.2 Conservation of aromatic residues in the cbEGF41-43 region of fibrillin. A) Alignment of sequences from the cbEGF41-43 region. Alignments were performed in ClustalW2. Cysteine residues are highlighted in yellow and residues involved in calcium binding or interdomain packing are highlighted in red. The last EGF-like domain in the red flour beetle lacks several key calcium-binding consensus residues and so is predicted not to bind calcium ions. Additional conserved aromatic residues are highlighted in blue. The structure of the interface between the last EGF/cbEGF and the unique C-terminal region is unknown, so interdomain packing residues have not been identified. Dotted lines show the domain boundaries and arrows indicate residues S2660 and Y2639 in the human fibrillin-1 sequence. B) Homology model of the cbEGF41-43 region showing the highlighted conserved aromatic residues (Part A) in the stick representation, including Y2639. These residues are surface-exposed in the model. Calcium ions are shown as red spheres.

6.4 Chapter aims

The aim of this chapter is to characterise the effects of several MFS-associated substitutions. In particular, the molecular and structural effects of the MFS-associated residue substitutions N57D, W71R and P148S have not yet been studied. As the affected residues lie at interdomain interfaces in the FUN-EGF3 region, it is possible that the substitutions cause structural perturbations. They might also disrupt the interaction of FUN-EGF3 with cbEGF41-43. Here, the effect of these substitutions was examined using two independent approaches. First, they were incorporated into the FUN-EGF3 fragment using site-directed mutagenesis. Modified fragments were then characterised by SDS-PAGE, mass spectrometry and NMR spectroscopy, and tested in the pull-down interaction assay (Section 3.7.1).

Second, the same mutations were individually introduced into a vector encoding a longer N-terminal fibrillin-1 fragment. MSU1.1 human fibroblast cells were then transfected to produce stable cell lines expressing the WT and modified fragments. This well-established transformed cell line expresses fibrillin-1 endogenously and assembles microfibrils in culture (Cardy and Handford, 1998; Kettle *et al.*, 2000). Secretion of the N-terminal fragments by the MSU1.1 cells was then studied as described previously (Schneider *et al.*, 2010; Whiteman and Handford, 2003; Whiteman *et al.*, 2007). The Y2639N substitution was also introduced into the cbEGF41-43-BirA fragment, in order to investigate whether it affected the N-C interaction.

6.5 Production and characterisation of FUN-EGF3 fragments with MFS-associated substitutions

6.5.1 Site-directed mutagenesis

Residue substitutions N57D, W71R and P148S were incorporated into the FUN-EGF3 fragment by site-directed mutagenesis. Point mutations were introduced into pQE-30 vector encoding FUN-EGF3 (Section 3.4) using the QuikChange Lightning site-directed mutagenesis kit (Stratagene; Section 2.1.10). The method is illustrated in Figure 6.3A and primer sequences used in mutagenesis are given in Appendix B. The pQE-30 N57D FUN-EGF3 construct was made by Joanne McNaught-Davis (Joanne McNaught-Davis, Part II project 2011). DNA sequencing confirmed the successful introduction of mutations (Figure 6.3B).

6.5.2 Protein production

FUN-EGF3 constructs with MFS-associated substitutions were produced by the same method used for WT FUN-EGF3 (Sections 2.2 and 3.4.2.1). A small amount of reduced WT FUN-EGF3 was refolded alongside each of the modified FUN-EGF3 proteins. This allowed direct comparisons with the WT protein when refolded material was purified by HPLC (Section 2.2.6). Differences in HPLC elution profiles can indicate that a protein with a substitution is incorrectly folded. The shape of the HPLC profile of refolded N57D FUN-EGF3 protein was broadly similar to that of the WT protein (Figure 6.4A). However, there were some differences. First, the major fraction of the N57D protein eluted later than the WT. Second, another major species eluted earlier than the expected major peak. These two fractions were collected separately. By contrast, the HPLC elution profile of W71R FUN-EGF3 was very different from the WT (Figure 6.4B). The W71R eluted as a broad peak, earlier than the major WT species, and was collected as a single fraction.

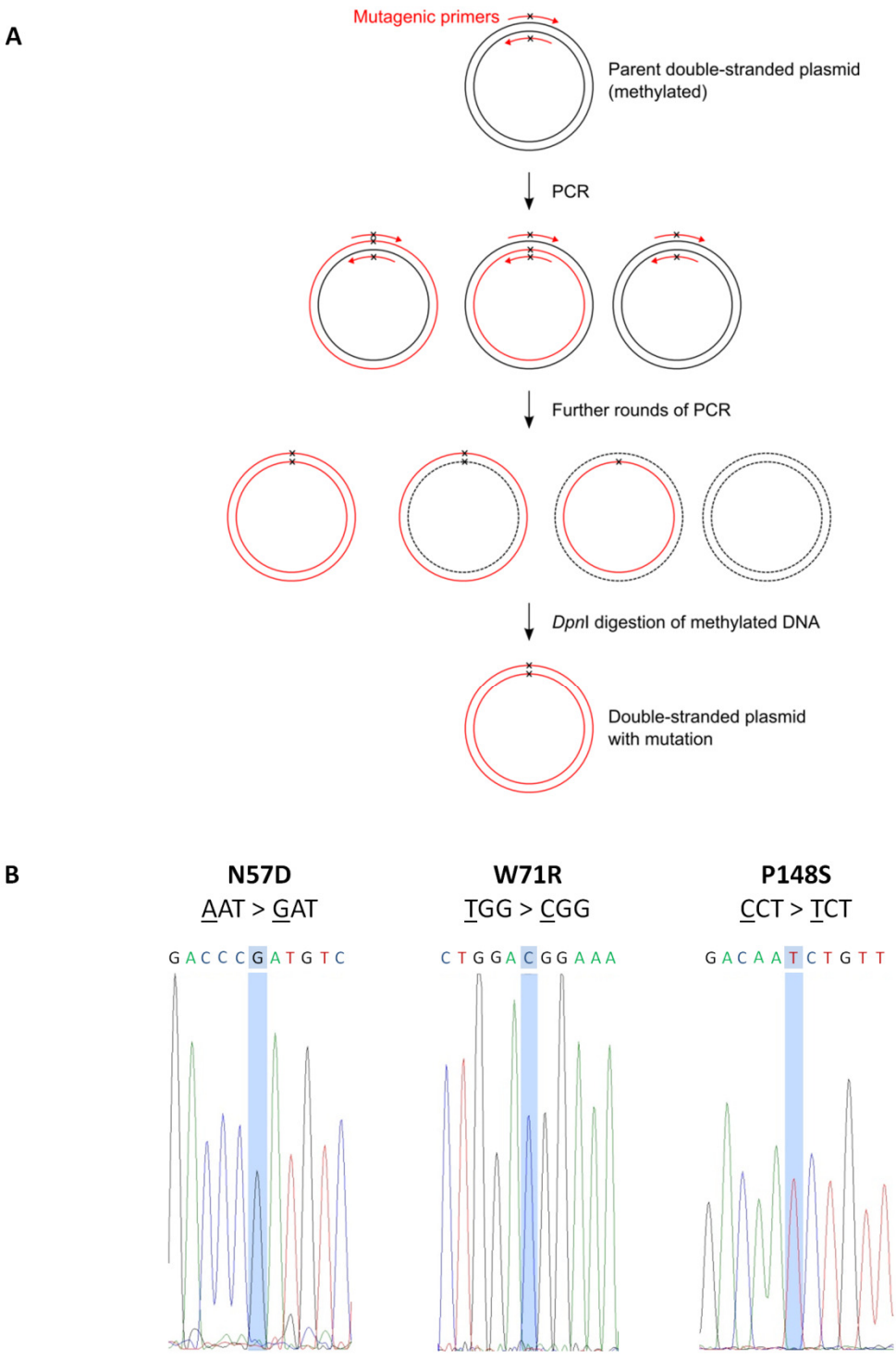


Figure 6.3 Mutagenesis using the QuikChange Lightning Kit. A) Illustration of QuikChange mutagenesis. Point mutations are introduced by PCR using primers containing the desired nucleotide changes. The template DNA, which is methylated, is degraded using the *Dpn*I enzyme. The remaining unmethylated DNA is transformed into bacteria, amplified and then purified. Diagram modified from the QuikChange Lightning handbook (Agilent). B) DNA sequencing traces of the N57D, W71R and P148S FUN-EGF3 constructs showing the correct, replaced nucleotides.

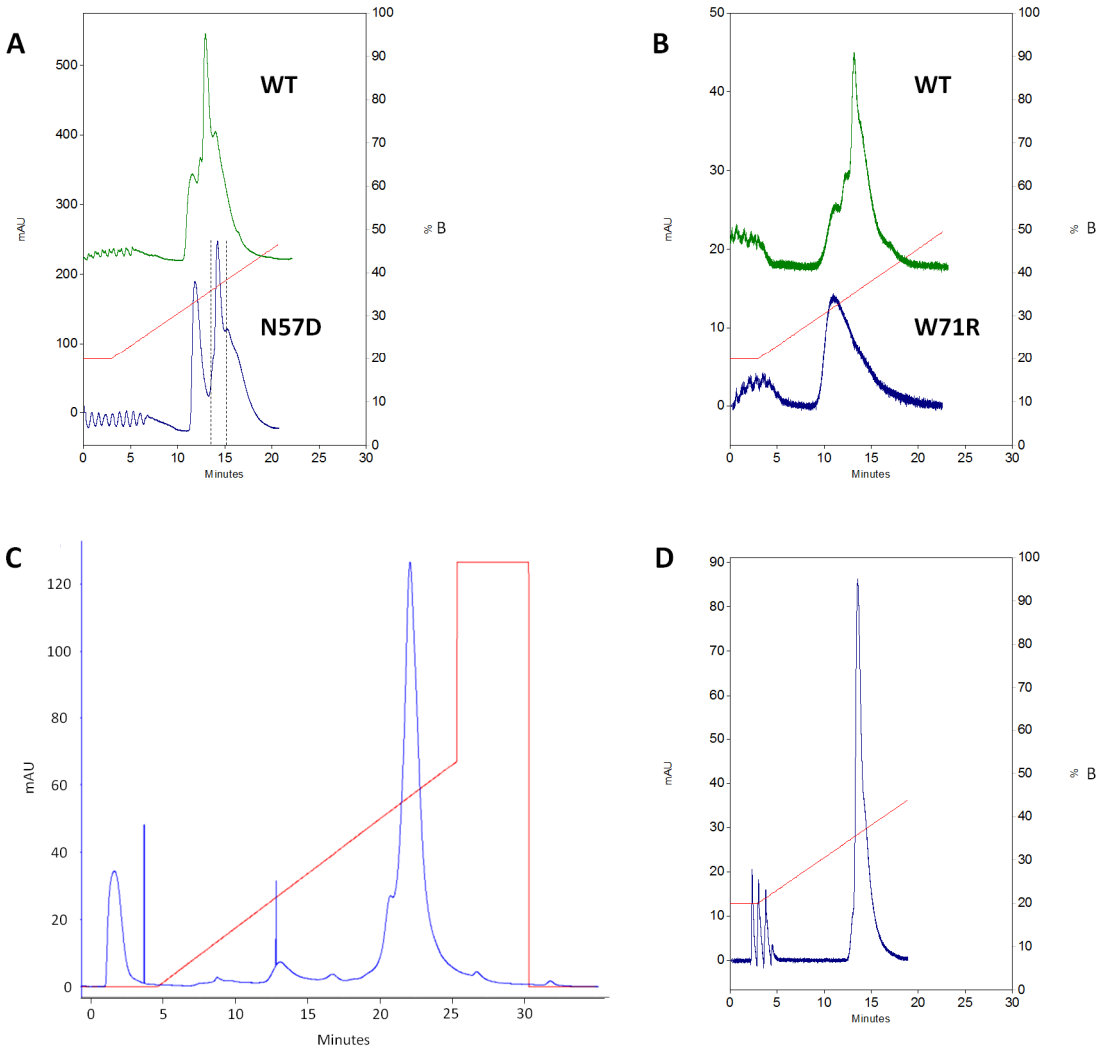


Figure 6.4 Refolding and purification of N57D and W71R FUN-EGF3 proteins. Blue and red lines indicate the absorbance at 280 nm and % of solution B respectively at a given time point. Solution B was 80% acetonitrile, 0.1% TFA for reverse-phase HPLC and 20 mM MES pH 6.0, 2 M NaCl for MonoS FPLC purification. A) HPLC chromatograms of refolded N57D (blue) and WT (green) FUN-EGF3. The collection boundaries of the major fraction of N57D FUN-EGF3 are indicated by dashed lines. B) HPLC chromatograms of refolded W71R (blue) and WT (green) FUN-EGF3. C) FPLC chromatogram of N57D His-FUN-EGF3. D) Final HPLC chromatogram of N57D His-FUN-EGF3.

Attempts were made to purify refolded material further after the first HPLC step. While the typical yield of WT protein after refolding and HPLC was ~30%, it was only ~3.5% for N57D FUN-EGF3. A second, independent refolding attempt also had a low yield (Joanne McNaught-Davis, Part II Project, 2011). Therefore, the N57D-substituted protein was not treated with factor Xa, so that material from the same preparation could be both tested in the pull-down assay and characterised by NMR spectroscopy. The major fraction was purified by cation exchange FPLC (Section 2.2.8; Figure 6.4C), followed by a final HPLC step (Figure 6.4D). There was one major peak in both chromatograms, although some material eluted in the void volume during FPLC purification. Attempts to purify the early-eluting peak were unsuccessful, yielding only a very small amount of final product (not shown). Furthermore, refolded W71R FUN-EGF3 could not be purified further by FPLC, as material eluted in the void volume (not shown). Combined with the HPLC elution profile, this suggested that folding of the W71R FUN-EGF3 fragment was abnormal.

The behaviour of P148S FUN-EGF3 was more similar to the WT protein during purification. HPLC elution profiles of refolded P148S and WT FUN-EGF3 were very similar and the major species eluted at similar acetonitrile concentrations (Figure 6.5A). SDS-PAGE analysis of the early-eluting peaks showed that they contained disulphide-bonded aggregates and higher-mobility species that were likely to be contaminants or degradation products (not shown). Some of material from the major HPLC fraction was treated with factor Xa (Section 2.2.7) and purified further by FPLC (Section 2.2.8; Figure 6.5B) and HPLC (Figure 6.5C). The remaining material, still containing the His_6 tag (P148S His-FUN-EGF3), was also purified by FPLC (Figure 6.5D) and HPLC (Figure 6.5E). Although the FPLC elution profiles appeared very similar to those of the WT protein (Figure 3.5, Section 3.4.1), the final HPLC elution profiles of the P148S proteins seemed broader. This hinted at a possible difference in folding compared to WT FUN-EGF3.

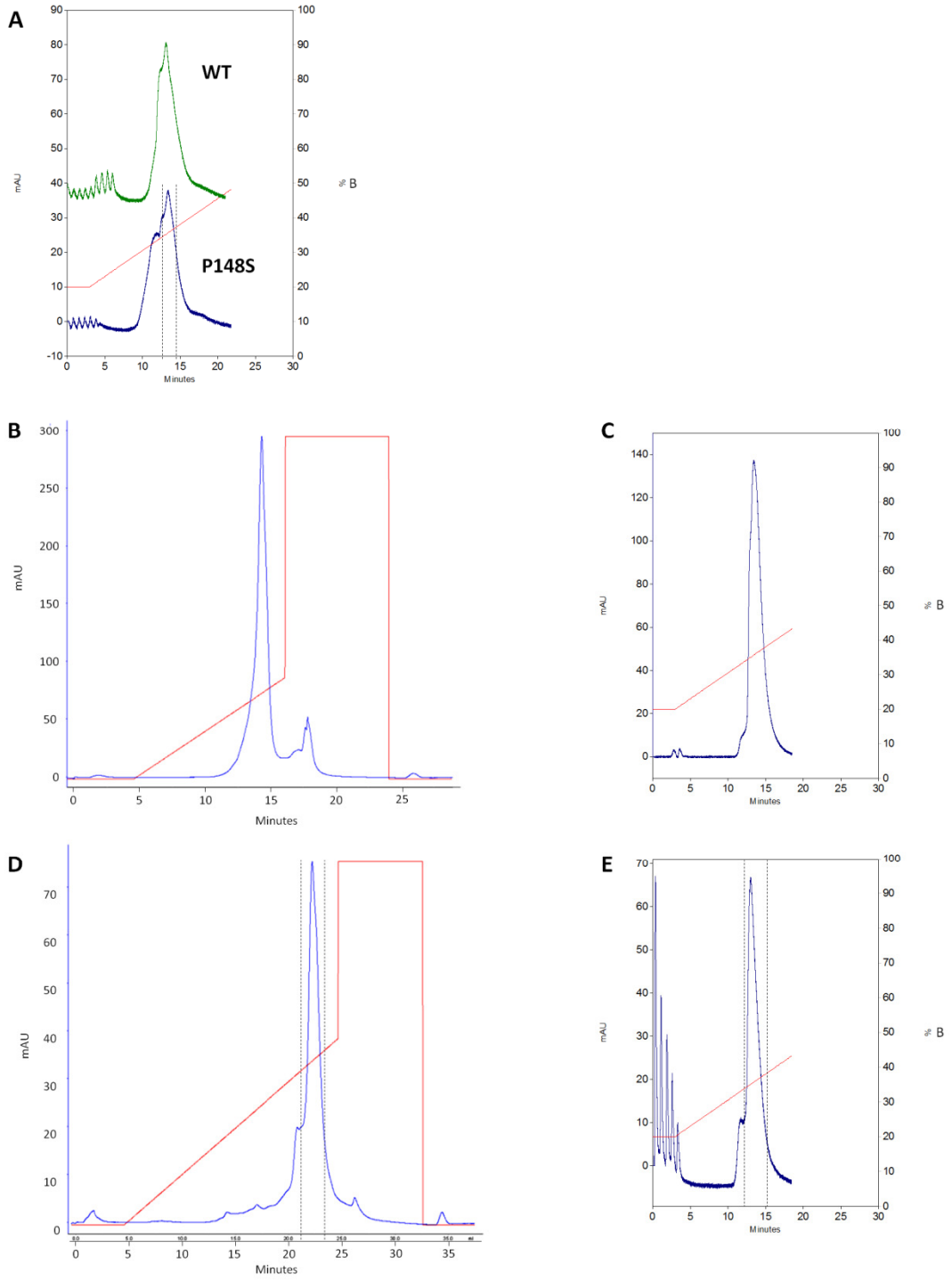


Figure 6.5 Refolding and purification of P148S FUN-EGF3. A) HPLC chromatograms of refolded P148S (blue) and WT (green) FUN-EGF3. B) FPLC chromatogram of refolded P148S FUN-EGF3 after treatment with factor Xa. C) HPLC chromatogram of P148S FUN-EGF3 after FPLC purification. D) FPLC chromatogram of P148S His-FUN-EGF3. E) HPLC chromatogram of P148S His-FUN-EGF3 after FPLC purification. Buffer conditions are as in Figure 6.4. Dotted lines indicate collection boundaries of the major fraction where there were significant other species.

6.5.3 SDS-PAGE and mass spectrometry analysis

Purified proteins were analysed by SDS-PAGE and mass spectrometry. SDS-PAGE analysis of WT, N57D, W71R and P148S His-FUN-EGF3 fragments revealed differences (Figure 6.6A).

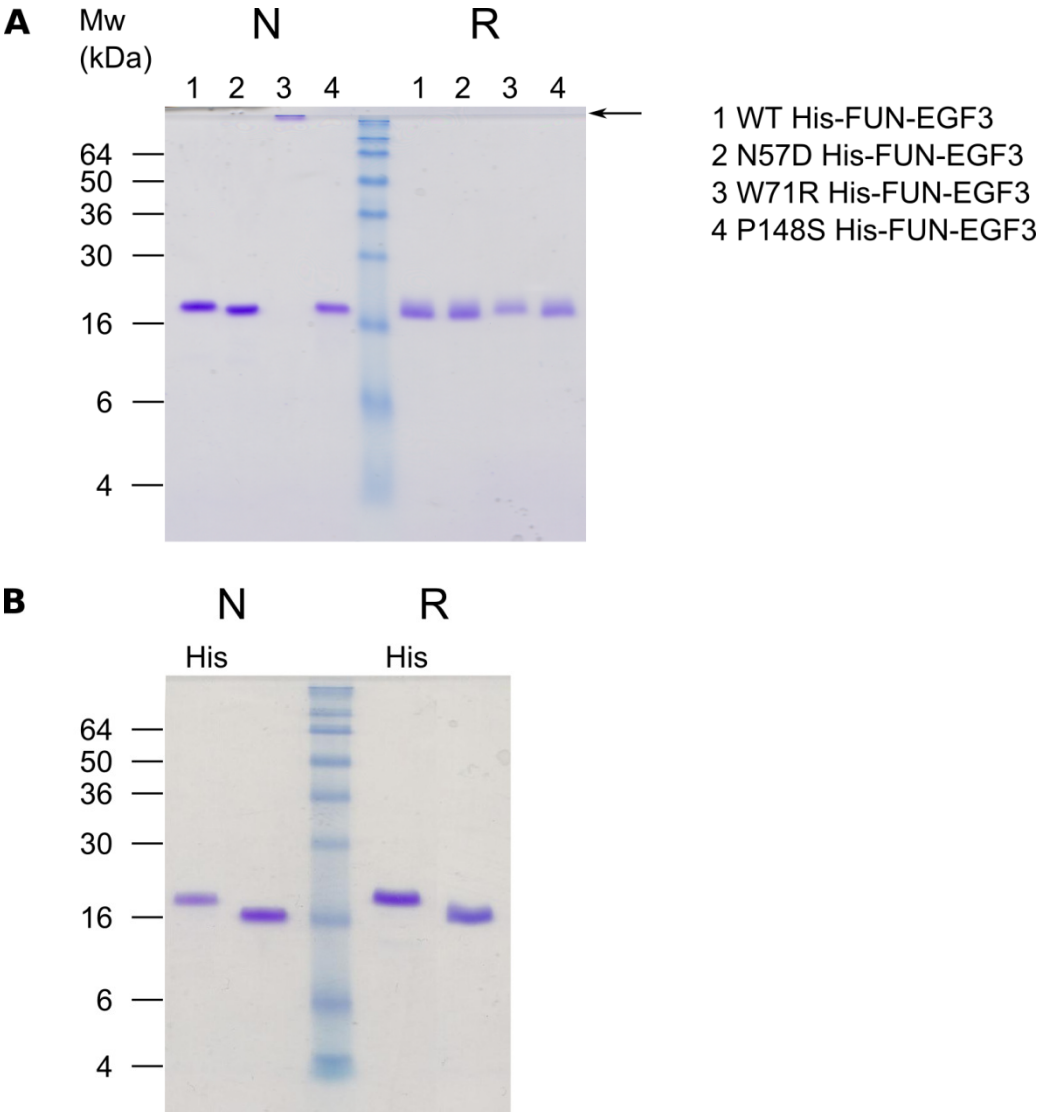


Figure 6.6 16% SDS-PAGE gels of FUN-EGF3 fragments with MFS-associated substitutions. A) His-FUN-EGF3 proteins with samples prepared in non-reducing (N) or reducing (R) sample buffer. The arrow indicates the high-molecular weight band in the W71R FUN-EGF3 sample under non-reducing conditions. B) His-FUN-EGF3 (His) and factor Xa-treated P148S FUN-EGF3 proteins.

N57D His-FUN-EGF3 ran as a single band, but its mobility was slightly higher than the WT protein under non-reducing conditions. W71R His-FUN-EGF3 ran as a low-mobility band under non-reducing conditions, but at a similar molecular weight to WT protein under reducing conditions. This suggested that the protein formed disulphide-bonded aggregates and therefore was misfolded. By contrast, P148S His-FUN-EGF3 appeared to be similar to WT His-FUN-EGF3. A gel of the factor Xa-treated material also showed that it ran as a single band and confirmed successful proteolytic cleavage (Figure 6.6B).

Mass spectrometry confirmed the expected masses of the refolded N57D His-FUN-EGF3, P148S His-FUN-EGF3 and P148S FUN-EGF3 fragments (Table 6.1). As W71R His-FUN-EGF3 appeared to form disulphide-bonded aggregates during refolding, it was impractical to analyse the oxidised form by mass spectrometry. Therefore the mass of the reduced protein was measured, confirming the expected molecular weight, which was different from the mass of reduced WT His-FUN-EGF3. This helped to exclude the possibility that misfolding resulted from the production of protein with the wrong sequence.

Table 6.1 Mass spectrometry results for FUN-EGF3 and His-FUN-EGF3 proteins with MFS-associated substitutions

Protein	Expected molecular weight (Da)	Measured molecular weight (Da)
N57D His-FUN-EGF3	15937	15938
P148S His-FUN-EGF3	15926	15928
P148S FUN-EGF3	14072	14072
W71R His-FUN-EGF3 (reduced)	15928	15930
WT His-FUN-EGF3 (reduced)	15958	15960

6.5.4 Characterisation of N57D and P148S FUN-EGF3 by NMR spectroscopy

N57D and P148S FUN-EGF3 proteins were characterised further using NMR spectroscopy. Markers for folding in the 2D ^{1}H NOESY spectrum of WT FUN-EGF3 were previously identified, allowing comparison with FUN-EGF3 proteins containing substitutions (Section 5.6.4). Relatively little N57D His-FUN-EGF3 protein was available because of the low refolding yield (Section 6.5.2). Nonetheless, a 2D spectrum was acquired and cross-peaks were evident (Figure 6.7). 296 scans were acquired for each block of the spectrum, compared with 96 scans for the spectrum of WT FUN-EGF3, in order to obtain a reasonable signal (Section 2.6.2.2). A summary of the analysis is given in Table 6.2.

The pattern of cross-peaks between aromatic side-chain resonances in the N57D spectrum was different from in the WT spectrum (Figure 6.7A). Notably, many of the cross-peaks between nuclei of the W71 side chain were absent, with the exception of the $^{1}H^{\eta 2}$-$^{1}H^{\zeta 2}$ cross-peak, which appeared to have shifted. There were also differences in the positions of the Y66 and Y63/Y170 (ambiguous) $H^{\varepsilon *}$-$H^{\delta *}$ cross-peaks. An additional cross-peak in the N57D spectrum, absent from the WT spectrum, could not be assigned. In addition, the V58-C67 and W71-P83 $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peaks were absent from the N57D spectrum (Figure 6.7B). Differences between the WT and N57D spectra could not be explained by the additional presence of the N-terminal His_6-tag in the N57D FUN-EGF3 protein, as the missing and shifted cross-peaks in the N57D spectum were unaffected in the spectrum of D131A His-FUN-EGF3 protein (data not shown). As the affected marker peaks were all assigned to residues in the FUN domain and at the FUN-EGF1 interface (Table 6.2), it was very likely that the FUN domain was misfolded in N57D His-FUN-EGF3. However, this only represented the relatively small amount of protein that could be purified from the refolding mixture.

Table 6.2 Analysis of cross-peak markers in 2D ^{1}H homonuclear NOESY spectra of N57D and P148S FUN-EGF3 proteins. Cross-peak markers for WT FUN-EGF3 were defined in Section 5.7.4 and are listed here as in Table 5.4. Red indicates that a peak was absent. Peaks that had a combined chemical shift for both dimensions of at least 0.03 ppm are highlighted in yellow. 'Unclear' indicates that the presence of the peak was not clear, as it was obscured by noise from the H_2O peak.

	WT	N57D	P148S
FUN domain			
V58 H^{α}/C67 H^{α}	4.06, 5.23		4.06, 5.23
Y66 $H^{\varepsilon*}$/$H^{\delta*}$	6.70, 7.00		6.70, 7.00
T73 H^{β}/H^{α}	4.44, 5.30		
FUN-EGF1 interface			
W71 $H^{\zeta3}$/$H^{\eta2}$	6.74, 7.21		6.74, 7.21
W71 H^{α}/P83 H^{α}	4.17, 5.39		4.17, 5.38
EGF1 domain			
F93 $H^{\varepsilon*}$/$H^{\delta*}$	7.08, 7.31	7.08, 7.31	7.07, 7.30
N98 H^{α}/C85 H^{α}	4.46, 4.95	4.46, 4.96	4.45, 4.95
T101 H^{α}/I107 H^{α}	4.43, 5.17	4.42, 5.17	4.43, 5.17
EGF2 domain			
D132 H^{α}/$H^{\beta*}$	2.79, 4.37	2.79, 4.37	2.79, 4.37
EGF2-EGF3 interface			
Y140 $H^{\varepsilon*}$/$H^{\delta*}$	6.66, 6.71	6.66, 6.71	6.68, 6.81
Y140 H^{α}/P148 H^{α}	4.72, 5.26	Unclear	
V149 $H^{\gamma*}$/H^{α}	0.74, 3.94	0.74, 3.94	
EGF3 domain			
C160 H^{α}/C166 H^{α}	4.87, 5.31	4.87, 5.31	
N164 H^{α}/C150 H^{α}	4.47, 5.10	4.47, 5.10	4.48, 5.09

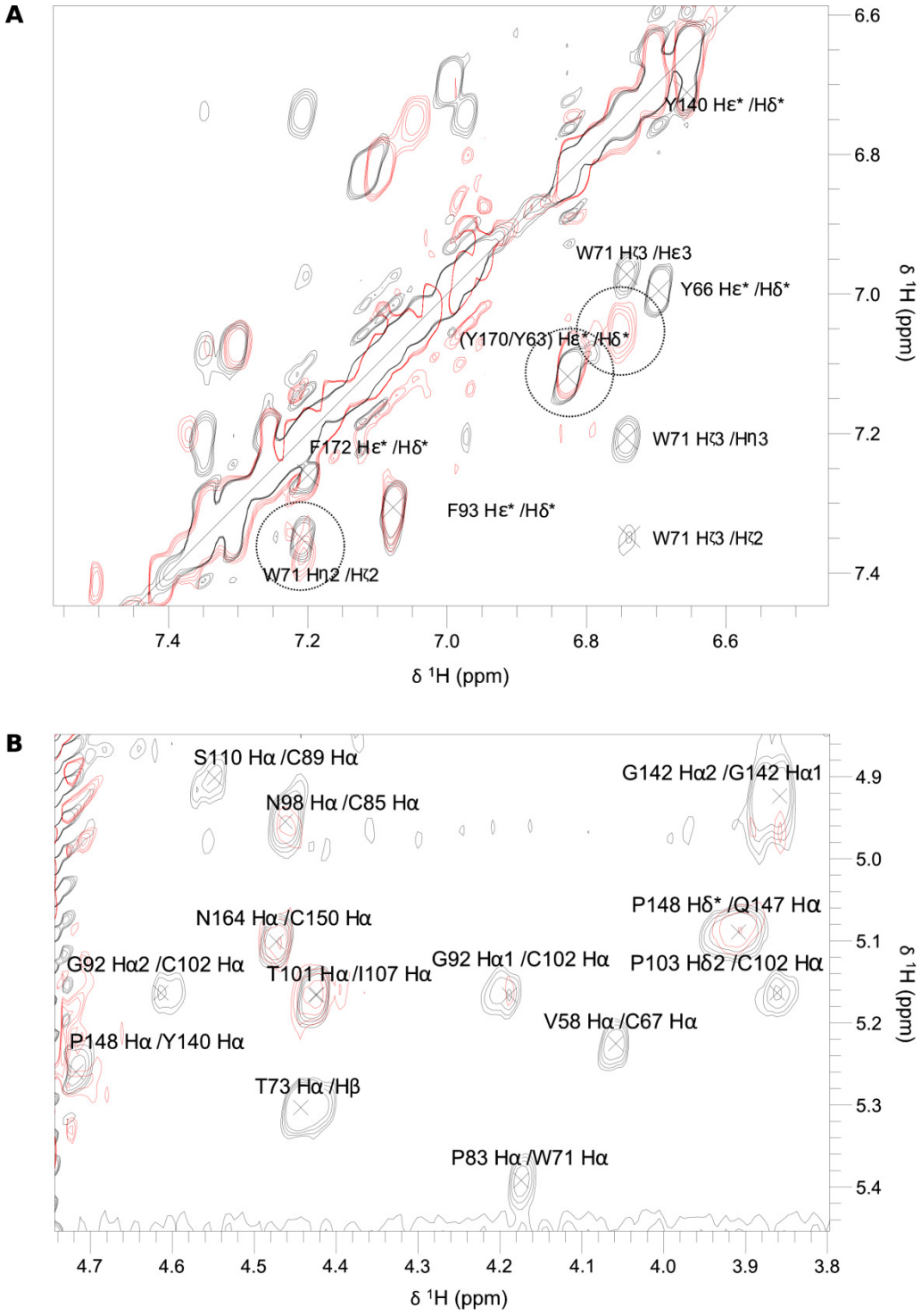


Figure 6.7 2D ^{1}H homonuclear NOESY spectrum of N57D His-FUN-EGF3: comparison with WT FUN-EGF3. The spectrum of N57D His-FUN-EGF3 was acquired with 0.1 mM protein in 100% D_2O at pH 5.4 and 25°C. A) Region of the spectrum showing cross-peaks between resonances in aromatic side chains. Contours for the N57D and WT spectra are coloured red and black respectively. Circles highlight peaks in the spectrum of the N57D protein that are absent from the WT spectrum, or were apparently shifted. B) Region of the spectrum showing predominantly $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peaks.

The 2D ^{1}H NOESY spectrum of P148S FUN-EGF3 also differed from the spectrum of WT FUN-EGF3. A summary of the analysis of the cross-peak markers is given in Table 6.2. One of most obvious differences was that the Y140 $H^{\varepsilon*}$-$H^{\delta*}$ cross-peak had shifted compared to the WT spectrum (Figure 6.8A). There was also a broad cross-peak in the P148S spectrum, which was not present in the WT spectrum and could not be assigned (circled in Figure 6.8A). Analysis of other regions of the spectrum did not conclusively identify which domains were correctly folded. Some of the cross-peaks had weak intensities (Figure 6.8B). Although the Y140-P148 $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ peak, a marker of folding at the EGF2-EGF3 interdomain interface, was not apparent in the P148S spectrum, there was a weak peak coinciding with the position of the N164-C150 $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peak in the WT spectrum. Moreover, the N98-C85 $^{1}H^{\alpha}$-$^{1}H^{\alpha}$ cross-peak appeared to be shifted, even though the residues involved were unlikely to be in proximity to the P148S substitution. The small number of markers of folding for the EGF2 domain in the NOESY spectrum also meant it was unclear whether this domain adopted the correct fold.

In order to investigate the fold of P148S FUN-EGF3 further, the protein was labelled with ^{15}N (Section 2.2). The final HPLC chromatogram of the ^{15}N-labelled protein showed a left shoulder peak that was larger than in the same chromatogram of the final unlabelled product (Figures 6.5C and 6.9A). It is possible that this other species was inadvertently removed from the unlabelled protein during the first HPLC purification step after refolding or during FPLC cation exchange purification. Therefore, the two fractions were collected and analysed separately. SDS-PAGE analysis showed that both fractions ran as a single band under non-reducing conditions, but the second fraction appeared to have slightly higher mobility than the first (Figure 6.9B). Mass spectrometry showed that the masses of both fractions were close to the expected molecular weight, although they differed by 4 Da (Table 6.3). It was unclear whether this difference arose from instrument error or a true difference in mass.

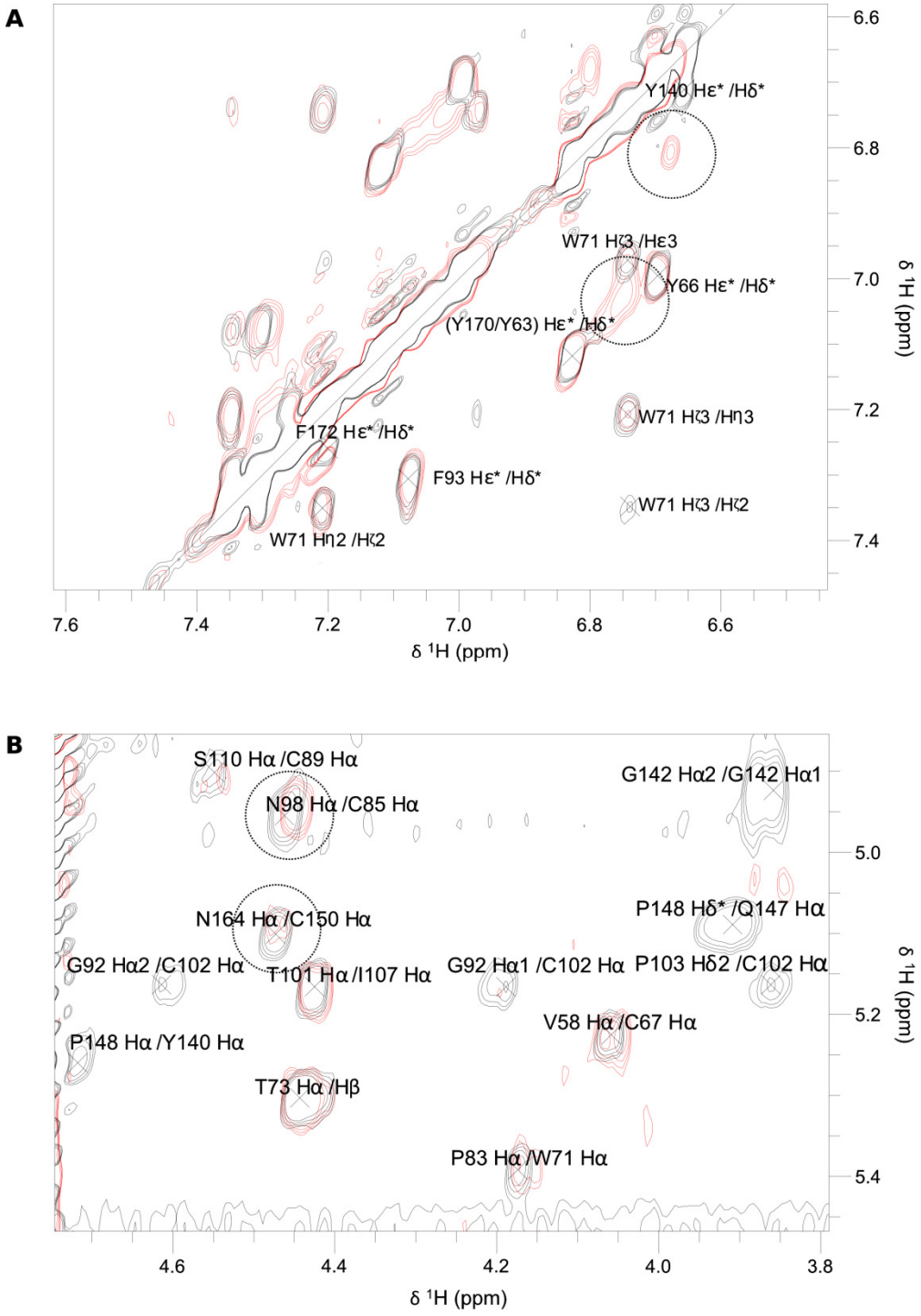


Figure 6.8 2D ^{1}H homonuclear NOESY spectrum of P148S FUN-EGF3: comparison with WT FUN-EGF3. The spectrum of P148S FUN-EGF3 was acquired with 0.5 mM protein in 100% D_2O at pH 5.4 and 25°C. A) Region of the spectrum showing cross-peaks between resonances in aromatic side chains. Contours for the P148S and WT spectra are coloured red and black respectively. Circles highlight peaks in the spectrum of the P148S protein that are absent from or shifted compared to the WT spectrum. B) Region of the spectrum showing mainly $^{1}H^{\alpha-1}H^{\alpha}$ cross-peaks with the same colour scheme as Part A.

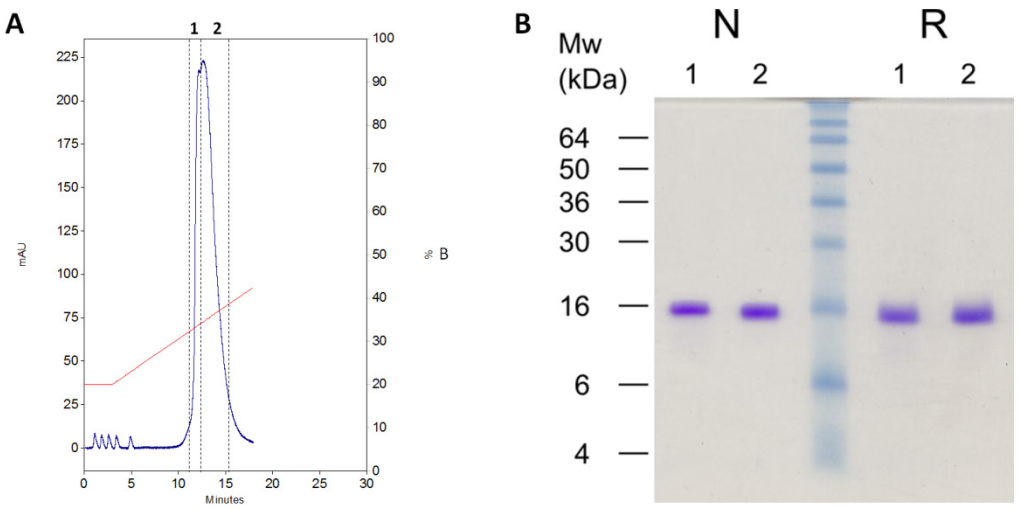


Figure 6.9 Final products of ^{15}N-labelled P148S FUN-EGF3 purification. A) HPLC chromatogram after factor Xa treatment and FPLC purification. Fraction boundaries are indicated by dotted lines and the two fractions are numbered. B) SDS-PAGE gel of the two fractions from the final HPLC purification, numbered as in Part A.

Table 6.3 Mass spectrometry results for ^{15}N-labelled P148S FUN-EGF3 fractions

Protein	Expected molecular weight (Da)	Measured molecular weight (Da)
Fraction 1	14259 (100% labelling)	14256
Fraction 2	14259 (100% labelling)	14260

HSQC spectra were acquired of both fractions (Figure 6.10A). They were not identical, but the majority of cross-peaks coincided. Slight differences suggested that were minor variations in folding, although without further peak assignment it was not possible to work out which residues were affected. Most importantly, the cross-peaks in the WT spectrum absent in the P148S spectra were consistent between fractions (comparison shown for fraction 1 in Figure 6.10B). Cross-peaks assigned to residues in the FUN-EGF1 region coincided with those in the WT FUN-EGF3 spectrum.

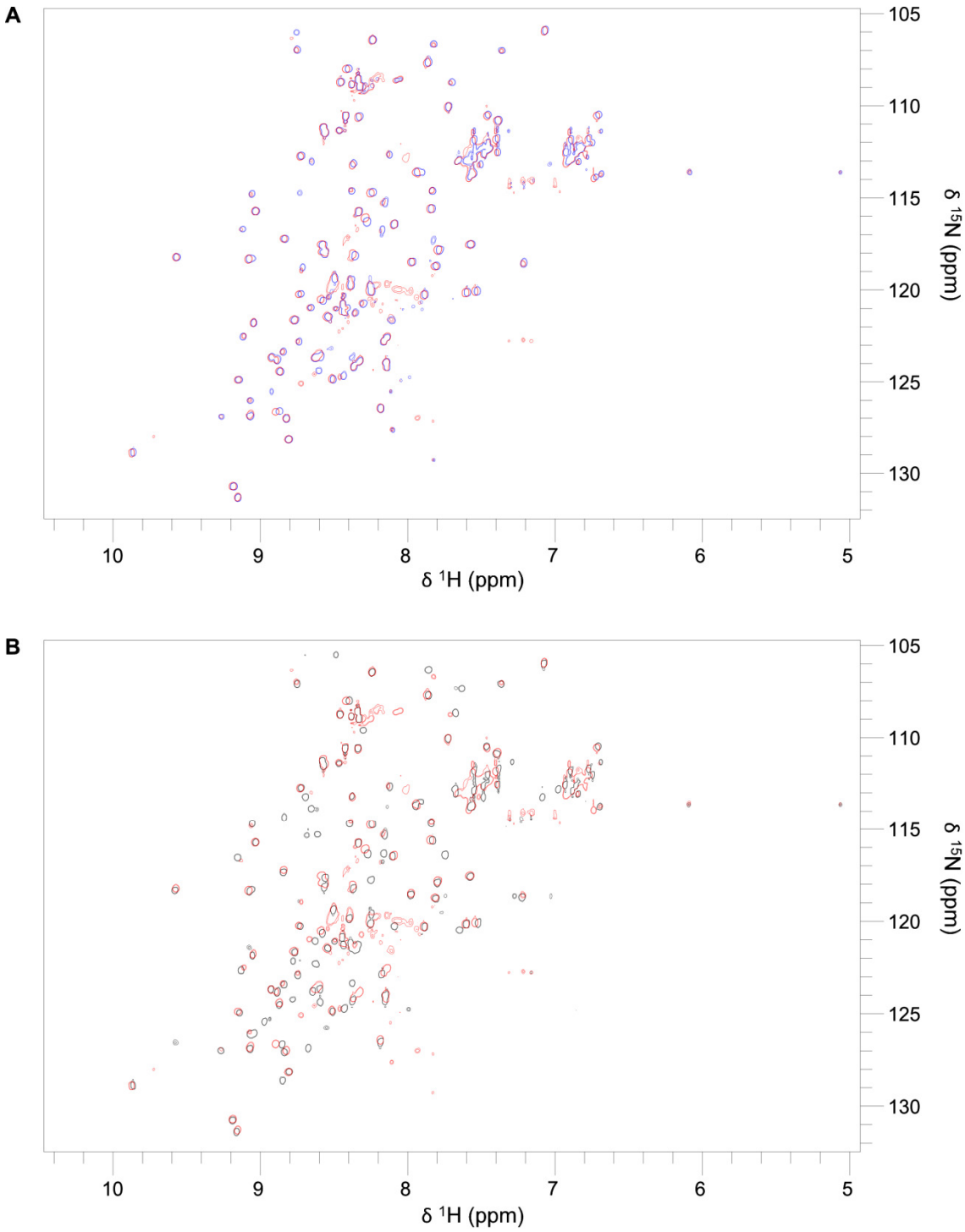


Figure 6.10 ^{1}H-^{15}N HSQC spectra of P148S FUN-EGF3. A) Superimposed spectra of the two final HPLC fractions of P148S FUN-EGF3. Contours of fractions 1 and 2 are coloured red and blue respectively. Both spectra were acquired using protein in 5% D_2O/95% H_2O (v/v) at pH 5.4 and 25°C. The concentrations of fractions 1 and 2 were 0.2 mM and 0.5 mM respectively. B) Spectrum of fraction 1 (red) superimposed on the spectrum of WT FUN-EGF3 (black).

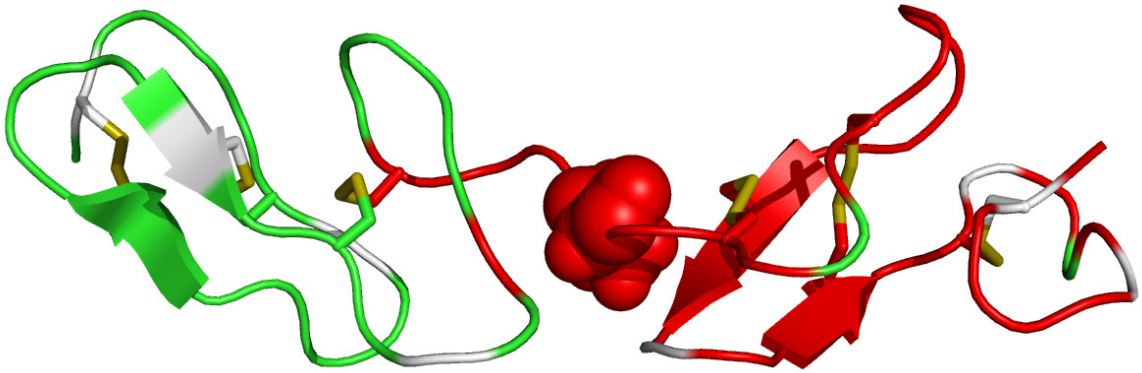


Figure 6.11 Residues affected by the P148S substitution in the EGF2-EGF3 region of FUN-EGF3. Atoms of residue P148 are represented by spheres. The backbone is coloured according to the presence of the cross-peaks from the WT HSQC spectrum in the P148S HSQC spectra: residues for which cross-peaks coincided are coloured green, while residues with absent cross-peaks are coloured red. White indicates that there was no information for that residue.

The major differences were in the EGF2-EGF3 region (Figure 6.11). Although the majority of cross-peaks assigned to residues in the EGF2 domain coincided with peaks in the WT spectrum, several arising from residues near the EGF2-EGF3 interdomain interface were absent. Furthermore, all peaks assigned to residues in the EGF3 domain, with the exception of two glycines, appeared to be absent from the spectra of both P148S FUN-EGF3 fractions. This suggested that the P148S substitution perturbed the fold of the EGF3 domain, but not the EGF2 domain. Moreover, it indicated that the effect was long-range, rather than localised to the EGF2-EGF3 interdomain interface.

6.6 Effects of MFS-associated substitutions on the N-C terminal interaction

N57D and P148S His-FUN-EGF3 proteins were tested for binding to cbEGF41-43-BirA-bio in the pull-down assay (Sections 2.4 and 3.7.1). Western blots detected WT, N57D and P148S His-FUN-EGF3 proteins when they were incubated with bead-immobilised cbEGF41-43-BirA-bio. The only noticeable difference was the greater intensity of the higher molecular weight species in the sample containing P148S His-FUN-EGF3. None of the fragments bound non-specifically to the beads alone. This result suggested that neither substitution significantly perturbed binding to cbEGF41-43-BirA. P148S FUN-EGF3 was also tested in the SPR assay (Section 3.7.2) and again, there was no significant difference between response levels when the same concentrations of WT and P148S FUN-EGF3 were injected over the flow cell with immobilised cbEGF41-43-BirA (data not shown).

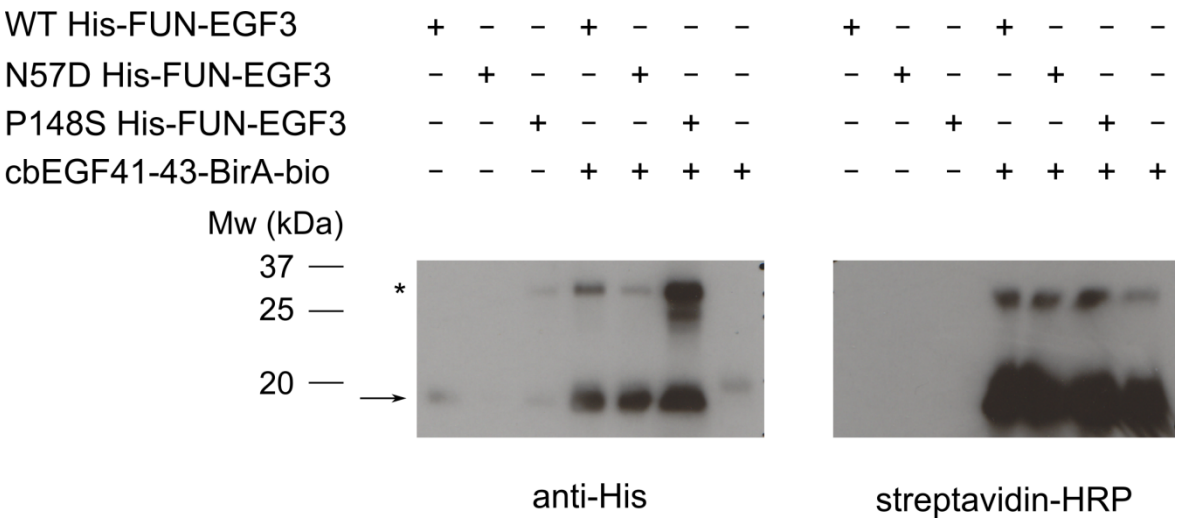


Figure 6.12 Results of the pull-down assay with N57D and P148S His-FUN-EGF3 proteins. Western blots are shown for supernatant samples after boiling beads from the pull-down in reducing SDS-PAGE buffer. The arrow indicates the bands running at the expected molecular weight of His-FUN-EGF3. The lower-mobility band is highlighted by *.

6.7 Effects of substitutions on protein secretion by MSU1.1 cells

6.7.1 pKG Nterm-Pro construct

The N57D, W71R and P148S substitutions were incorporated into a larger N-terminal human fibrillin-1 fragment, comprising the N-terminal sequence up to the end of the Pro-rich region (Nterm-Pro), including the N-terminal signal sequence and propeptide (Figure 6.13). The vector encoding the Nterm-Pro fragment (pKG-52 Nterm-Pro; hereafter referred to as pKG Nterm-Pro) was previously transfected into MSU1.1 fibroblast cells to create a stably expressing cell line (Kettle *et al.*, 2000). The expressed Nterm-Pro fragment was detected in the extracellular medium by Western blotting with a polyclonal antibody raised against a peptide corresponding to the Pro-rich region of human fibrillin-1 (Kettle *et al.*, 1999). This system was also used in several studies examining the effect of disease-linked substitutions on protein secretion, by adding other fragments of fibrillin-1 (Whiteman and Handford, 2003; Whiteman *et al.*, 2007) or fibulin-5 (Schneider *et al.*, 2010) C-terminally to the Pro-rich region in the Nterm-Pro fragment.

6.7.2 Site-directed mutagenesis of pKG Nterm-Pro and transfection into MSU1.1 cells

Mutations were incorporated into the pKG Nterm-Pro vector, kindly provided by Dr Pat Whiteman, using the QuikChange Lightning Kit (Section 2.1.10). The same primers were used as in mutagenesis of the pQE-30 FUN-EGF3 plasmid (Appendix B; Section 6.5.1). DNA sequencing confirmed the correct sequence of the N57D, W71R and P148S pKG Nterm-Pro constructs (Section 2.1.14). The WT and three mutant pKG Nterm-Pro vectors were individually transfected into MSU1.1 cells (Section 2.8.4). Transfected cells were cultured in medium containing puromycin, which selected against cells lacking the plasmid DNA. This resulted in clones of transfected cells.

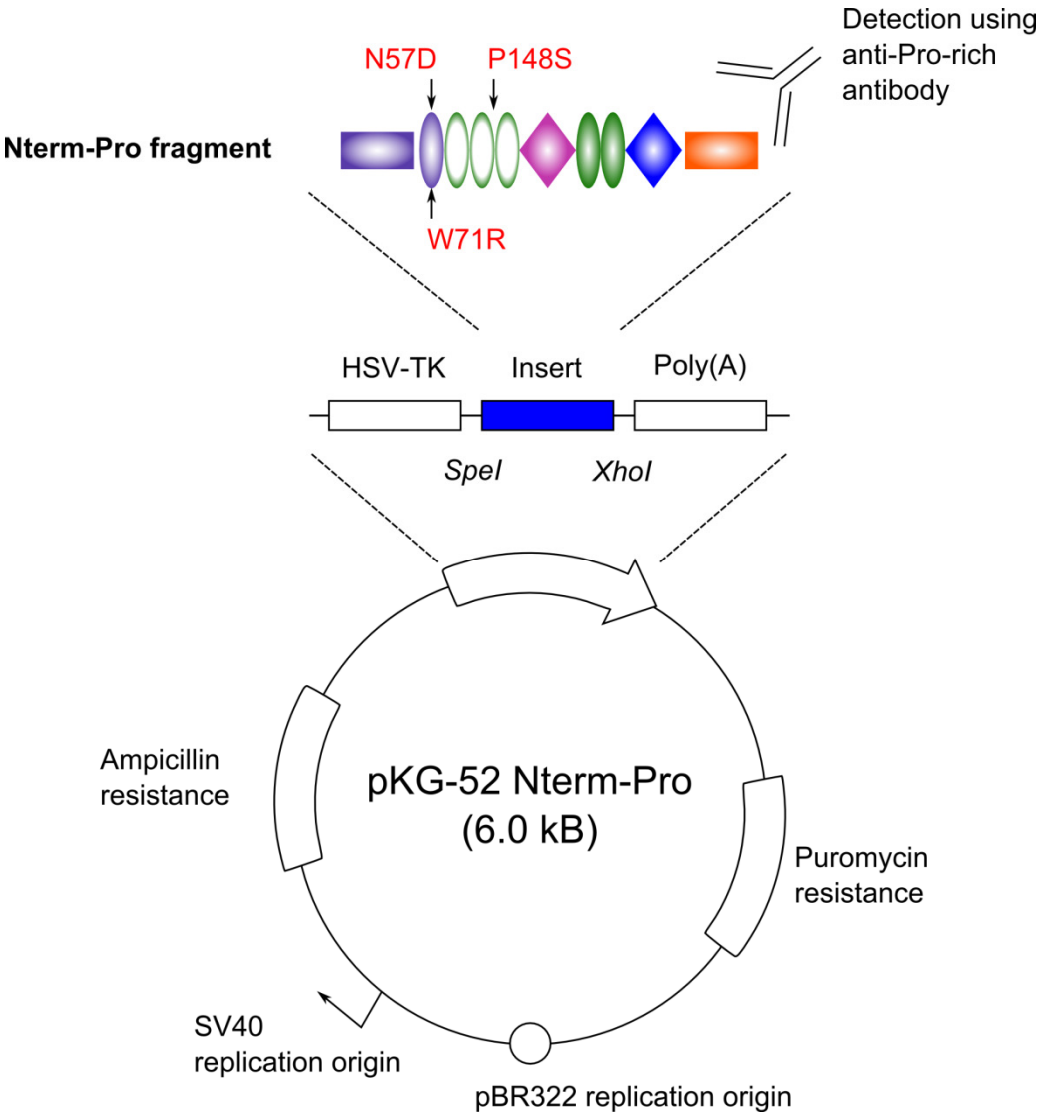


Figure 6.13. Expression and detection of the human fibrillin-1 Nterm-Pro fragment in MSU1.1 fibroblast cells. The pKG-52 plasmid contains replication origins that allow its replication in both bacterial (pBR322) and mammalian (SV40) cells. It also contains genes that confer ampicillin and puromycin resistance on cells, allowing selection of bacterial and mammalian cells, respectively, containing the plasmid. The insert DNA, comprising exons 1-11 of *FBN1*, encodes the Nterm-Pro fragment. It is expressed under the control of a Herpes Simplex Virus-Thymidine Kinase (HSV-TK) promoter and contains a 3' poly-adenine sequence needed for correct processing of the nascent RNA transcript into a mature mRNA. The locations of the MFS-associated substitutions N57D, W71R and P148S are indicated on the protein domain structure (domains illustrated as in Figure 1.2). A polyclonal antibody raised against the Pro-rich region allows detection of the fragment. Figure after Sarah Iqbal, D.Phil. Thesis, 2011. Vector described in Kettle *et al.*, 2000.

6.7.3 Nterm-Pro secretion by pools of MSU1.1 clones

Cell lysate and medium samples from pooled colonies of transfected cells were analysed by Western blotting (Section 2.8.5; Figure 6.15). As the expression level of the recombinant construct is likely to vary between clones, combining multiple clones should minimise differences in the total level of protein production between batches of transfected cells. Samples were collected every day for three days after splitting cells. Medium samples from both untransfected and transfected cells both contained a large molecular weight protein (>200 kDa) recognised by the anti-Pro-rich antibody, likely to be fibrillin-1 (Figure 6.15A). The band was faint, as the transfer efficiency of large proteins (>100 kDa) from the 10% SDS-PAGE gel to the nitrocellulose membrane was low. However, only the transfected cells produced a ~50 kDa protein, matching the expected size of the Nterm-Pro fragment.

There were differences in the relative amounts of recombinant protein in the cell medium and lysate samples between the cells transfected with different constructs (Figure 6.15B). In all cases, recombinant protein was detected in the medium and cell lysates two days after the cells had been passaged. However, the fraction of protein in the cell lysate samples was greater in the N57D- and W71R-transfected cells than in the WT-transfected cells. By contrast, the secretion profile of P148S-transfected cells appeared similar to the WT-transfected cells. The assay was repeated with the same batches of transfected cells and again showed an increased amount of protein in the cell lysate in the N57D- and W71R-transfected cells (Figure 6.16A).

Western blots were also probed with an anti-actin antibody, to exclude the possibility that the observed effects resulted from unequal loading of cell lysate samples. The quantity of recombinant protein in the medium in the N57D- and W71R-transfected cells was also reduced compared to the WT Nterm-Pro cells (Figure 6.16B). Protein levels in both the medium and cell

lysate samples of the P148S-transfected cells were higher compared to the WT, but the relative quantities were similar. In addition, protein in the medium samples appeared as two bands of similar size, most likely resulting from N-terminal processing by furin, as observed previously (Kettle *et al.*, 2000).

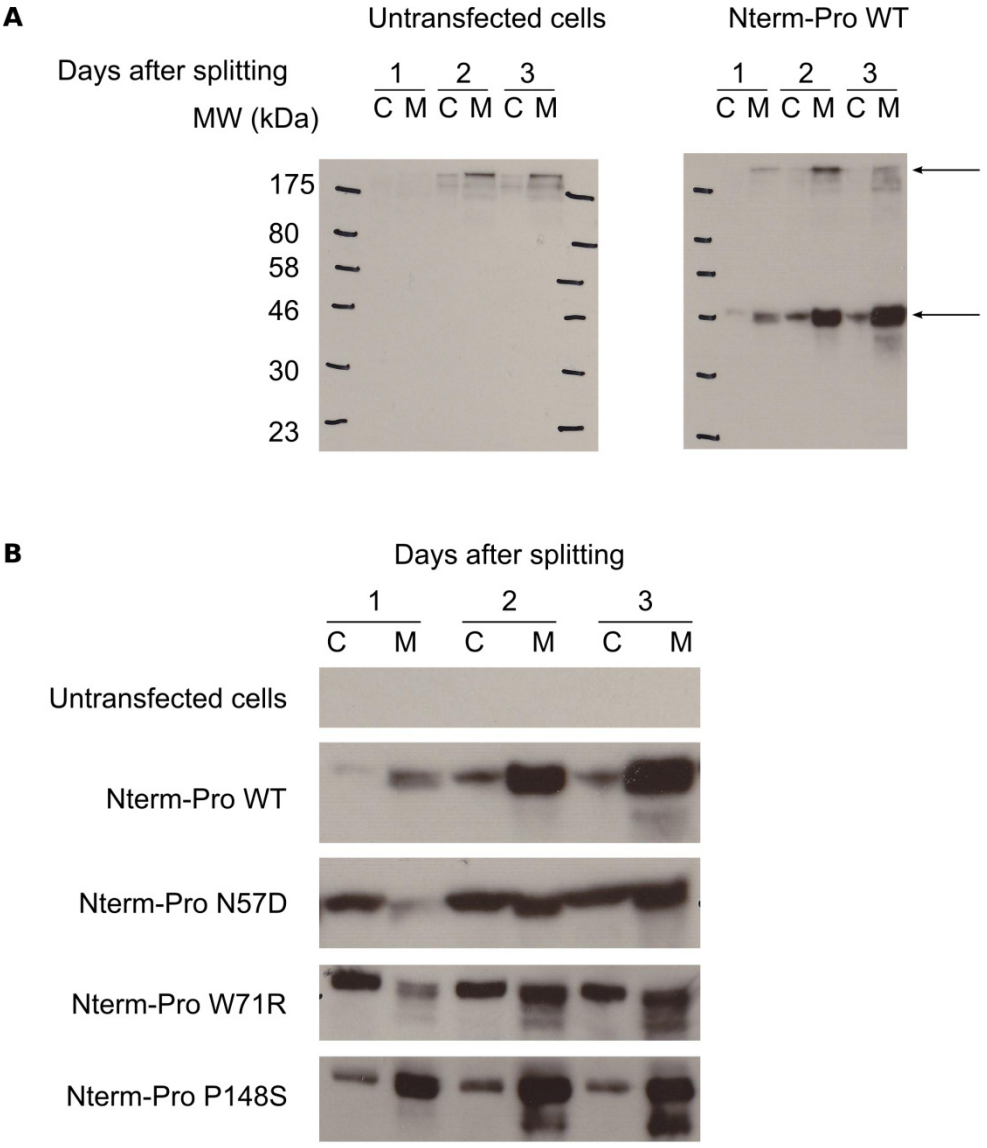


Figure 6.15 Secretion of Nterm-Pro fragments by MSU1.1 cells. A) Western blots of MSU1.1 cell lysate (C) and medium (M) samples, which were collected every day for three days after splitting. Protein detection was achieved using anti-Pro-rich antibody. The high molecular weight (>200 kDa) and lower molecular weight (~50 kDa) bands are indicated by arrows. B) Western blots of untransfected MSU1.1 cells, and cells transfected with WT, N57D, W71R and P148S pKG Nterm-Pro plasmids, showing the bands corresponding to the recombinant Nterm-Pro proteins.

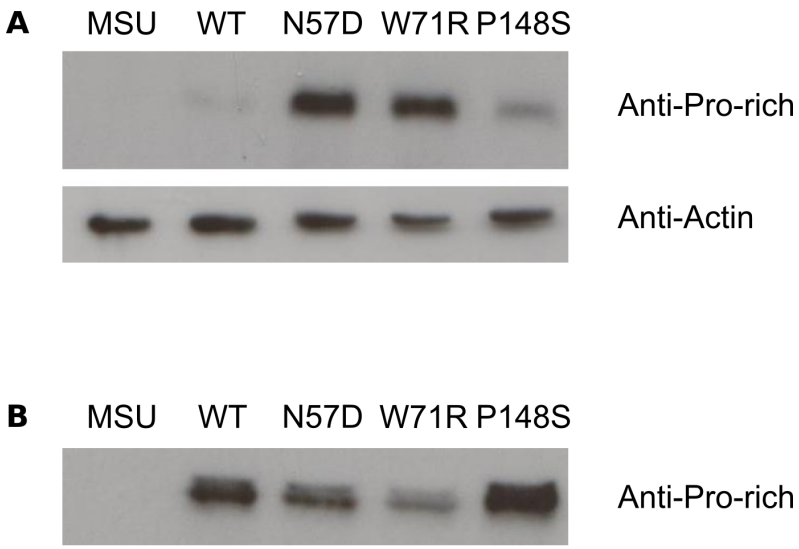


Figure 6.16 Partial apparent intracellular retention of Nterm-Pro constructs with N57D and W71R substitutions. A) Western blots of cell lysate samples from untransfected (MSU) and transfected MSU1.1 cells cultured for three days after splitting. Top: blot with anti-Pro-rich antibody, showing the Nterm-Pro bands; bottom: blot with anti-α-actin antibody, showing the band running at a molecular weight of ~42 kDa, corresponding to actin. B) Western blot of medium samples using the anti-Pro-rich antibody, showing the Nterm-Pro bands in samples from transfected cells.

6.8 Production and characterisation of the Y2639N cbEGF41-43-BirA fragment

6.8.1 Site-directed mutagenesis

The Y2639N substitution was introduced into the cbEGF41-43-BirA fragment by site-directed mutagenesis, as described for the FUN-EGF3 fragment (Sections 2.2.10 and 6.5.1). Primer sequences for mutagenesis are given in Appendix B. DNA sequencing confirmed the correct nucleotide change (Figure 6.17A).

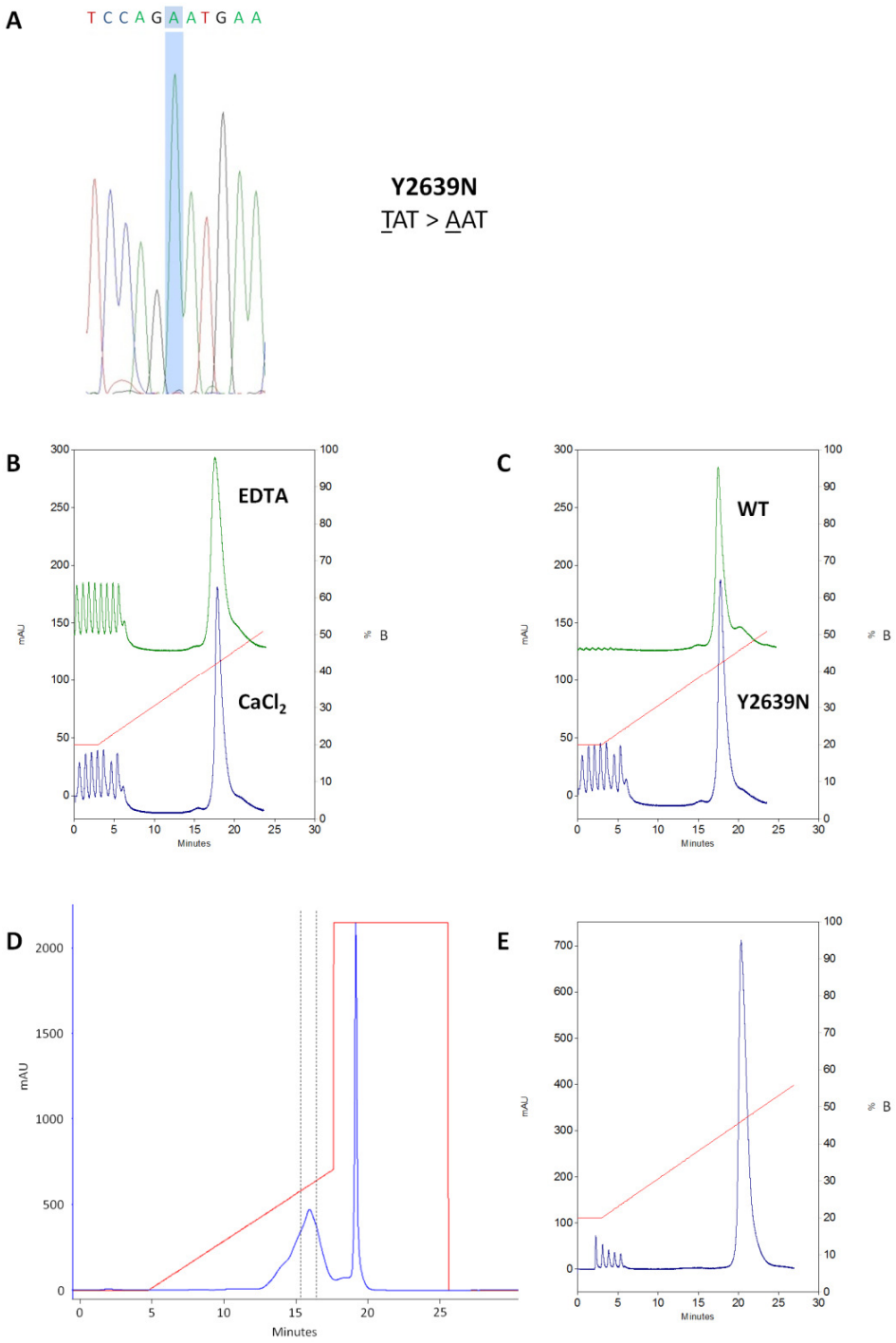


Figure 6.17 Production of Y2639N cbEGF41-43-BirA. A) Sequencing trace confirming the correct nucleotide substitution. B) Calcium-dependence of Y2639N cbEGF41-43-BirA refolding. HPLC chromatograms of the protein refolded in the presence of 50 mM $CaCl_2$ (blue) and 10 mM EDTA (green) are shown. C) HPLC chromatograms of refolded Y2639N-substituted protein (blue) and WT cbEGF41-43-BirA (green). D) FPLC chromatogram of refolded Y2639N cbEGF41-43-BirA treated with factor Xa. E) HPLC chromatogram of FPLC-purified protein. The dotted lines indicate collection boundaries of the major fraction.

6.8.2 Protein production

The purification method for WT cbEGF41-43-BirA (Sections 2.2 and 3.4.2.3) was also used for the Y2639N-substituted protein. As for the unmodified fragment, the HPLC elution profile was slightly broader when the protein was refolded in the presence of 10 mM EDTA, compared with 50 mM $CaCl_2$ (Figures 3.7 and 6.17B). WT and Y2639N cbEGF41-43-BirA were refolded concomitantly with the same reagents and had very similar elution profiles (Figure 6.17C). The FPLC (Figure 6.17D) and final HPLC (Figure 6.17E) elution profiles of factor Xa-treated Y2639N cbEGF-41-43-BirA were also very similar to the unmodified protein (Figure 3.7, Section 3.4.2.3). Y2639N cbEGF41-43-BirA was biotinylated using the BirA enzyme for use in the pull-down assay, as described for the WT protein (Section 2.2.9).

6.8.3 Characterisation by SDS-PAGE and mass spectrometry

SDS-PAGE suggested that the Y2639N cbEGF41-43-BirA and cbEGF41-43-BirA-bio proteins were homogeneous under reducing and non-reducing sample conditions (Figure 6.18). Mass spectrometry confirmed the expected sizes, although the measured mass of biotinylated protein was 6 Da lower than expected.

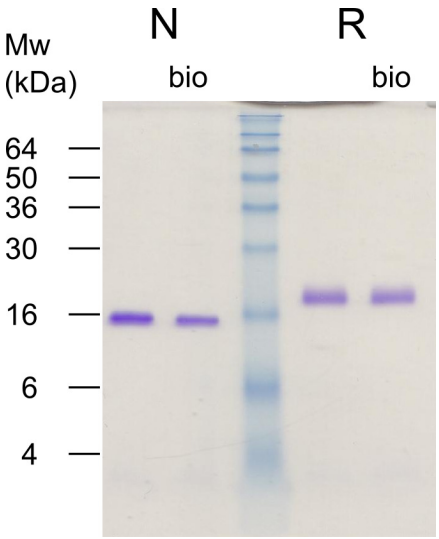


Figure 6.18 16% SDS-PAGE gel of Y2639N cbEGF41-43-BirA and cbEGF41-43-BirA-bio. Samples were prepared under non-reducing (N) and reducing (R) conditions. The biotinylated protein is marked 'bio'.

Table 6.4 Mass spectrometry results for Y2639N cbEGF41-43-BirA

Protein	**Expected molecular weight (Da)**	**Measured molecular weight (Da)**
Y2639N cbEGF4143-BirA	15256	15254
Y2639N cbEGF41-43-BirA-bio	15482	15288

6.9 Interaction of Y2639N cbEGF41-43-BirA-bio with His-FUN-EGF3

The interaction of Y2639N cbEGF41-43-BirA-bio with His-FUN-EGF3 was compared with the WT in the pull-down assay (Sections 2.4 and 3.7.1). His-FUN-EGF3 bound equally well to Y2639N and WT cbEGF41-43-BirA-bio (Figure 6.19). The SPR assay also showed that the Y2639N substitution had no effect on the binding of FUN-EGF3 to cbEGF41-43-BirA (data not shown).

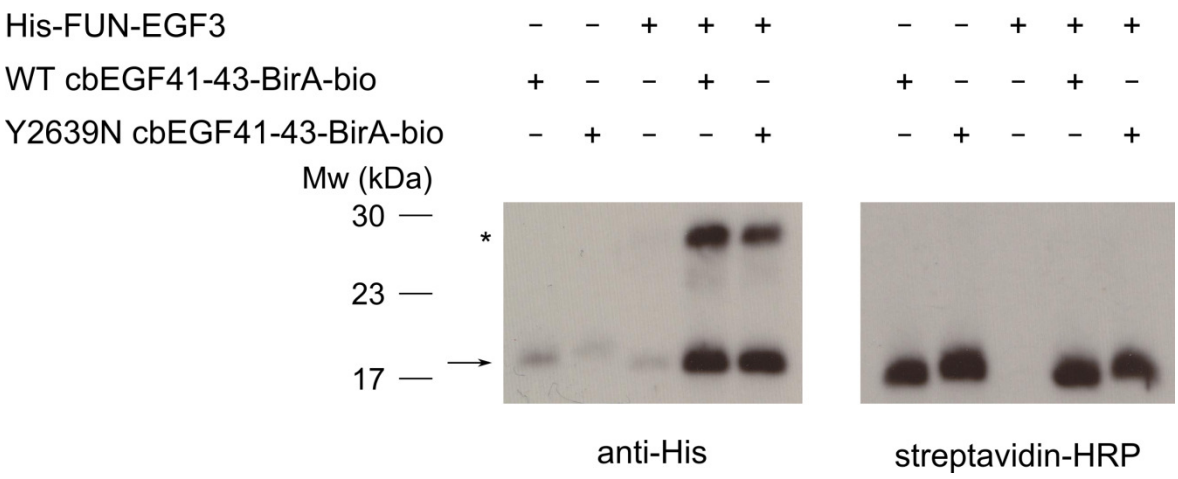


Figure 6.19 Pull-down results for the interaction of WT His-FUN-EGF3 with Y2639N cbEGF41-43-BirA-bio. Western blots are shown for supernatant samples after boiling beads from the pull-down in reducing SDS-PAGE sample buffer. The arrow indicates bands running at the expected molecular weight of His-FUN-EGF3, while * indicates the lower mobility bands detected in the anti-His blot.

6.10 Discussion

The structure of the FUN-EGF3 fragment (determined in Chapter 4) allowed the effects of some MFS-linked missense mutations affecting the FUN-EGF3 region of human fibrillin-1 to be predicted. For the cbEGF41-43 region, a homology model and sequence conservation data permitted the potential effects of MFS-associated substitutions to be assessed. Three residue substitutions in the FUN-EGF3 region associated with MFS or MFS-like phenotypes were studied: N57D, W71R and P148S. Site-directed mutagenesis was employed to introduce these substitutions into the FUN-EGF3 fragment, produced by prokaryotic expression and *in vitro* refolding, and the Nterm-Pro fragment, which was expressed and secreted by MSU1.1 fibroblast cells. FUN-EGF3 fragments were characterised by HPLC, SDS-PAGE, mass spectrometry and NMR spectroscopy, and were tested for binding to cbEGF41-43-BirA in the pull-down assay. Secretion of the Nterm-Pro fragments was analysed by Western blotting of cell lysate and extracellular

medium fractions. The Y2639N MFS-associated substitution was also incorporated into the cbEGF41-43-BirA fragment and the protein was tested in the pull-down assay.

While N57 in the FUN domain does not form significant contacts with residues in the EGF1 domain, its side chain packs against the aromatic ring of W71 (Section 4.11). In other species' fibrillins, or in EGF-EGF domain pairs in other proteins, the asparagine residue may be replaced by histidine or arginine, both of which contain basic groups (Sections 4.11 and 4.13). Therefore, the replacement of N57 with a negatively-charged residue might destabilise its packing with W71. A substitution of this type, N57D, is caused by a missense mutation found in a patient with classical MFS (Chung *et al.*, 2009).

The N57D FUN-EGF3 fragment was produced and refolded *in vitro*. It was possible to purify a species with apparently similar properties to WT FUN-EGF3. However, the yield of the N57D-substituted protein was substantially lower than that of the WT. Moreover, analysis by ^{1}H NMR suggested that the fold of the FUN domain was different in the N57D-substituted protein compared to WT FUN-EGF3. It was likely that the purified species represented a fraction of material that refolded into a stable form, while the majority of the protein molecules failed to fold. An increased fraction of intracellular N57D Nterm-Pro compared to the WT protein was consistent with this hypothesis. Combined, the results indicated that residue N57 is important for folding of the N-terminal domains and that the N57D substitution causes misfolding.

The aromatic side chain of W71 forms extensive contacts with several residues in the EGF1 domain, as well as others in the FUN domain (Section 4.11). Therefore, the W71R substitution, found in a patient with some MFS-like features (Sakai *et al.*, 2006), would be expected to disrupt the FUN-EGF1 interface by perturbing interdomain packing. Attempts to refold a FUN-EGF3 fragment with the W71R substitution were unsuccessful, as the protein formed disulphide-

bonded aggregates. Favourable interactions between residues at the interdomain interface are likely to stabilise conformations that allow the correct formation of disulphide bonds in the WT construct. These conformations would be disrupted by the replacement of W71 by arginine, resulting in the formation of incorrect disulphide bonds and an increased frequency of intermolecular reactions between cysteine side-chain thiol groups. An apparent partial increase in intracellular W71R Nterm-Pro in the secretion assay, compared with the WT construct, was consistent with the *in vitro* result.

Although the MFS-associated P148S substitution (Stheneur *et al.*, 2009) affects a conserved residue at the EGF2-EGF3 interdomain interface, it is not necessarily easy to predict its effects. The equivalent residue in the HHIP and WIF-1 EGF-EGF domain pairs (Section 4.13) is alanine instead of proline. It is possible that the substitution of P148 by a residue with a small side chain, such as serine, could be tolerated and might not cause misfolding. The P148S FUN-EGF3 fragment was refolded *in vitro* and a protein with WT-like properties was purified. Closer investigation by NMR using ^{15}N-labelled protein suggested that the P148S substitution had long-range effects on folding. Although the EGF2 domain appeared to adopt a WT-like fold, the profile of the EGF3 domain was completely different. However, this apparent misfolding did not significantly affect the yield of refolding.

In the secretion assay, the P148S Nterm-Pro fragment behaved like the WT protein, with no obvious increase in the intracellular protein fraction. It is possible that the hyb1 domain, which was missing from the FUN-EGF3 fragment, stabilised the EGF3 domain in the Nterm-Pro fragment. Previous studies have shown the importance of domain context in the *in vitro* refolding of fragments containing disease-associated substitutions. For example, the G1127S substitution in the cbEGF13 domain causes misfolding of the isolated domain or cbEGF13, but not cbEGF14, in the cbEGF13-cbEGF14 domain pair. However, the cbEGF12-cbEGF13 pair with the G1127S

substitution has a WT-like fold (Whiteman *et al.*, 1998; Whiteman *et al.*, 2001). In a similar secretion assay to the one used here, the G1127S substitution did not cause intracellular retention (Whiteman and Handford, 2003). In the case of the P148S substitution, effects on folding may only be local when the hyb1 domain is present. This warrants further investigation in a construct comprising EGF2-cbEGF1, which has been shown to adopt a native-like fold (Ian Robertson, personal communication).

Further work is needed to validate the results of the secretion assay. Comparisons of RNA levels by reverse-transcriptase PCR have previously been used to check that differences observed in the assay are not accounted for by expression levels (Schneider et al., 2010; Whiteman and Handford, 2003). Furthermore, the intracellular location of the protein retained in the cell lysate fraction could not be determined using this assay. Previous studies have demonstrated that modified proteins are retained in the endoplasmic reticulum, by degrading particular N-linked glycans with enzymes and identifying interactions with chaperones (Whiteman and Handford, 2003). The Nterm-Pro region does not contain any sites for N-linked glycosylation and so a larger fragment would be needed to study the effects of the N57D, W71R and P148S substitutions on intracellular trafficking by the former method.

None of the residue substitutions studied here disrupted the N-C interaction in the pull-down assay. His-FUN-EGF3 fragments with N57D or P148S substitutions were still pulled down by cbEGF41-43-BirA-bio in the same manner as the WT protein. This might suggest that the FUN and EGF3 domains are not required for binding, and that the P148S substitution has a different mechanism of pathogenesis, such as the disruption interactions with microfibril-associated proteins. However, these results must be interpreted cautiously as no single residue substitution has yet been found that abolishes binding. In addition, the Y2639N substitution in the cbEGF41-43-BirA fragment did not have any effect on the interaction. Likewise, this does necessarily mean

that these MFS mutations do not affect microfibril assembly or function. An alternative approach for studying the interaction is discussed in Section 7.4.

In summary, although none of the disease-linked mutations studied here affected the N-C interaction, possible mechanisms of pathogenesis were identified for the N57D and W71R substitutions. Protein misfolding, resulting in intracellular retention, may reduce the amount of fibrillin secreted into the extracellular medium and therefore reduce the deposition of the protein into microfibrils in the ECM.

Chapter 7

Discussion, conclusions
and future work

7.1 Summary

The aim of this thesis was to characterise the interaction between the N- and C-terminal regions of human fibrillin-1, a proposed initial step in microfibril assembly (Section 1.10). This involved the production of the recombinant fragments corresponding to the minimal regions required for the interaction, FUN-EGF3 and cbEGF41-43. These fragments were then characterised by biochemical methods and NMR spectroscopy, and were shown to interact specifically. NMR spectroscopy was also employed to determine the structure of the FUN-EGF3 fragment. On the basis of this structure, smaller N-terminal fragments and FUN-EGF3 proteins with single substitutions were designed, produced and tested in interaction assays with the cbEGF41-43-BirA fragment. Moreover, the structure suggested a basis for the effects of three MFS-associated missense mutations. The consequences of the resulting residue substitutions were tested in fragments produced *in vitro* and in a cell-based secretion assay.

7.2 Production, characterisation of N- and C-terminal fibrillin-1 fragments

Results of previous *in vitro* experiments suggested that head-to-tail interactions between fibrillin monomers were involved in their assembly into microfibrils (Lin *et al.*, 2002; Marson *et al.*, 2005). The minimal regions of human fibrillin-1 required for the interaction were subsequently narrowed down to the FUN-EGF3 region at the N-terminus (El-Hallous *et al.*, 2007) and cbEGF41-43 at the C-terminus (Hubmacher *et al.*, 2008). Here, fragments corresponding to these regions were produced using an established prokaryotic expression, purification and *in vitro* refolding method (Knott *et al.*, 1996). This method was previously employed to produce fragments of human fibrillin-1 (Jovanovic *et al.*, 2007; Lee *et al.*, 2004), Notch-1 (Cordle *et al.*, 2008b; Hambleton *et al.*, 2004) and Jagged-1 (Cordle *et al.*, 2008a), which are involved in extracellular protein-protein interactions. These fragments retained the functional properties of the full-length proteins.

FUN-EGF3 and cbEGF41-43-BirA fragments were successfully produced and characterised (Chapter 3). A C-terminal BirA tag was included in C-terminal fibrillin-1 fragment, to facilitate its immobilisation by the streptavidin-biotin interaction. Previous studies have examined calcium binding to assess the structural integrity of cbEGF-containing fibrillin fragments, as this property is dependent on the correct formation of disulphide bonds (Handford *et al.*, 1990). The lack of calcium-binding consensus sequences in the FUN-EGF3 region necessitated alternative means of characterisation. 1D and 2D spectra of the corresponding fragment showed elements of folding (Section 3.5), but definitive evidence came from structure determination (Chapter 4).

Characterisation of the cbEGF41-43-BirA fragment was more challenging because of its low solubility, particularly in the presence of calcium ions. Calcium-mediated protection from proteolysis (Section 3.6.1) and calcium-specific, concentration-dependent peak shifts in NMR spectra (Section 3.6.2) were demonstrated. However, assignments for peaks in the NMR spectra could not be obtained, so it was not clear if all three domains adopted native-like folds. In addition, the low pH conditions used in the experiment meant it was unlikely that the calcium-binding sites were saturated, even when $CaCl_2$ was added to a final concentration of 100 mM. Domain-specific calcium-binding information might be obtained in future through N-terminal sequencing of fragments from limited proteolysis and studying calcium-binding-deficient cbEGF41-43-BirA fragments, as previously carried out for fibrillin-1 (Whiteman *et al.*, 2007) and Notch-1 fragments (Cordle *et al.*, 2008b). Mutations resulting in the loss of putative calcium-binding ligands, such as the MFS-associated E2610K substitution in the cbEGF42 domain (Section 1.9.6) (Liu *et al.*, 1997), could be introduced into the fragment.

7.3 Interaction of N- and C-terminal fibrillin-1 fragments

The FUN-EGF3 and cbEGF41-43-BirA fragments interacted in pull-down and SPR assays, again suggesting that they had native-like structures (Section 3.7). A pull-down assay was developed using bead-immobilised cbEGF41-43-BirA and His-tagged FUN-EGF3 in solution. This allowed binding experiments to be performed under more physiologically relevant conditions than with immobilised FUN-EGF-BirA and His-cbEGF41-43 in solution. Specificity was demonstrated through the inclusion of negative control fragments, which did not bind significantly to FUN-EGF3 or cbEGF41-43-BirA. In the SPR assay, the interaction appeared to be less specific, as a concentration-dependent response was also observed when FUN-EGF3 was injected over the flow cell containing the immobilised negative control fragment cbEGF32-34-BirA.

In contrast to earlier studies that examined larger N- and C-terminal fragments, the interaction in the pull-down was calcium-independent. An interaction between N- and C-terminal 'half' fragments was completely abolished by EDTA (Lin *et al.*, 2002), while smaller fragments showed partially calcium-dependent binding (Marson *et al.*, 2005). However, there are no published studies showing that the interaction of the minimal regions is dependent on calcium. Domains outside the cbEGF41-43 region might be responsible for the observed calcium dependence. Nonetheless, without evidence of calcium binding by all three domains in cbEGF41-43-BirA (Sections 3.6.2 and 7.1), partial misfolding of the protein, resulting in a calcium-independent interaction with FUN-EGF3, cannot be ruled out.

Aggregation of the N- and C-terminal regions is likely to have implications for the interaction. The same study that identified cbEGF41-43 as the minimal region for binding to the N-terminus also found that disulphide-dependent aggregation enhanced the apparent affinity of the interaction (Hubmacher *et al.*, 2008). Deletion of the cbEGF41-43 region from the C-terminal half fragment

also reduced the proportion of multimers. Although direct comparisons cannot be made with the different fragments and assays used here, aggregation may affect interpretation of the results. It is possible that clustering of cbEGF41-43-BirA molecules on the beads had a similar effect to aggregation, enhancing the apparent binding affinity for FUN-EGF3. The oligomeric state of both fragments is still unknown and preliminary investigations into the behaviour of cbEGF41-43-BirA in solution by size-exclusion chromatography were unsuccessful (data not shown). This requires further study, for example by dynamic light scattering or analytical ultracentrifugation, including the effect of different conditions on aggregation.

7.4 Structure of the FUN-EGF3 fragment

Before this study, there were no high-resolution structures of the fibrillin N-terminal domains. NMR spectroscopy was used to determine the structure of the FUN-EGF3 fragment and investigate its dynamics (Chapter 4). This provided insights into the novel fold of the FUN domain, structures of the non-calcium-binding EGF-like domains and the interdomain interfaces. The FUN domain is unique to the fibrillins and LTBPs, although it is not present in all LTBP isoforms (Robertson *et al.*, 2011). In contrast to EGF-like domains, it has four cysteine residues, which form two disulphide bonds in a C1-C3, C2-C4 arrangement. However, the C-terminal portions of the FUN and EGF-like domains are similar (Section 4.10). The FUN domain is a more dynamic structure than the EGF-like domain, with a flexible N-terminal segment and a flexible loop between the first and second cysteine residues (Section 4.4). Part of this loop region (Y63-Y66) is absent from the LTBP-1 FUN-like domain (Section 1.3.5), suggesting a possible functional role.

The structure also showed how the domains in the FUN-EGF3 region pack together. Interdomain interactions determine the overall conformation of multi-proteins. Previous studies of fibrillin fragments have focused mainly on calcium-dependent interdomain interfaces. Calcium binding

rigidifies cbEGF-cbEGF domain pairs by stabilising the interdomain interfaces (Werner *et al.*, 2000) and calcium affinity is also likely to regulate interface stability in TB-cbEGF and hyb-cbEGF domain pairs (Jensen *et al.*, 2005; Jensen *et al.*, 2009; Yuan *et al.*, 2002). Until now, relatively little was known about calcium-independent interfaces in fibrillin. Structures of cbEGF-TB (Lee *et al.*, 2004) and cbEGF-hyb interfaces (Jensen *et al.*, 2009) have been determined by X-ray crystallography, but their dynamics have not been characterised.

The FUN-EGF1 and EGF2-EGF3 interdomain interfaces are structurally homologous, but they are distinct from others in fibrillin. Like the cbEGF-cbEGF, TB-cbEGF and hyb-cbEGF interfaces, an aromatic residue (W71 or Y140) in the N-terminal domain of the pair (FUN or EGF2) packs with residues in the C-terminal domain (EGF1 or EGF3). However, unlike the previously characterised interfaces, two of the principal packing residues in the C-terminal domain are conserved prolines (Section 4.11). An asparagine residue from the N-terminal domain (N57 or N125) packs against the other face of the aromatic ring, a feature also not seen in other fibrillin interdomain interfaces. Despite the differences, the FUN-EGF1 and EGF2-EGF3 pairs adopt rod-like structures, similar to cbEGF-cbEGF pairs (Figure 7.1). The surface area buried by the interfaces is also similar, but the tilt angle of the EGF2-EGF3 pair is larger than in the cbEGF12-cbEGF13 or cbEGF32-33 pair (Section 4.12). These rigid interfaces help to maintain the structural integrity of fibrillin. This is important because microfibrils are subjected to significant forces as connective tissues are stretched and compressed. Moreover, the rod-like structure of multi-domain arrays has implications for microfibril organisation (Section 7.7).

EGF-EGF pairs in other extracellular proteins (Hedgehog interacting protein [HHIP], Wnt inhibitory factor and Jagged-1) contain elements of the packing motifs found in the fibrillin FUN-EGF3 region (Section 4.13). The similarity of the interdomain interfaces, particularly between the fibrillin-1 EGF2-EGF3 and HHIP EGF2-EGF3 pairs, supports the assertion that the FUN-EGF3 fragment adopts

a native-like structure. Despite the similarities, there is variation between the EGF-EGF pairs in different proteins. In particular, the conserved proline residues before the first cysteine of EGF1 and EGF3 are only present in fibrillin. This might indicate that the FUN-EGF1 pair arose from the duplication of exons encoding an EGF-EGF pair, one of which subsequently evolved into a FUN-EGF pair. However, without the observation of an evolutionary intermediate protein containing an EGF domain instead of the FUN domain, this is speculative.

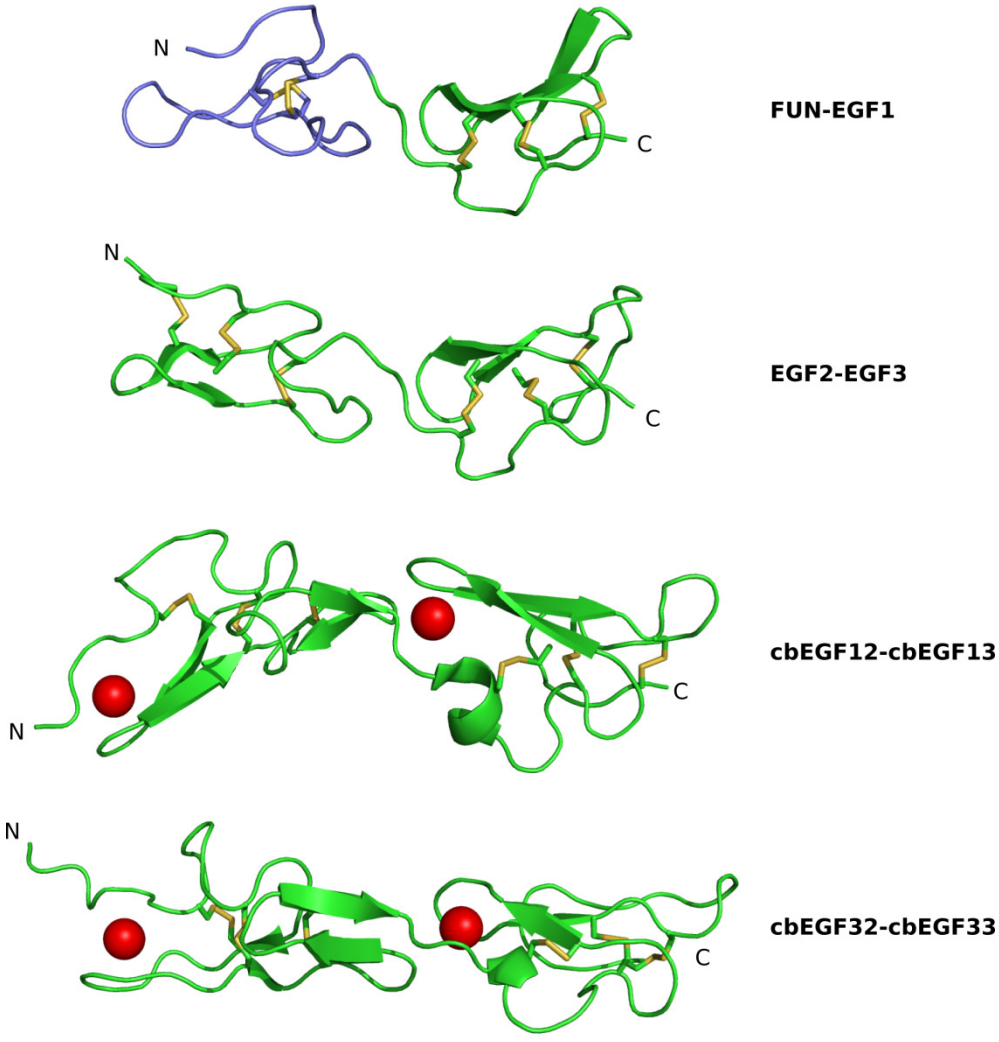


Figure 7.1 Structures of FUN-EGF, EGF-EGF and cbEGF-cbEGF domain pairs in human fibrillin-1. Cartoon representations of energy-minimised average structures are shown for the FUN-EGF1, EGF2-EGF3, cbEGF12-cbEGF13 (Smallridge *et al.*, 2003) and cbEGF32-cbEGF33 (Downing *et al.*, 1996) domain pairs. The FUN domain is coloured lilac, while the EGF-like and cbEGF domains are coloured green. Calcium ions are shown as red spheres. All four domain pairs adopt rod-like structures, but there is variation in the interdomain tilt angle.

While the FUN-EGF1 and EGF2-EGF3 interfaces are well defined, there is no observable interface between the EGF1 and EGF2 domains. The poor definition of the EGF1-EGF2 interface in the structure was due to the lack of experimental restraints. This is consistent with the results of the heteronuclear NOE experiment, which showed that the EGF1-EGF2 interdomain linker sequence (G112-H118) was flexible (Section 4.4). Analysis of secondary chemical shifts by TALOS+ also suggested that these residues were unstructured (Section 4.6.2) and separate alignment tensors were needed to define the RDC restraints for residues in the FUN-EGF1 and EGF2-EGF3 regions (Section 4.8.6). In addition, the key packing residues found at the FUN-EGF1 and EGF2-EGF3 interfaces are absent from the EGF1-EGF2 region and the length of the EGF1-EGF2 linker sequence is not conserved (Section 4.11). This is only the second flexible interdomain sequence to have been identified in fibrillin-1 − the first being the TB6-cbEGF32 linker (Yuan *et al.*, 2002).

7.5 Dissection of the N-C terminal interaction

Although the minimal interacting regions of the N- and C-termini are well established, specific regions or residues involved in the interaction have not yet been identified. Here, attempts were made to narrow down the N-terminal domains required for the interaction and target individual residues by site-directed mutagenesis. The apparent absence of an interface between the EGF1 and EGF2 domains of FUN-EGF3 suggested that smaller fragments corresponding to these regions would refold independently. These fragments were produced and characterised, although the EGF2-EGF3 domain pair required a C-terminal BirA tag to refold efficiently (Sections 5.2 and 5.3).

FUN-EGF1 and EGF2-EGF3-BirA behaved differently from FUN-EGF3 in the pull-down and SPR assays. A greater quantity of His-tagged FUN-EGF1 or EGF2-EGF3-BirA than FUN-EGF3 had to be used to observe an interaction with bead-immobilised cbEGF41-43-BirA in the pull-down assay (Section 5.4.1). In the SPR assay, FUN-EGF1 and EGF2-EGF3-BirA generated much lower responses

than FUN-EGF3 at the same concentration when flowed over a flow cell containing immobilised cbEGF41-43-BirA (Section 5.4.2). However, aggregation of FUN-EGF3 might have been responsible for the observed differences from the smaller fragments. Therefore, it cannot be concluded that both halves of FUN-EGF3 bind cbEGF41-43 until the oligomeric states of all three N-terminal fragments have been characterised.

5 residues in FUN-EGF3 were replaced with alanine by site-directed mutagenesis (Section 5.6). They were selected using several criteria: evolutionary conservation, including in species lacking LTBP and MAGP proteins that may also bind the FUN-EGF3 region; they were unlikely to be structurally important and so their replacement would not disrupt the structure of FUN-EGF3; they were all surface-exposed in the structure. FUN-EGF3 fragments with single alanine substitutions were made and NMR spectroscopy showed that they had a similar fold to the unmodified protein. However, none of the substitutions abolished binding to cbEGF41-43-BirA. It is possible that none of the targeted residues are involved in the N-C interaction or that the assays could not detect small perturbations. Moreover, if there are multiple binding sites for cbEGF41-43 on FUN-EGF3, a single substitution may not be sufficient to abolish the interaction.

There are several potential lines of enquiry for future investigations into the N-C interaction. In addition to further single residue substitutions, larger-scale changes could be made in FUN-EGF3 fragment, such as the deletion or replacement of entire loop regions. This includes the Y-N-A-Y sequence in the flexible loop in the FUN domain (Section 7.4). An interaction assay using other fragments could also be developed. In previous studies, larger N- and C-terminal fibrillin fragments produced in mammalian cells interacted with high affinity (Lin *et al.*, 2002; Marson *et al.*, 2005), possibly as a result of multimerisation (Section 7.3). Employing similar C-terminal fragments instead of cbEGF41-43 in interaction assays might improve the specificity, as lower concentrations of FUN-EGF3 would be required. Moreover, studying cbEGF41-43 in its native

context would alleviate concerns about misfolding. Initial attempts to produce a fragment comprising domains TB7 to cbEGF43 were unsuccessful, as mammalian cells did not secrete the protein into the extracellular medium (data not shown). Therefore, it may be necessary to include the unique C-terminal region in these larger fragments. A scheme for possible future experiments is shown in Figure 7.2.

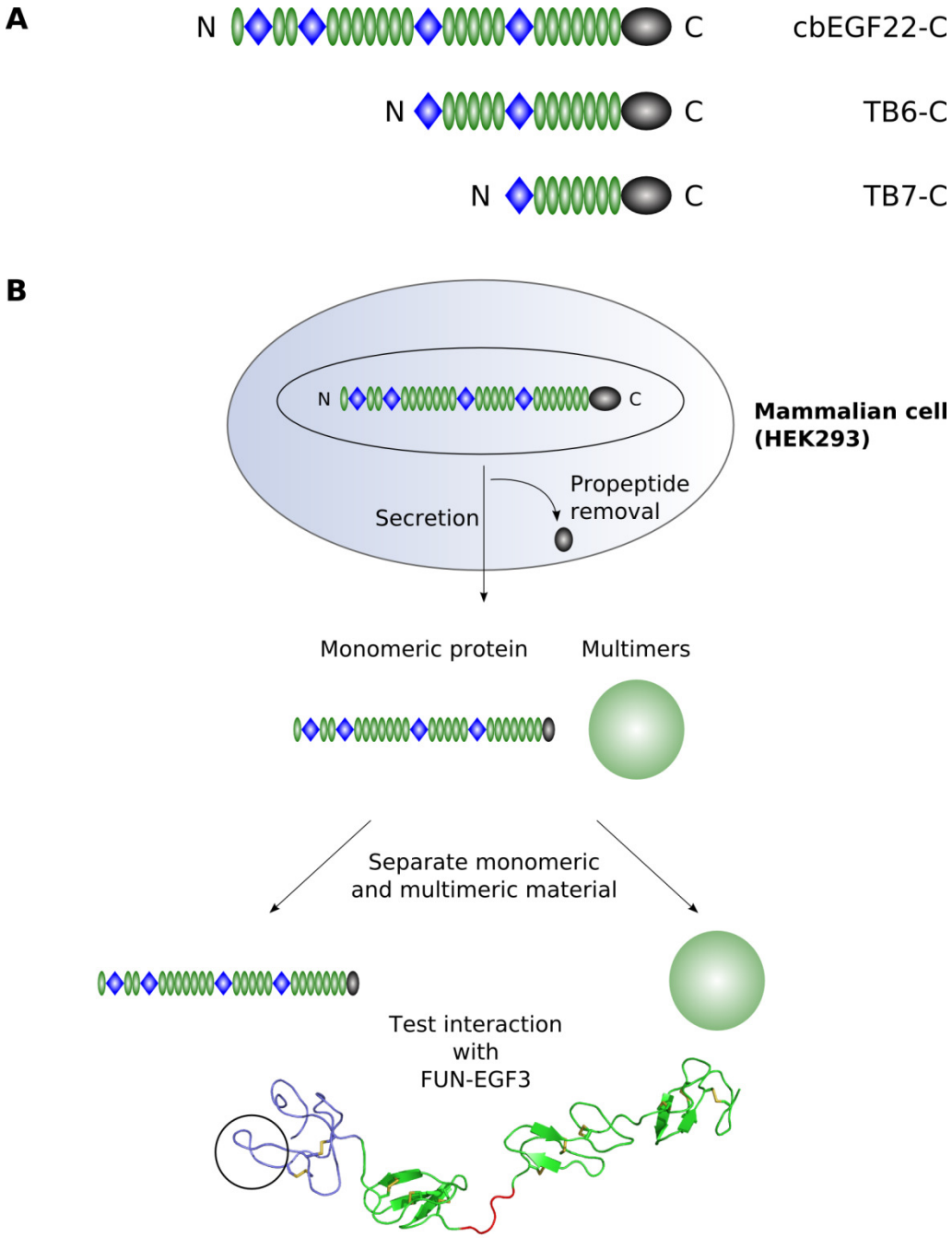


Figure 7.2 Possible future studies of the fibrillin N-C terminal interaction. A) Potential C-terminal fibrillin-1 fragments for use in interaction studies: cbEGF22-C (Lin *et al.*, 2002), TB6-C, TB7-C (Marson *et al.*, 2005). B) Recombinant fragments are produced by human embyonic kidney (HEK) 293 cells and secreted into the extracellular medium. They are then purified from the medium and separated according to size using size-exclusion chromatography. The monomeric and multimeric proteins are tested for interactions with FUN-EGF3 produced by the prokaryotic expression, purification and *in vitro* refolding method. Mutagenesis of both fragments, including deletion of loop regions such as the Y-N-A-Y sequence in FUN-EGF3 (highlighted by black circle), could be used to identify regions important for the interaction.

7.6 Effects of MFS-associated mutations on the fibrillin-1 N- and C-termini

The FUN-EGF3 structure suggested that several mutations found in patients with MFS or MFS-like features could have structural effects (Section 6.2). Missense mutations resulting in the N57D, W71R and P148S substitutions affect residues involved in interdomain packing (Figure 7.3A). Previous studies have demonstrated that disease-linked missense mutations affect the structure of cbEGF domains. Mutations resulting in the loss of calcium-binding ligands (Whiteman *et al.*, 2007), interdomain packing residues (Schneider *et al.*, 2010) or cysteine residues (Suk *et al.*, 2004) can cause misfolding of fragments produced by prokaryotic expression and *in vitro* refolding. In some cases, these mutations also result in defective secretion of recombinant fragments by fibroblast cells (Schneider *et al.*, 2010; Whiteman *et al.*, 2007). Lower levels of fibrillin-1 secretion in MFS patients with these mutations may give rise to a reduced number of microfibrils in the extracellular matrix.

Two different methods were used to test whether the N57D, W71R, P148S substitutions affected protein folding. First, FUN-EGF3 fragments containing the substitutions were characterised by reverse phase HPLC, SDS-PAGE and NMR spectroscopy (Section 6.5). Second, the same substitutions were introduced into the Nterm-Pro fragment and their effects on secretion by MSU1.1 fibroblast cells were assessed by Western blotting (Section 6.7). The results are summarised in Figure 7.3B. All three substitutions caused misfolding of the FUN-EGF3 fragment, but to different extents, while only the N57D and W71R substitutions affected secretion of the Nterm-Pro fragment. Although P148S resulted in misfolding of the EGF3 domain in the FUN-EGF3 fragment, it may only have a local effect in full-length fibrillin-1, as the hyb1 domain might stabilise the EGF3 domain.

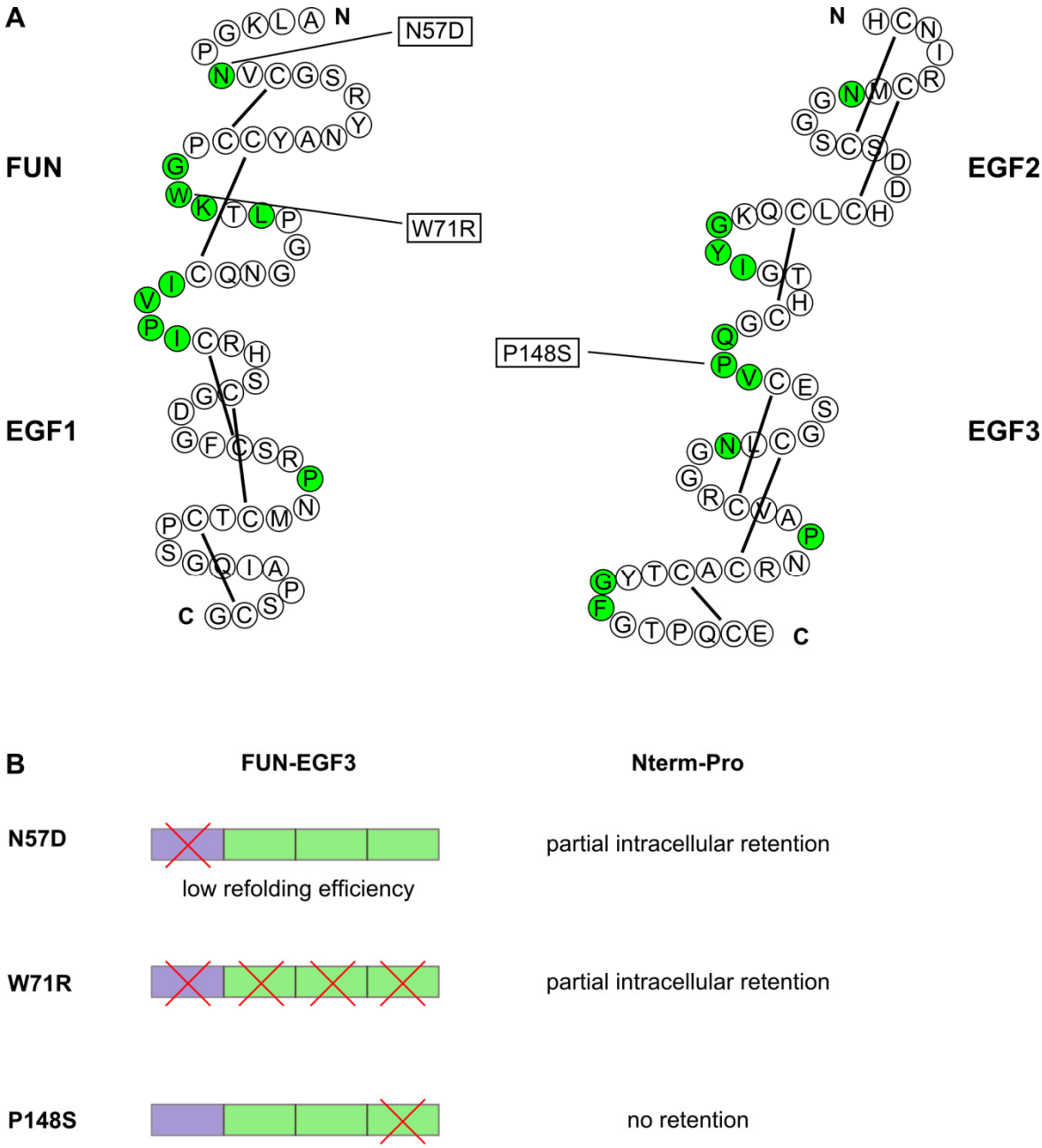


Figure 7.3 Effects of disease-linked residue substitutions in the FUN-EGF3 region. A) Schematic representation of the structures of the FUN-EGF1 and EGF2-EGF3 regions. Residues involved in interdomain packing are highlighted in green, including residues in the EGF3 domain potentially involved in packing with the hyb1 domain (Figure 5.12, Chapter 5). The positions of the N57D, W71R and P148S residue substitutions are indicated. B) Summary of results for the three residue substitutions. Left column: effect on folding of the FUN-EGF3 fragment. Red crosses indicate that a domain is misfolded (domain colour scheme as in Figure 7.1). Right column: summary of secretion assay results using MSU1.1 cells transfected with Nterm-Pro fragments.

The N57D and W71R substitutions did not cause complete intracellular retention of the Nterm-Pro fragment. Other mutations tested in the same assay system also had a partial retention effect: the C1017R MFS-linked substitution (Ji Young Suk, D.Phil. thesis 2003), a WMS-linked deletion in the TB5 domain (Sarah Iqbal, D.Phil. thesis 2011) and the C1719W GD-linked substitution (Alicjia Bulsiewicz, Sacha Jensen and Penny Handford, unpublished results). It is possible that partial retention is caused by delayed secretion. Substitutions resulting in the gain or loss of a cysteine residue have previously been shown to produce such an effect (Aoyama *et al.*, 1994; Halliday *et al.*, 1999). Pulse-chase analysis using a recombinant fibrillin protein distinguishable from endogenous fibrillin-1 would be needed to determine whether the effect was partial retention or delayed secretion.

The effect on secretion suggests that functional haploinsufficiency is a possible mechanism for disease pathogenesis for the mutations resulting in the N57D and W71R substitutions. Unlike the N57D substitution (Chung *et al.*, 2009), the patient with the W71R substitution did not fulfil all the clinical criteria for MFS (Sakai *et al.*, 2006). It is possible that this patient's symptoms were milder because unaffected fibrillin-1 protein compensated for the reduced secretion of the mutant molecules. A definitive assessment cannot be made without studying fibrillin secretion by patient cells.

Effects beyond defective secretion cannot be excluded. None of the N-terminal substitutions abrogated binding of FUN-EGF3 to cbEGF41-43-BirA in the pull-down assay (Section 6.6); the C-terminal Y2639N substitution also had no effect (Section 6.9). However, concerns about the sensitivity of the assay to single residue changes (Section 7.3 and 7.5) mean that effects on the N-C interaction and microfibril assembly cannot be ruled out. In particular, a possible pathogenic mechanism was not identified for the P148S substitution, as it did not perturb secretion. A potential explanation is that the associated mutation is close to an intron-exon boundary

(Stheneur *et al.*, 2009). Although it does not affect the consensus splice site, it is unknown whether it affects splicing efficiency. Exon skipping could result in the loss of domains and consequent protein misfolding. This might cause reduced fibrillin-1 expression and microfibril deposition, as has been observed previously (Halliday *et al.*, 1999).

The N-terminal disease-linked substitutions might also affect interactions with microfibril-associated proteins. SSS-associated mutations in the TB4 domain of fibrillin-1 result in defective interactions with integrins (Loeys *et al.*, 2010) and a WMS-associated deletion of the TB1, Pro-rich and EGF4 domains abrogates binding of ADAMTSL proteins (Sengle *et al.*, 2012). In the FUN-EGF3 region, mutations might perturb fibrillin binding to the BMPs, LTBPs or MAGPs. Without detailed information about their binding sites on fibrillin, it is not possible to make further predictions.

7.7 Implications for microfibril assembly

Although it was not possible to narrow down the N-terminal regions involved in the N-C interaction, structure determination of the FUN-EGF3 fragment laid the foundations for future work. Experiments described earlier (Section 7.5), informed by the FUN-EGF3 structure, could help to dissect the binding site for the cbEGF41-43 region. It is important to study the interaction in context of microfibril assembly (Figure 7.4). If aggregation of C-terminal domains is a prerequisite for a high-affinity interaction with the N-terminal domains (Hubmacher *et al.*, 2008), its molecular basis also needs to be understood. At present, it is unclear how these domains could form disulphide-bonded aggregates. A rearrangement of disulphide bonds would almost certainly be required. Aggregation would also need to occur in a controlled manner, so that the cbEGF41-43 region was exposed and therefore able to interact with the FUN-EGF3 region of other fibrillin monomers. The propeptide regions may regulate assembly. In particular, furin-mediated cleavage of the C-terminal propeptide appears to be essential for fibrillin deposition into microfibrils

(Milewicz *et al.*, 1995; Raghunath *et al.*, 1999). Removal of the propeptides might expose sites involved in aggregation or the N-C interaction, but this is still speculative.

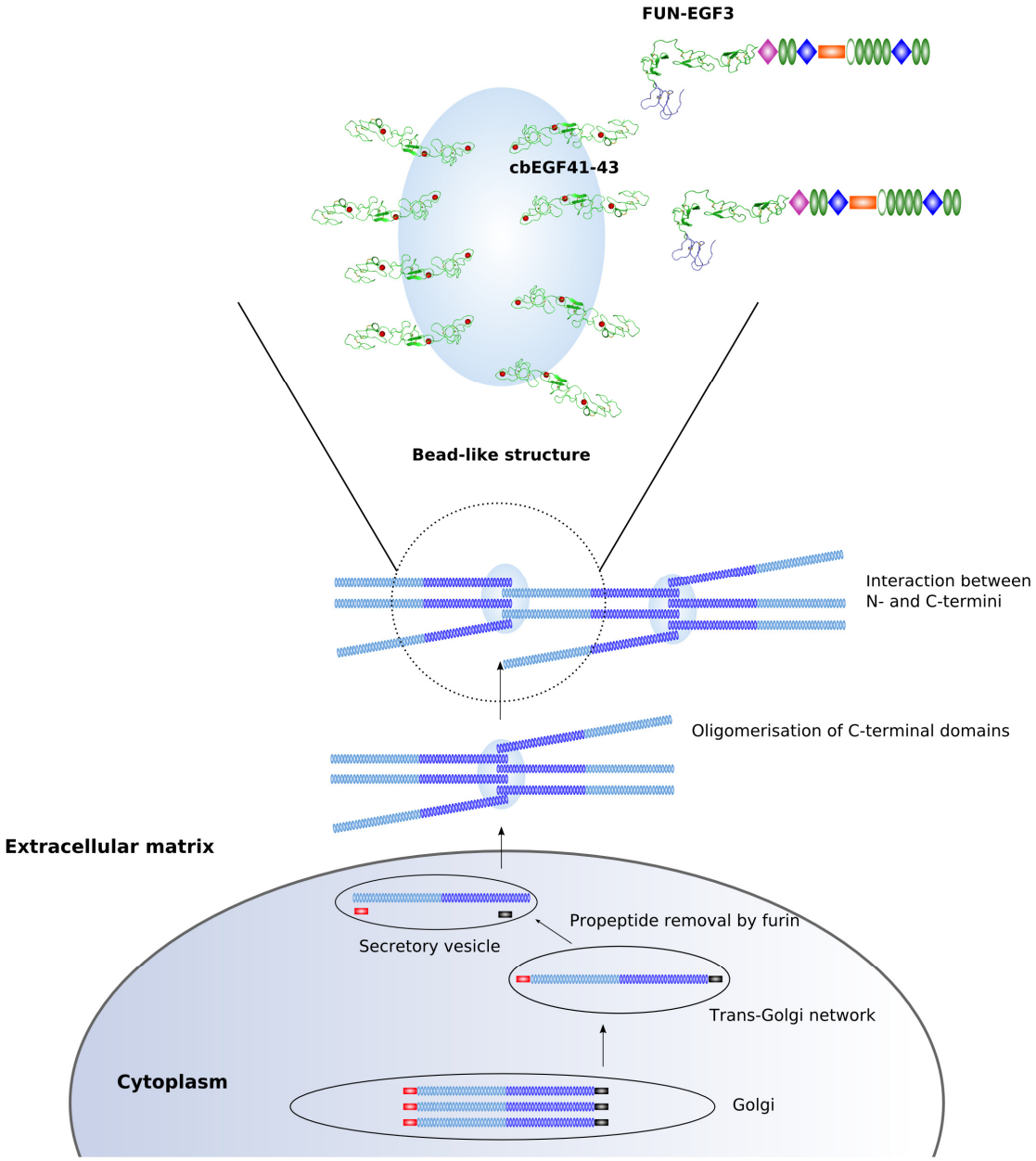


Figure 7.4 The fibrillin N-C interaction in context of microfibil assembly. The initial steps of fibrillin processing and assembly are the cleavage of the N- and C-terminal propeptides and aggregation of the C-terminal domains. The N-terminal domains bind to the C-terminal aggregates which resemble bead-like structures (Hubmacher *et al.*, 2008). cbEGF41-43 (homology model structure shown) forms the binding site for the FUN-EGF3 region. The N-C interaction could link the bead-like structures together to form microfibril-like assemblies. This figure is modified from Figure 1.6, Section 1.7.

Until recently, it was not possible to assess the binding of recombinant, full-length fibrillin molecules to microfibrils. An assay using GFP-fibrillin-1 will help to relate the results of *in vitro* interaction studies to microfibril assembly (Figure 7.5; Sacha Jensen, personal communication). The construct was designed on the basis of the NMR data obtained in this thesis, showing that the N-terminal portion of the FUN domain after the furin cleavage site is unstructured (Section 4.4). A co-culture system is used as HEK293 cells do not assemble microfibrils and it is challenging to obtain MSU1.1 cells that express full-length GFP-fibrillin. Initial results showed that deletion of the FUN-EGF3 region resulted in a partial reduction in the binding of GFP-fibrillin to microfibrils (Sacha Jensen, personal communication). Regions or residues in the N- and C-termini found to be important in N-C interaction could also be modified in GFP-fibrillin-1, and their effects on microfibril binding tested.

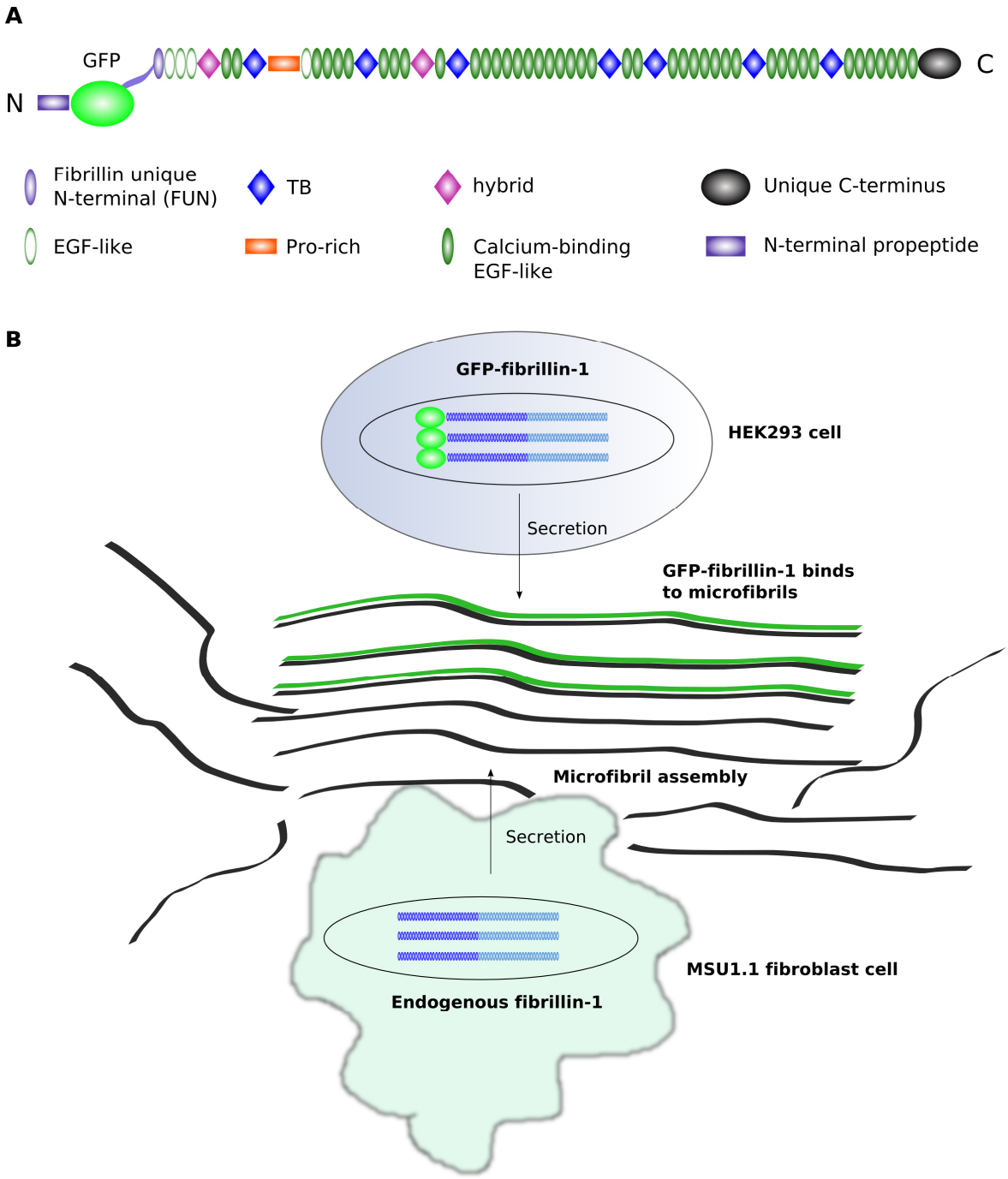


Figure 7.5 Microfibril binding assay using GFP-fibrillin-1. A) Fibrillin-1 with an N-terminal GFP tag. GFP is inserted after the N-terminal furin cleavage site, before the structured region of the FUN domain. B) Microfibril binding assay. HEK293 cells are transfected with a vector encoding GFP-fibrillin and are co-cultured with MSU1.1 cells. The HEK293 cells produce and secrete GFP-fibrillin into the medium, where it associates with microfibrils assembled by the MSU1.1 cells. Microfibril binding is assessed by immunofluorescence, using an anti-Pro-rich antibody to detect fibrillin (endogenous and recombinant) and an anti-GFP antibody to detect GFP-fibrillin. Figure modified from Sacha Jensen, personal communication.

7.8 Implications for microfibril organisation

Models for microfibril organisation differ in their predictions about the organisation of fibrillin monomers. These models seek to explain how fibrillin, with an extended length of ~150 nm, assembles into beaded microfibril structures with a distance of 50-60 nm between the bead-like structures (Section 1.6). Staggered models predict a linear arrangement, while folded models require bends in fibrillin monomers to accommodate the observed periodicity. Evidence for a folded arrangement comes from antibody epitope mapping data (Baldock *et al.*, 2001), SAXS-derived structural models of recombinant fibrillin fragments (Baldock *et al.*, 2006) and studies of microfibril extension (Wang *et al.*, 2009; Wess *et al.*, 1998). However, high-resolution structures of fibrillin-1 fragments (Downing *et al.*, 1996; Jensen *et al.*, 2009; Lee *et al.*, 2004; Smallridge *et al.*, 2003) indicate that the majority of the fibrillin molecule is likely to adopt a linear structure. Furthermore, a recent antibody epitope mapping study showed that a staggered arrangement is more likely and that the earlier data could have resulted from microfibril degradation (Kuo *et al.*, 2007).

Work in this thesis adds FUN-EGF3 to the set of high-resolution fibrillin-1 fragment structures (Figure 7.6A). Although the linker region between the EGF1 and EGF2 domains is flexible (Section 4.4), this is unlikely to have a bearing on the overall organisation of fibrillin as the FUN-EGF1 pair is at the N-terminus (Figure 7.6B). Nonetheless, the EGF2-EGF3 region adopts a rod-like structure (Figure 7.1), again consistent with a linear arrangement of fibrillin monomers. Elsewhere, the flexible TB6-cbEGF32 linker may allow a kink in the structure, but otherwise there are few candidates for flexible or bent regions.

The most recent version of the folded model includes a bend between the TB4 and TB6 domains (Baldock *et al.*, 2006). However, this seems unlikely, as high-resolution structures show that

cbEGF-cbEGF domain pairs and cbEGF-TB-cbEGF triples form linear, rod-like structures. It has been proposed that the Pro-rich region could act as a hinge (Pereira *et al.*, 1993), but there is currently no information about its structure and dynamics. Even if the Pro-rich region is flexible, a linear array of domains spanning EGF4 to TB6 has an expected length of ~95 nm – longer than the inter-bead distance (Figure 7.6B). A complete understanding of the organisation of the N-terminal domains awaits structural information about the EGF3-hyb1 interface and the TB1-EGF4 region (Figure 7.6A).

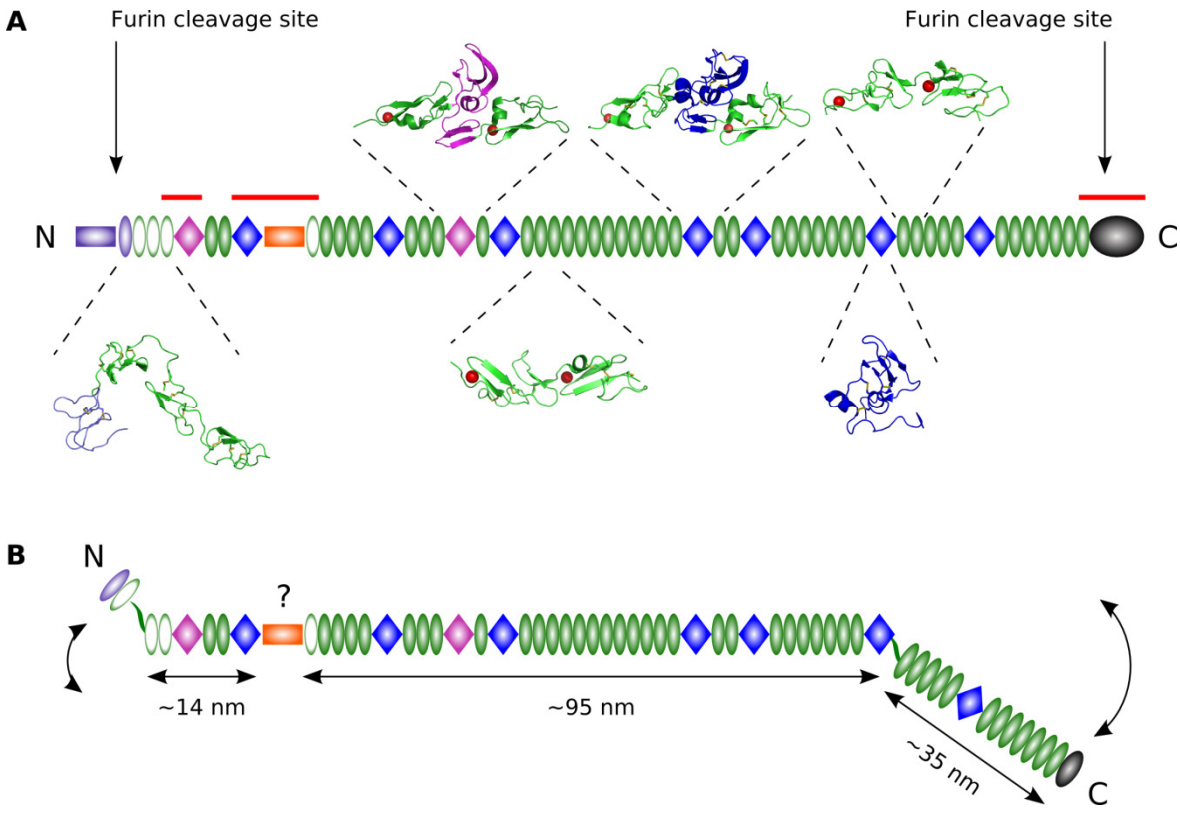


Figure 7.6 Structure and organisation of human fibrillin-1. A) Structures of fibrillin-1 fragments, including FUN-EGF3. Other structures are as in Figure 1.2, Section 1.3 and the legend for domain structures is as in Figure 7.5. Red bars indicate regions of unknown structure. B) Organisation of fibrillin-1. Flexible regions have been identified between the EGF1 and EGF2 domains, as well as between the TB6 and cbEGF32 domains (flexibility indicated by arrows). It is unknown whether the Pro-rich region is flexible or rigid. Approximate estimated dimensions are given for the three segments separated by the Pro-rich region and the TB6 domain, using lengths of 2.7 nm, 2.3 nm, 2.0 nm and 2.1 nm for the cbEGF, TB, hyb and EGF-like domains respectively.

7.9 Implications for interactions of fibrillin-1 with microfibril-associated factors

Many microfibril-associated proteins have been shown to bind the fibrillin N-terminal domains, although few interaction sites have been characterised (Section 1.8). The FUN-EGF3 structure means that their binding sites can be studied in greater detail. In particular, the prodomains of BMP-5 and -7 interact with the FUN-EGF3 region (Sengle *et al.*, 2008a; Sengle *et al.*, 2011). Mapping the binding sites of accessory factors will be important for understanding the regulation of growth factor sequestration and elastic fibre assembly, which are defective in elastic fibre disorders. It is likely that some of these proteins have overlapping interaction sites and so may compete for binding, as has been previously observed for the fibulins and LTBPs (Ono *et al.*, 2009). Characterisation of interactions of these proteins with fibrillin will also allow a model of elastic fibre structure to be developed.

7.10 Conclusions

In conclusion, work in this thesis has provided novel information about the structure and dynamics of the four N-terminal domains of human fibrillin-1, the minimal region required for binding to the C-terminal cbEGF41-43 region. The FUN-EGF3 fragment comprises two rod-like domain pairs, FUN-EGF1 and EGF2-EGF3, connected by a flexible linker sequence. Evidence for native-like folding of this fragment came from its interaction with the cbEGF41-43-BirA fragment and the similarity of the FUN-EGF1 and EGF2-EGF3 interdomain interfaces to EGF-EGF domain pairs in other extracellular proteins. The structure also explained the effects of some disease-associated missense mutations, three of which result in residue substitutions at the FUN-EGF1 and EGF2-EGF3 interdomain interfaces. Two of these substitutions, N57D and W71R, caused defective secretion of an N-terminal fibrillin fragment from MSU1.1 fibroblast cells, suggesting a possible mechanism of disease pathogenesis.

The FUN-EGF3 structure provides the starting point for dissecting the N-C interaction, a likely initial step in fibrillin microfibril assembly. Although initial attempts to narrow down the binding site for cbEGF41-43 on FUN-EGF3 were unsuccessful, alternative approaches using larger C-terminal fragments have been proposed to study the interaction. Moreover, the FUN-EGF3 structure and complete resonance assignments allow the structural effects of modifications to be assessed. In future, combining the results of *in vitro* interaction studies with cell-based microfibril binding assays using recombinant fibrillin proteins will provide insights into the role of the N-C interaction in microfibril assembly.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool', *J Mol Biol,* 215(3), 403-10.

Aoyama, T., Francke, U., Dietz, H. C. and Furthmayr, H. (1994) 'Quantitative differences in biosynthesis and extracellular deposition of fibrillin in cultured fibroblasts distinguish five groups of Marfan syndrome patients and suggest distinct pathogenetic mechanisms.', *J Clin Invest,* 94(1), 130-7.

Aoyama, T., Francke, U., Gasner, C. and Furthmayr, H. (1995) 'Fibrillin abnormalities and prognosis in Marfan syndrome and related disorders.', *Am J Med Genet,* 58(2), 169-76.

Arbustini, E., Grasso, M., Ansaldi, S., Malattia, C., Pilotto, A., Porcu, E., Disabella, E., Marziliano, N., Pisani, A., Lanzarini, L., Mannarino, S., Larizza, D., Mosconi, M., Antoniazzi, E., Zoia, M. C., Meloni, G., Magrassi, L., Brega, A., Bedeschi, M. F., Torrente, I., Mari, F. and Tavazzi, L. (2005) 'Identification of sixty-two novel and twelve known FBN1 mutations in eighty-one unrelated probands with Marfan syndrome and other fibrillinopathies', *Hum Mutat,* 26(5), 494.

Arnold, K., Kiefer, F., Kopp, J., Battey, J. N., Podvinec, M., Westbrook, J. D., Berman, H. M., Bordoli, L. and Schwede, T. (2009) 'The Protein Model Portal', *J Struct Funct Genomics,* 10(1), 1-8.

Ashworth, J. L., Kelly, V., Rock, M. J., Shuttleworth, C. A. and Kielty, C. M. (1999) 'Regulation of fibrillin carboxy-terminal furin processing by N-glycosylation, and association of amino- and carboxy-terminal sequences', *J Cell Sci,* 112 (Pt 22), 4163.

Baldock, C., Koster, A. J., Ziese, U., Rock, M. J., Sherratt, M. J., Kadler, K. E., Shuttleworth, C. A. and Kielty, C. M. (2001) 'The supramolecular organization of fibrillin-rich microfibrils', *J Cell Biol,* 152(5), 1045.

Baldock, C., Siegler, V., Bax, D. V., Cain, S. A., Mellody, K. T., Marson, A., Haston, J. L., Berry, R., Wang, M. C., Grossmann, J. G., Roessle, M., Kielty, C. M. and Wess, T. J. (2006) 'Nanostructure of fibrillin-1 reveals compact conformation of EGF arrays and mechanism for extensibility', *Proc Natl Acad Sci U S A,* 103(32), 11922.

Bashir, M. M., Abrams, W. R., Rosenbloom, J., Kucich, U., Bacarra, M., Han, M. D., Brown-Augsberger, P. and Mecham, R. (1994) 'Microfibril-associated glycoprotein: characterization of the bovine gene and of the recombinantly expressed protein.', *Biochemistry,* 33(2), 593-600.

Baumgartner, C., Mátyás, G., Steinmann, B. and Baumgartner, D. (2005) 'Marfan syndrome--a diagnostic challenge caused by phenotypic and genetic heterogeneity', *Methods Inf Med,* 44(4), 487-97.

Bax, A., Clore, G. and Gronenborn, A. (1990) 'H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13-enriched proteins', *J Magn Reson,* 88(2), 425-431.

Bax, D. V., Bernard, S. E., Lomas, A., Morgan, A., Humphries, J., Shuttleworth, C. A., Humphries, M. J. and Kielty, C. M. (2003) 'Cell adhesion to fibrillin-1 molecules and microfibrils is mediated by alpha 5 beta 1 and alpha v beta 3 integrins.', *J Biol Chem,* 278(36), 34605-16.

Biggin, A., Holman, K., Brett, M., Bennetts, B. and Adès, L. (2004) 'Detection of thirty novel FBN1 mutations in patients with Marfan syndrome or a related fibrillinopathy', *Hum Mutat,* 23(1), 99.

Biot, C., Buisine, E. and Rooman, M. (2003) 'Free-energy calculations of protein-ligand cation-pi and amino-pi interactions: From vacuum to proteinlike environments', *J Am Chem Soc,* 125(46), 13988-13994.

Birk, D., Bruckner, P. and Mecham, R. (2011) 'Collagens, Suprastructures, and Collagen Fibril Assembly', *Extracellular Matrix: an Overview*, 77-115.

Birnboim, H. C. and Doly, J. (1979) 'A rapid alkaline extraction procedure for screening recombinant plasmid DNA', *Nucleic Acids Res,* 7(6), 1513-23.

Bishop, B., Aricescu, A. R., Harlos, K., O'Callaghan, C. A., Jones, E. Y. and Siebold, C. (2009) 'Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP.', *Nat Struct Mol Biol,* 16(7), 698-703.

Bjellqvist, B., Basse, B., Olsen, E. and Celis, J. E. (1994) 'Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions', *Electrophoresis,* 15(3-4), 529-39.

Black, C., Withers, A. P., Gray, J. R., Bridges, A. B., Craig, A., Baty, D. U. and Boxer, M. (1998) 'Correlation of a recurrent FBN1 mutation (R122C) with an atypical familial Marfan syndrome phenotype', *Hum Mutat,* Suppl 1, S198-200.

Bodenhausen, G. and Ruben, D. (1980) 'Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy', *Chemical Physics Letters,* 69(1), 185-189.

Booms, P., Pregla, R., Ney, A., Barthel, F., Reinhardt, D. P., Pletschacher, A., Mundlos, S. and Robinson, P. N. (2005) 'RGD-containing fibrillin-1 fragments upregulate matrix metalloproteinase expression in cell culture: a potential factor in the pathogenesis of the Marfan syndrome', *Hum Genet,* 116(1-2), 51-61.

Bork, P., Downing, A., Kieffer, B. and Campbell, I. (1996) 'Structure and distribution of modules in extracellular proteins.', *Q Rev Biophys,* 29(2), 119-67.

Braunschweiler, L. and Ernst, R. (1983) 'Coherence transfer by isotropic mixing - application to proton correlation spectroscopy', *J Magn Reson,* 53(3), 521-528.

Brünger, A. T., Clore, G. M., Gronenborn, A. M. and Karplus, M. (1986) 'Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: application to crambin', *Proc Natl Acad Sci U S A,* 83(11), 3801-5.

Buck, M., Boyd, J., Redfield, C., Mackenzie, D., Jeenes, D., Archer, D. and Dobson, C. (1995) 'Structural determinations of protein dynamics - analysis of N-15 NMR relaxation measurements for main-chain and side-chain nuclei of hen egg-white lysozyme', *Biochemistry,* 34(12), 4041-4055.

Cain, S., Baldock, C., Gallagher, J., Morgan, A., Bax, D., Weiss, A., Shuttleworth, C. and Kielty, C. (2005) 'Fibrillin-1 interactions with heparin. Implications for microfibril and elastic fiber assembly.', *J Biol Chem,* 280(34), 30526-37.

Cain, S. A., Baldwin, A. K., Mahalingam, Y., Raynal, B., Jowitt, T. A., Shuttleworth, C. A., Couchman, J. R. and Kielty, C. M. (2008) 'Heparan sulfate regulates fibrillin-1 N- and C-terminal interactions.', *J Biol Chem,* 283(40), 27017-27.

Campbell, I. D. and Humphries, M. J. (2011) 'Integrin structure, activation, and interactions', *Cold Spring Harb Perspect Biol,* 3(3).

Cardy, C. M. and Handford, P. A. (1998) 'Metal ion dependency of microfibrils supports a rod-like conformation for fibrillin-1 calcium-binding epidermal growth factor-like domains.', *J Mol Biol,* 276(5), 855-60.

Cañadas, V., Vilacosta, I., Bruna, I. and Fuster, V. (2010a) 'Marfan syndrome. Part 1: pathophysiology and diagnosis.', *Nat Rev Cardiol,* 7(5), 256-65.

Cañadas, V., Vilacosta, I., Bruna, I. and Fuster, V. (2010b) 'Marfan syndrome. Part 2: treatment and management of patients.', *Nat Rev Cardiol,* 7(5), 266-76.

Charbonneau, N., Carlson, E., Tufa, S., Sengle, G., Manalo, E., Carlberg, V., Ramirez, F., Keene, D. and Sakai, L. (2010) 'In vivo studies of mutant fibrillin-1 microfibrils.', *J Biol Chem,* 285(32), 24943-55.

Chung, B., Lam, S., Tong, T., Li, S., Lun, K., Chan, D., Fok, S., Or, J., Smith, D., Yang, W. and Lau, Y. (2009) 'Identification of novel FBN1 and TGFBR2 mutations in 65 probands with Marfan syndrome or Marfan-like phenotypes.', *Am J Med Genet A,* 149A(7), 1452-9.

Clore, G. and Garrett, D. (1999) 'R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures', *J Am Chem Soc,* 121(39), 9008-9012.

Clore, G. M., Gronenborn, A. M. and Bax, A. (1998) 'A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information', *J Magn Reson,* 133(1), 216-21.

Clubb, R., Thanabal, V. and Wagner, G. (1992) 'A constant-time 3-dimensional triple-resonance pulse scheme to correlate intraresidue H-1(N), N-15, and C-13(') chemical-shifts in N-15-C-13-labeled proteins', *J Magn Reson,* 97(1), 213-217.

Collaborative Computational Project, N. m. (1994) 'The CCP4 suite: programs for protein crystallography', *Acta Crystallogr D Biol Crystallogr,* 50(Pt 5), 760-3.

Collod-Béroud, G., Béroud, C., Adès, L., Black, C., Boxer, M., Brock, D. J., Godfrey, M., Hayward, C., Karttunen, L., Milewicz, D., Peltonen, L., Richards, R. I., Wang, M., Junien, C. and Boileau, C. (1997) 'Marfan Database (second edition): software and database for the analysis of mutations in the human FBN1 gene', *Nucleic Acids Res,* 25(1), 147-50.

Comeglio, P., Evans, A., Brice, G., Cooling, R. and Child, A. (2002) 'Identification of FBN1 gene mutations in patients with ectopia lentis and marfanoid habitus.', *Br J Ophthalmol,* 86(12), 1359-62.

Cordier, F. and Grzesiek, S. (1999) 'Direct observation of hydrogen bonds in proteins by interresidue (3h)J(NC ') scalar couplings', *J Am Chem Soc,* 121(7), 1601-1602.

Cordle, J. (2005). 'A molecular analysis of Notch-Ligand binding'. Biochemistry D.Phil. thesis, University of Oxford.

Cordle, J., Johnson, S., Tay, J. Z., Roversi, P., Wilkin, M. B., de Madrid, B. H., Shimizu, H., Jensen, S., Whiteman, P., Jin, B., Redfield, C., Baron, M., Lea, S. M. and Handford, P. A. (2008a) 'A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition', *Nat Struct Mol Biol,* 15(8), 849.

Cordle, J., Redfieldz, C., Stacey, M., van der Merwe, P. A., Willis, A. C., Champion, B. R., Hambleton, S. and Handford, P. A. (2008b) 'Localization of the delta-like-1-binding site in human Notch-1 and its modulation by calcium affinity', *J Biol Chem,* 283(17), 11785-93.

Cornilescu, G., Marquardt, J., Ottiger, M. and Bax, A. (1998) 'Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase', *J Am Chem Soc,* 120(27), 6836-6837.

Corson, G. M., Chalberg, S. C., Dietz, H. C., Charbonneau, N. L. and Sakai, L. Y. (1993) 'Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5' end.', *Genomics,* 17(2), 476-84.

Corson, G. M., Charbonneau, N. L., Keene, D. R. and Sakai, L. Y. (2004) 'Differential expression of fibrillin-3 adds to microfibril variety in human and avian, but not rodent, connective tissues', *Genomics,* 83(3), 461.

De Bernardez Clark, E., Schwarz, E. and Rudolph, R. (1999) 'Inhibition of aggregation side reactions during in vitro protein folding', *Methods Enzymol,* 309, 217-36.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) 'NMRPipe: a multidimensional spectral processing system based on UNIX pipes.', *J Biomol NMR,* 6(3), 277-93.

Dietz, H. C., Cutting, G. R., Pyeritz, R. E., Maslen, C. L., Sakai, L. Y., Corson, G. M., Puffenberger, E. G., Hamosh, A., Nanthakumar, E. J. and Curristin, S. M. (1991) 'Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene.', *Nature,* 352(6333), 337-9.

Dietz, H. C. and Pyeritz, R. E. (1995) 'Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders', *Hum Mol Genet,* 4 Spec No, 1799-809.

Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D. and Handford, P. A. (1996) 'Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders', *Cell,* 85(4), 597.

El-Hallous, E., Sasaki, T., Hubmacher, D., Getie, M., Tiedemann, K., Brinckmann, J., Bätge, B., Davis, E. C. and Reinhardt, D. P. (2007) 'Fibrillin-1 interactions with fibulins depend on the first hybrid domain and provide an adaptor function to tropoelastin.', *J Biol Chem,* 282(12), 8935-46.

Eldadah, Z. A., Brenn, T., Furthmayr, H. and Dietz, H. C. (1995) 'Expression of a mutant human fibrillin allele upon a normal human or murine genetic background recapitulates a Marfan cellular phenotype.', *J Clin Invest,* 95(2), 874-80.

Engel, J., Chiquet, M. and Mecham, R. (2011) 'An Overview of Extracellular Matrix Structure and Function', *Extracellular Matrix: an Overview*, 1-39.

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. and Sali, A. (2006) 'Comparative protein structure modeling using Modeller', *Curr Protoc Bioinformatics,* Chapter 5, Unit 5.6.

Evans, P. (2012). 'Identifying key sites at the fibrillin N-terminus involved in the N- to C-terminal interaction.' Biochemistry Final Honour School Part II Project, University of Oxford

Faivre, L., Collod-Beroud, G., Adès, L., Arbustini, E., Child, A., Callewaert, B. L., Loeys, B., Binquet, C., Gautier, E., Mayer, K., Arslan-Kirchner, M., Grasso, M., Beroud, C., Hamroun, D., Bonithon-Kopp, C., Plauchu, H., Robinson, P. N., De Backer, J., Coucke, P., Francke, U., Bouchot, O., Wolf, J. E., Stheneur, C., Hanna, N., Detaint, D., De Paepe, A., Boileau, C. and Jondeau, G. (2012) 'The new Ghent criteria for Marfan syndrome: what do they change?', *Clin Genet,* 81(5), 433-42.

Faivre, L., Collod-Beroud, G., Callewaert, B., Child, A., Binquet, C., Gautier, E., Loeys, B. L., Arbustini, E., Mayer, K., Arslan-Kirchner, M., Stheneur, C., Kiotsekoglou, A., Comeglio, P., Marziliano, N., Wolf, J. E., Bouchot, O., Khau-Van-Kien, P., Beroud, C., Claustres, M., Bonithon-Kopp, C., Robinson, P. N., Adès, L., De Backer, J., Coucke, P., Francke, U., De Paepe, A., Jondeau, G. and Boileau, C. (2009) 'Clinical and mutation-type analysis from an international series of 198 probands with a pathogenic FBN1 exons 24-32 mutation.', *Eur J Hum Genet,* 17(4), 491-501.

Faivre, L., Collod-Beroud, G., Loeys, B. L., Child, A., Binquet, C., Gautier, E., Callewaert, B., Arbustini, E., Mayer, K., Arslan-Kirchner, M., Kiotsekoglou, A., Comeglio, P., Marziliano, N., Dietz, H. C., Halliday, D., Beroud, C., Bonithon-Kopp, C., Claustres, M., Muti, C., Plauchu, H., Robinson, P. N., Adès, L. C., Biggin, A., Benetts, B., Brett, M., Holman, K. J., De Backer, J., Coucke, P., Francke, U., De Paepe, A., Jondeau, G. and Boileau, C. (2007) 'Effect of mutation type and location on clinical outcome in 1,013 probands with Marfan syndrome or related phenotypes and FBN1 mutations: an international study.', *Am J Hum Genet,* 81(3), 454-66.

Faivre, L., Gorlin, R. J., Wirtz, M. K., Godfrey, M., Dagoneau, N., Samples, J. R., Le Merrer, M., Collod-Beroud, G., Boileau, C., Munnich, A. and Cormier-Daire, V. (2003) 'In frame fibrillin-1 gene deletion in autosomal dominant Weill-Marchesani syndrome.', *J Med Genet,* 40(1), 34-6.

Farmer, B. T., Constantine, K. L., Goldfarb, V., Friedrichs, M. S., Wittekind, M., Yanchunas, J., Robertson, J. G. and Mueller, L. (1996) 'Localizing the NADP+ binding site on the MurB enzyme by NMR', *Nat Struct Biol,* 3(12), 995-7.

Folmer, R. H., Hilbers, C. W., Konings, R. N. and Nilges, M. (1997) 'Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data', *J Biomol NMR,* 9(3), 245-58.

Francke, U., Berg, M. A., Tynan, K., Brenn, T., Liu, W., Aoyama, T., Gasner, C., Miller, D. C. and Furthmayr, H. (1995) 'A Gly1127Ser mutation in an EGF-like domain of the fibrillin-1 gene is a risk factor for ascending aortic aneurysm and dissection', *Am J Hum Genet,* 56(6), 1287-96.

Frenkiel, T., Bauer, C., Carr, M., Birdsall, B. and Feeney, J. (1990) 'HMQC-NOESY-HMQC, a 3-dimensional NMR experiment which allows detection of nuclear Overhauser effects between protons with overlapping signals', *J Magn Reson,* 90(2), 420-425.

Gibson, M. A., Finnis, M. L., Kumaratilake, J. S. and Cleary, E. G. (1998) 'Microfibril-associated glycoprotein-2 (MAGP-2) is specifically associated with fibrillin-containing microfibrils but exhibits

more restricted patterns of tissue localization and developmental expression than its structural relative MAGP-1.', *J Histochem Cytochem,* 46(8), 871-86.

Gibson, M. A., Hatzinikolas, G., Davis, E. C., Baker, E., Sutherland, G. R. and Mecham, R. P. (1995) 'Bovine latent transforming growth factor beta 1-binding protein 2: molecular cloning, identification of tissue isoforms, and immunolocalization to elastin-associated microfibrils.', *Mol Cell Biol,* 15(12), 6932-42.

Gibson, M. A., Hughes, J. L., Fanning, J. C. and Cleary, E. G. (1986) 'The major antigen of elastin-associated microfibrils is a 31-kDa glycoprotein.', *J Biol Chem,* 261(24), 11429-36.

Gill, S. C. and von Hippel, P. H. (1989) 'Calculation of protein extinction coefficients from amino acid sequence data', *Anal Biochem,* 182(2), 319-26.

Giltay, R., Timpl, R. and Kostka, G. (1999) 'Sequence, recombinant expression and tissue localization of two novel extracellular matrix proteins, fibulin-3 and fibulin-4.', *Matrix Biol,* 18(5), 469-80.

Glab, J. and Wess, T. (2008) 'Changes in the molecular packing of fibrillin microfibrils during extension indicate intrafibrillar and interfibrillar reorganization in elastic response.', *J Mol Biol,* 383(5), 1171-80.

Gregory, K. E., Ono, R. N., Charbonneau, N. L., Kuo, C. L., Keene, D. R., Bachinger, H. P. and Sakai, L. Y. (2005) 'The prodomain of BMP-7 targets the BMP-7 complex to the extracellular matrix', *J Biol Chem,* 280(30), 27970.

Grishaev, A. and Bax, A. (2004) 'An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation', *J Am Chem Soc,* 126(23), 7281-92.

Grzesiek, S., Anglister, J. and Bax, A. (1993) 'Correlation of backbone amide and aliphatic side-chain resonances in C-13/N-15-enriched proteins by isotropic mixing of C-13 magnetization', *J Magn Reson Ser B,* 101(1), 114-119.

Habashi, J. P., Judge, D. P., Holm, T. M., Cohn, R. D., Loeys, B. L., Cooper, T. K., Myers, L., Klein, E. C., Liu, G., Calvi, C., Podowski, M., Neptune, E. R., Halushka, M. K., Bedja, D., Gabrielson, K., Rifkin, D. B., Carta, L., Ramirez, F., Huso, D. L. and Dietz, H. C. (2006) 'Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome.', *Science,* 312(5770), 117-21.

Habeck, M., Rieping, W. and Nilges, M. (2005) 'Bayesian estimation of Karplus parameters and torsion angles from three-bond scalar couplings constants', *J Magn Reson,* 177(1), 160-165.

Halliday, D., Hutchinson, S., Kettle, S., Firth, H., Wordsworth, P. and Handford, P. A. (1999) 'Molecular analysis of eight mutations in FBN1.', *Hum Genet,* 105(6), 587-97.

Halliday, D. (2000).' Molecular analysis of *FBN1* mutations in Marfan syndrome and related disorders'. Biochemistry D.Phil. thesis, University of Oxford.

Hambleton, S., Valeyev, N. V., Muranyi, A., Knott, V., Werner, J. M., McMichael, A. J., Handford, P. A. and Downing, A. K. (2004) 'Structural and functional properties of the human notch-1 ligand binding region', *Structure,* 12(12), 2173-83.

Handford, P., Mayhew, M., Baron, M., Winship, P., Campbell, I. and Brownlee, G. (1991) 'Key residues involved in calcium-binding motifs in EGF-like domains.', *Nature,* 351(6322), 164-7.

Handford, P. A., Baron, M., Mayhew, M., Willis, A., Beesley, T., Brownlee, G. G. and Campbell, I. D. (1990) 'The first EGF-like domain from human factor IX contains a high-affinity calcium binding site', *EMBO J,* 9(2), 475-80.

Handford, P. A., Downing, A. K., Reinhardt, D. P. and Sakai, L. Y. (2000) 'Fibrillin: from domain structure to supramolecular assembly', *Matrix Biol,* 19(6), 457.

Hanssen, E., Franc, S. and Garrone, R. (1998) 'Atomic force microscopy and modeling of natural elastic fibrillin polymers', *Biol Cell,* 90(3), 223.

Hanssen, E., Hew, F. H., Moore, E. and Gibson, M. A. (2004) 'MAGP-2 has multiple binding regions on fibrillins and has covalent periodic association with fibrillin-containing microfibrils', *J Biol Chem,* 279(28), 29185-94.

Hayward, C., Porteous, M. E. and Brock, D. J. (1997) 'Mutation screening of all 65 exons of the fibrillin-1 gene in 60 patients with Marfan syndrome: report of 12 novel mutations', *Hum Mutat,* 10(4), 280-9.

Henderson, M., Polewski, R., Fanning, J. C. and Gibson, M. A. (1996) 'Microfibril-associated glycoprotein-1 (MAGP-1) is specifically located on the beads of the beaded-filament structure for fibrillin-containing microfibrils as visualized by the rotary shadowing technique.', *J Histochem Cytochem,* 44(12), 1389-97.

Higman, V. A., Boyd, J., Smith, L. J. and Redfield, C. (2004) 'Asparagine and glutamine side-chain conformation in solution and crystal: a comparison for hen egg-white lysozyme using residual dipolar couplings', *J Biomol NMR,* 30(3), 327-46.

Hirai, M., Horiguchi, M., Ohbayashi, T., Kita, T., Chien, K. R. and Nakamura, T. (2007) 'Latent TGF-beta-binding protein 2 binds to DANCE/fibulin-5 and regulates elastic fiber assembly.', *EMBO J,* 26(14), 3283-95.

Hirani, R., Hanssen, E. and Gibson, M. A. (2007) 'LTBP-2 specifically interacts with the amino-terminal region of fibrillin-1 and competes with LTBP-1 for binding to this microfibrillar protein.', *Matrix Biol,* 26(4), 213-23.

Hollister, D. W., Godfrey, M., Sakai, L. Y. and Pyeritz, R. E. (1990) 'Immunohistologic abnormalities of the microfibrillar-fiber system in the Marfan syndrome.', *N Engl J Med,* 323(3), 152-9.

Holm, L. and Rosenström, P. (2010) 'Dali server: conservation mapping in 3D', *Nucleic Acids Res,* 38(Web Server issue), W545-9.

Howarth, R., Yearwood, C. and Harvey, J. F. (2007) 'Application of dHPLC for mutation detection of the fibrillin-1 gene for the diagnosis of Marfan syndrome in a National Health Service Laboratory', *Genet Test,* 11(2), 146-52.

Hubmacher, D., El-Hallous, E. I., Nelea, V., Kaartinen, M. T., Lee, E. R. and Reinhardt, D. P. (2008) 'Biogenesis of extracellular microfibrils: Multimerization of the fibrillin-1 C terminus into bead-like structures enables self-assembly', *Proc Natl Acad Sci U S A,* 105(18), 6548.

Hubmacher, D., Reinhardt, D. and Mecham, R. (2011) 'Microfibrils and Fibrillin', *Extracellular Matrix: an Overview*, 233-265.

Hutchinson, S., Furger, A., Halliday, D., Judge, D. P., Jefferson, A., Dietz, H. C., Firth, H. and Handford, P. A. (2003) 'Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype?', *Hum Mol Genet,* 12(18), 2269-76.

Hynes, R. O. (2009) 'The extracellular matrix: not just pretty fibrils.', *Science,* 326(5957), 1216-9.

Ikura, M., Bax, A., Clore, G. and Gronenborn, A. (1990) 'Detection of nuclear Overhauser effects between degenerate amide proton resonances by heteronuclear 3-dimensional nuclear-magnetic-resonance spectroscopy', *J Am Chem Soc,* 112(24), 9020-9022.

Iqbal, S. (2011).' Molecular studies of stiff skin syndrome'. Biochemistry D.Phil. thesis, University of Oxford.

Isogai, Z., Ono, R. N., Ushiro, S., Keene, D. R., Chen, Y., Mazzieri, R., Charbonneau, N. L., Reinhardt, D. P., Rifkin, D. B. and Sakai, L. Y. (2003) 'Latent transforming growth factor beta-binding protein 1 interacts with fibrillin and is a microfibril-associated protein', *J Biol Chem,* 278(4), 2750.

Jeener, J., Meier, B., Bachmann, P. and Ernst, R. (1979) 'Investigation of exchange processes by 2-dimensional NMR-spectroscopy', *Journal of Chemical Physics,* 71(11), 4546-4553.

Jensen, S. A., Corbett, A. R., Knott, V., Redfield, C. and Handford, P. A. (2005) 'Ca2+-dependent interface formation in fibrillin-1', *J Biol Chem,* 280(14), 14076.

Jensen, S. A., Iqbal, S., Lowe, E. D., Redfield, C. and Handford, P. A. (2009) 'Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins', *Structure,* 17(5), 759.

Jensen, S. A., Reinhardt, D. P., Gibson, M. A. and Weiss, A. S. (2001) 'Protein interaction studies of MAGP-1 with tropoelastin and fibrillin-1.', *J Biol Chem,* 276(43), 39661-6.

Jensen, S. A., Robertson, I. B. and Handford, P. A. (2012) 'Dissecting the fibrillin microfibril: structural insights into organization and function.', *Structure,* 20(2), 215-25.

Jordan, C. D., Bohling, S. D., Charbonneau, N. L. and Sakai, L. Y. (2010) 'Fibrillins in adult human ovary and polycystic ovary syndrome: is fibrillin-3 affected in PCOS?', *J Histochem Cytochem,* 58(10), 903-15.

Jovanovic, J., Takagi, J., Choulier, L., Abrescia, N. G., Stuart, D. I., van der Merwe, P. A., Mardon, H. J. and Handford, P. A. (2007) 'alphaVbeta6 is a novel receptor for human fibrillin-1. Comparative studies of molecular determinants underlying integrin-rgd affinity and specificity.', *J Biol Chem,* 282(9), 6743-51.

Judge, D. P., Biery, N. J., Keene, D. R., Geubtner, J., Myers, L., Huso, D. L., Sakai, L. Y. and Dietz, H. C. (2004) 'Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome.', *J Clin Invest,* 114(2), 172-81.

Kan, Z., Jaiswal, B. S., Stinson, J., Janakiraman, V., Bhatt, D., Stern, H. M., Yue, P., Haverty, P. M., Bourgon, R., Zheng, J., Moorhead, M., Chaudhuri, S., Tomsho, L. P., Peters, B. A., Pujara, K.,

Cordes, S., Davis, D. P., Carlton, V. E., Yuan, W., Li, L., Wang, W., Eigenbrot, C., Kaminker, J. S., Eberhard, D. A., Waring, P., Schuster, S. C., Modrusan, Z., Zhang, Z., Stokoe, D., de Sauvage, F. J., Faham, M. and Seshagiri, S. (2010) 'Diverse somatic mutation patterns and pathway alterations in human cancers', *Nature,* 466(7308), 869-73.

Kay, L., Brooks, B., Sparks, S., Torchia, D. and Bax, A. (1989a) 'Measurement of NH-C-alpha-H coupling-constants in staphylococcal nuclease by two-dimensional NMR and comparison with X-ray crystallographic result', *J Am Chem Soc,* 111(14), 5488-5490.

Kay, L., Ikura, M., Tsudin, R. and Bax, A. (1990) '3-dimensional triple-resonance NMR-spectroscopy of isotopically enriched proteins', *J Magn Reson,* 89(3), 496-514.

Kay, L., Torchia, D. and Bax, A. (1989b) 'Backbone dynamics of proteins as studied by N-15 inverse detected heteronuclear NMR-spectroscopy - application to staphylococcal nuclease', *Biochemistry*, 8972-8979.

Keene, D. R., Maddox, B. K., Kuo, H. J., Sakai, L. Y. and Glanville, R. W. (1991) 'Extraction of extendable beaded structures and their identification as fibrillin-containing extracellular matrix microfibrils.', *J Histochem Cytochem,* 39(4), 441-9.

Kettle, S., Cardy, C. M., Hutchinson, S., Sykes, B. and Handford, P. A. (2000) 'Characterisation of fibrillin-1 cDNA clones in a human fibroblast cell line that assembles microfibrils', *Int J Biochem Cell Biol,* 32(2), 201.

Kettle, S., Yuan, X., Grundy, G., Knott, V., Downing, A. K. and Handford, P. A. (1999) 'Defective calcium binding to fibrillin-1: consequence of an N2144S change for fibrillin-1 structure and function', *J Mol Biol,* 285(3), 1277-87.

Kielty, C. M., Cummings, C., Whittaker, S. P., Shuttleworth, C. A. and Grant, M. E. (1991) 'Isolation and ultrastructural analysis of microfibrillar structures from foetal bovine elastic tissues. Relative abundance and supramolecular architecture of type VI collagen assemblies and fibrillin', *J Cell Sci,* 99 (Pt 4), 797.

Kielty, C. M., Phillips, J. E., Child, A. H., Pope, F. M. and Shuttleworth, C. A. (1994) 'Fibrillin secretion and microfibril assembly by Marfan dermal fibroblasts.', *Matrix Biol,* 14(2), 191-9.

Kielty, C. M. and Shuttleworth, C. A. (1993) 'The role of calcium in the organization of fibrillin microfibrils.', *FEBS Lett,* 336(2), 323-6.

Kielty, C. M., Wess, T. J., Haston, L., Ashworth, J. L., Sherratt, M. J. and Shuttleworth, C. A. (2002) 'Fibrillin-rich microfibrils: elastic biopolymers of the extracellular matrix', *J Muscle Res Cell Motil,* 23(5-6), 581.

Kinsey, R., Williamson, M. R., Chaudhry, S., Mellody, K. T., McGovern, A., Takahashi, S., Shuttleworth, C. A. and Kielty, C. M. (2008) 'Fibrillin-1 microfibril deposition is dependent on fibronectin assembly.', *J Cell Sci,* 121(Pt 16), 2696-704.

Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983) 'Optimization by simulated annealing', *Science,* 220(4598), 671-80.

Kirschner, R., Hubmacher, D., Iyengar, G., Kaur, J., Fagotto-Kaufmann, C., Brömme, D., Bartels, R. and Reinhardt, D. P. (2011) 'Classical and neonatal Marfan syndrome mutations in fibrillin-1 cause differential protease susceptibilities and protein function.', *J Biol Chem,* 286(37), 32810-23.

Knott, V., Downing, A. K., Cardy, C. M. and Handford, P. (1996) 'Calcium binding properties of an epidermal growth factor-like domain pair from human fibrillin-1', *J Mol Biol,* 255(1), 22.

Kramer, F., Deshmukh, M., Kessler, H. and Glaser, S. (2004) 'Residual dipolar coupling constants: An elementary derivation of key equations', *Concepts in Magnetic Resonance Part a,* 21A(1), 10-21.

Kuo, C. L., Isogai, Z., Keene, D. R., Hazeki, N., Ono, R. N., Sengle, G., Bachinger, H. P. and Sakai, L. Y. (2007) 'Effects of fibrillin-1 degradation on microfibril ultrastructure', *J Biol Chem,* 282(6), 4007.

Kuszewski, J., Gronenborn, A. and Clore, G. (1996) 'Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases.', *Protein Sci,* 5(6), 1067-80.

Kutz, W. E., Wang, L. W., Bader, H. L., Majors, A. K., Iwata, K., Traboulsi, E. I., Sakai, L. Y., Keene, D. R. and Apte, S. S. (2011) 'ADAMTS10 protein interacts with fibrillin-1 and promotes its deposition in extracellular matrix of cultured fibroblasts.', *J Biol Chem,* 286(19), 17156-67.

Körkkö, J., Kaitila, I., Lönnqvist, L., Peltonen, L. and Ala-Kokko, L. (2002) 'Sensitivity of conformation sensitive gel electrophoresis in detecting mutations in Marfan syndrome and related conditions', *J Med Genet,* 39(1), 34-41.

Lack, J., O'Leary, J. M., Knott, V., Yuan, X., Rifkin, D. B., Handford, P. A. and Downing, A. K. (2003) 'Solution structure of the third TB domain from LTBP1 provides insight into assembly of the large latent complex that sequesters latent TGF-beta.', *J Mol Biol,* 334(2), 281-91.

Laemmli, U. K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature,* 227(5259), 680-5.

Laskowski, R., Rullmannn, J., MacArthur, M., Kaptein, R. and Thornton, J. (1996) 'AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.', *J Biomol NMR,* 8(4), 477-86.

Le Goff, C., Mahaut, C., Wang, L. W., Allali, S., Abhyankar, A., Jensen, S., Zylberberg, L., Collod-Beroud, G., Bonnet, D., Alanay, Y., Brady, A. F., Cordier, M. P., Devriendt, K., Genevieve, D., Kiper, P., Kitoh, H., Krakow, D., Lynch, S. A., Le Merrer, M., Mégarbane, A., Mortier, G., Odent, S., Polak, M., Rohrbach, M., Sillence, D., Stolte-Dijkstra, I., Superti-Furga, A., Rimoin, D. L., Topouchian, V., Unger, S., Zabel, B., Bole-Feysot, C., Nitschke, P., Handford, P., Casanova, J. L., Boileau, C., Apte, S. S., Munnich, A. and Cormier-Daire, V. (2011) 'Mutations in the TGFβ binding-protein-like domain 5 of FBN1 are responsible for acromicric and geleophysic dysplasias.', *Am J Hum Genet,* 89(1), 7-14.

Lee, B., Godfrey, M., Vitale, E., Hori, H., Mattei, M. G., Sarfarazi, M., Tsipouras, P., Ramirez, F. and Hollister, D. W. (1991) 'Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes.', *Nature,* 352(6333), 330-4.

Lee, S. S., Knott, V., Jovanovic, J., Harlos, K., Grimes, J. M., Choulier, L., Mardon, H. J., Stuart, D. I. and Handford, P. A. (2004) 'Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization', *Structure,* 12(4), 717.

Lin, G., Tiedemann, K., Vollbrandt, T., Peters, H., Batge, B., Brinckmann, J. and Reinhardt, D. P. (2002) 'Homo- and heterotypic fibrillin-1 and -2 interactions constitute the basis for the assembly of microfibrils', *J Biol Chem,* 277(52), 50795.

Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. and Nilges, M. (2003) 'Refinement of protein structures in explicit solvent', *Proteins,* 50(3), 496-506.

Lipsitz, R. S. and Tjandra, N. (2004) 'Residual dipolar couplings in NMR structure analysis', *Annu Rev Biophys Biomol Struct,* 33, 387-413.

Liu, W. O., Oefner, P. J., Qian, C., Odom, R. S. and Francke, U. (1997) 'Denaturing HPLC-identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders', *Genet Test,* 1(4), 237-42.

Loeys, B., De Backer, J., Van Acker, P., Wettinck, K., Pals, G., Nuytinck, L., Coucke, P. and De Paepe, A. (2004) 'Comprehensive molecular screening of the FBN1 gene favors locus homogeneity of classical Marfan syndrome', *Hum Mutat,* 24(2), 140-6.

Loeys, B., Nuytinck, L., Delvaux, I., De Bie, S. and De Paepe, A. (2001) 'Genotype and phenotype analysis of 171 patients referred for molecular study of the fibrillin-1 gene FBN1 because of suspected Marfan syndrome', *Arch Intern Med,* 161(20), 2447-54.

Loeys, B. L., Chen, J., Neptune, E. R., Judge, D. P., Podowski, M., Holm, T., Meyers, J., Leitch, C. C., Katsanis, N., Sharifi, N., Xu, F. L., Myers, L. A., Spevak, P. J., Cameron, D. E., De Backer, J., Hellemans, J., Chen, Y., Davis, E. C., Webb, C. L., Kress, W., Coucke, P., Rifkin, D. B., De Paepe, A. M. and Dietz, H. C. (2005) 'A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2.', *Nat Genet,* 37(3), 275-81.

Loeys, B. L., Gerber, E. E., Riegert-Johnson, D., Iqbal, S., Whiteman, P., McConnell, V., Chillakuri, C. R., Macaya, D., Coucke, P. J., De Paepe, A., Judge, D. P., Wigley, F., Davis, E. C., Mardon, H. J., Handford, P., Keene, D. R., Sakai, L. Y. and Dietz, H. C. (2010) 'Mutations in fibrillin-1 cause congenital scleroderma: stiff skin syndrome.', *Sci Transl Med,* 2(23), 23ra20.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) 'Structure validation by Calpha geometry: phi,psi and Cbeta deviation', *Proteins,* 50(3), 437-50.

Lönnqvist, L., Reinhardt, D., Sakai, L. and Peltonen, L. (1998) 'Evidence for furin-type activity-mediated C-terminal processing of profibrillin-1 and interference in the processing by certain mutations.', *Hum Mol Genet,* 7(13), 2039-44.

Malinauskas, T., Aricescu, A. R., Lu, W., Siebold, C. and Jones, E. Y. (2011) 'Modular mechanism of Wnt signaling inhibition by Wnt inhibitory factor 1', *Nat Struct Mol Biol,* 18(8), 886-93.

Marion, D., Driscoll, P., Kay, L., Wingfield, P., Bax, A., Gronenborn, A. and Clore, G. (1989) 'Overcoming the overlap problem in the assignment of H-1-NMR spectra of larger proteins by use of 3-dimensional heteronuclear H1-N-15 Hartmann-Hahn multiple quantum coherence and

nuclear Overhauser Quantum coherence spectroscopy - application to interleukin-1-beta', *Biochemistry,* 28(15), 6150-6156.

Marson, A., Rock, M. J., Cain, S. A., Freeman, L. J., Morgan, A., Mellody, K., Shuttleworth, C. A., Baldock, C. and Kielty, C. M. (2005) 'Homotypic fibrillin-1 interactions in microfibril assembly', *J Biol Chem,* 280(6), 5013.

Maslen, C. L., Corson, G. M., Maddox, B. K., Glanville, R. W. and Sakai, L. Y. (1991) 'Partial sequence of a candidate gene for the Marfan syndrome.', *Nature,* 352(6333), 334-7.

Massagué, J. (1998) 'TGF-beta signal transduction', *Annu Rev Biochem,* 67, 753-91.

Massam-Wu, T., Chiu, M., Choudhury, R., Chaudhry, S. S., Baldwin, A. K., McGovern, A., Baldock, C., Shuttleworth, C. A. and Kielty, C. M. (2010) 'Assembly of fibrillin microfibrils governs extracellular deposition of latent TGF beta.', *J Cell Sci,* 123(Pt 17), 3006-18.

Milewicz, D. M., Grossfield, J., Cao, S. N., Kielty, C., Covitz, W. and Jewett, T. (1995) 'A mutation in FBN1 disrupts profibrillin processing and results in isolated skeletal features of the Marfan syndrome.', *J Clin Invest,* 95(5), 2373-8.

McGettrick, A.J. (2002). 'Molecular consequences of mutations in *FBN1.*' Biochemistry D.Phil. thesis, University of Oxford.

McNaught-Davis, J. (2011). 'The role of the fibrillin- Nterm-EGF3 region in microfibril assembly.' Biochemistry Final Honour School Part II Project, University of Oxford

Miner, J. and Mecham, R. (2011) 'Basement Membranes', *Extracellular Matrix: an Overview*, 117-145.

Molloy, S. S., Anderson, E. D., Jean, F. and Thomas, G. (1999) 'Bi-cycling the furin pathway: from TGN localization to pathogen activation and embryogenesis.', *Trends Cell Biol,* 9(1), 28-35.

Morgan, T. L., Yang, D. J., Fry, D. G., Hurlin, P. J., Kohler, S. K., Maher, V. M. and McCormick, J. J. (1991) 'Characteristics of an infinite life span diploid human fibroblast cell strain and a near-diploid strain arising from a clone of cells expressing a transfected v-myc oncogene', *Exp Cell Res,* 197(1), 125-36.

Mátyás, G., De Paepe, A., Halliday, D., Boileau, C., Pals, G. and Steinmann, B. (2002) 'Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: Identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene', *Hum Mutat,* 19(4), 443-56.

Neptune, E. R., Frischmeyer, P. A., Arking, D. E., Myers, L., Bunton, T. E., Gayraud, B., Ramirez, F., Sakai, L. Y. and Dietz, H. C. (2003) 'Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome.', *Nat Genet,* 33(3), 407-11.

Nijbroek, G., Sood, S., McIntosh, I., Francomano, C. A., Bull, E., Pereira, L., Ramirez, F., Pyeritz, R. E. and Dietz, H. C. (1995) 'Fifteen novel FBN1 mutations causing Marfan syndrome detected by heteroduplex analysis of genomic amplicons', *Am J Hum Genet,* 57(1), 8-21.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988) 'Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a

random array of atoms. Circumventing problems associated with folding', *FEBS Lett,* 239(1), 129-36.

Norwood, T., Boyd, J., Heritage, J., Soffe, N. and Campbell, I. (1990) 'Comparison of techinques for H-1 detected heteronuclear H1-N-15 spectroscopy', *J Magn Reson,* 87(3), 488-501.

O'Callaghan, C. A., Byford, M. F., Wyer, J. R., Willcox, B. E., Jakobsen, B. K., McMichael, A. J. and Bell, J. I. (1999) 'BirA enzyme: production and application in the study of membrane receptor-ligand interactions by site-specific biotinylation', *Anal Biochem,* 266(1), 9-15.

Ono, R. N., Sengle, G., Charbonneau, N. L., Carlberg, V., Bachinger, H. P., Sasaki, T., Lee-Arteaga, S., Zilberberg, L., Rifkin, D. B., Ramirez, F., Chu, M. L. and Sakai, L. Y. (2009) 'Latent Transforming Growth Factor {beta}-binding Proteins and Fibulins Compete for Fibrillin-1 and Exhibit Exquisite Specificities in Binding Sites', *J Biol Chem,* 284(25), 16872.

Ottiger, M. and Bax, A. (1998) 'Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules', *J Biomol NMR,* 12(3), 361-72.

Ottiger, M. and Bax, A. (1999) 'Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values', *J Biomol NMR,* 13(2), 187-91.

Ottiger, M., Delaglio, F. and Bax, A. (1998) 'Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra', *J Magn Reson,* 131(2), 373-8.

Palz, M., Tiecke, F., Booms, P., Göldner, B., Rosenberg, T., Fuchs, J., Skovby, F., Schumacher, H., Kaufmann, U. C., von Kodolitsch, Y., Nienaber, C. A., Leitner, C., Katzke, S., Vetter, B., Hagemeier, C. and Robinson, P. N. (2000) 'Clustering of mutations associated with mild Marfan-like phenotypes in the 3' region of FBN1 suggests a potential genotype-phenotype correlation', *Am J Med Genet,* 91(3), 212-21.

Pepe, G., Giusti, B., Attanasio, M., Comeglio, P., Porciani, M. C., Giurlani, L., Montesi, G. F., Calamai, G. C., Vaccari, M., Favilli, S., Abbate, R. and Gensini, G. F. (1997) 'A major involvement of the cardiovascular system in patients affected by Marfan syndrome: novel mutations in fibrillin 1 gene', *J Mol Cell Cardiol,* 29(7), 1877-84.

Pereira, L., Andrikopoulos, K., Tian, J., Lee, S. Y., Keene, D. R., Ono, R., Reinhardt, D. P., Sakai, L. Y., Biery, N. J., Bunton, T., Dietz, H. C. and Ramirez, F. (1997) 'Targetting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome.', *Nat Genet,* 17(2), 218-22.

Pereira, L., D'Alessio, M., Ramirez, F., Lynch, J. R., Sykes, B., Pangilinan, T. and Bonadio, J. (1993) 'Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome', *Hum Mol Genet,* 2(7), 961.

Pereira, L., Lee, S. Y., Gayraud, B., Andrikopoulos, K., Shapiro, S. D., Bunton, T., Biery, N. J., Dietz, H. C., Sakai, L. Y. and Ramirez, F. (1999) 'Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1.', *Proc Natl Acad Sci U S A,* 96(7), 3819-23.

Pfaff, M., Reinhardt, D. P., Sakai, L. Y. and Timpl, R. (1996) 'Cell adhesion and integrin binding to recombinant human fibrillin-1.', *FEBS Lett,* 384(3), 247-50.

Piantini, U., Sorensen, O. and Ernst, R. (1982) 'Multiple quantum filters for elucidating NMR coupling networks', *J Am Chem Soc,* 104(24), 6800-6801.

Piha-Gossack, A., Sossin, W. and Reinhardt, D. P. (2012) 'The evolution of extracellular fibrillins and their functional domains.', *PLoS One,* 7(3), e33560.

Qian, R. and Glanville, R. (1997) 'Alignment of fibrillin molecules in elastic microfibrils is defined by transglutaminase-derived cross-links.', *Biochemistry,* 36(50), 15841-7.

Raghunath, M., Putnam, E. A., Ritty, T., Hamstra, D., Park, E. S., Tschodrich-Rotter, M., Peters, R., Rehemtulla, A. and Milewicz, D. M. (1999) 'Carboxy-terminal conversion of profibrillin to fibrillin at a basic site by PACE/furin-like activity required for incorporation in the matrix', *J Cell Sci,* 112 (Pt 7), 1093.

Raja-Khan, N., Kunselman, A. R., Demers, L. M., Ewens, K. G., Spielman, R. S. and Legro, R. S. (2010) 'A variant in the fibrillin-3 gene is associated with TGF-β and inhibin B levels in women with polycystic ovary syndrome.', *Fertil Steril,* 94(7), 2916-9.

Raleigh, E. A., Murray, N. E., Revel, H., Blumenthal, R. M., Westaway, D., Reith, A. D., Rigby, P. W., Elhai, J. and Hanahan, D. (1988) 'McrA and McrB restriction phenotypes of some E. coli strains and implications for gene cloning', *Nucleic Acids Res,* 16(4), 1563-75.

Rao, Z., Handford, P. A., Knott, V., Mayhew, M., Brownlee, G. G. and Stuart, D. (1995) 'Crystallization of a calcium-binding EGF-like domain.', *Acta Crystallogr D Biol Crystallogr,* 51(Pt 3), 402-3.

Reber-Müller, S., Spissinger, T., Schuchert, P., Spring, J. and Schmid, V. (1995) 'An extracellular matrix protein of jellyfish homologous to mammalian fibrillins forms different fibrils depending on the life stage of the animal.', *Dev Biol,* 169(2), 662-72.

Reinhardt, D. P., Gambee, J. E., Ono, R. N., Bachinger, H. P. and Sakai, L. Y. (2000a) 'Initial steps in assembly of microfibrils. Formation of disulfide-cross-linked multimers containing fibrillin-1', *J Biol Chem,* 275(3), 2205.

Reinhardt, D. P., Keene, D. R., Corson, G. M., Pöschl, E., Bächinger, H. P., Gambee, J. E. and Sakai, L. Y. (1996) 'Fibrillin-1: organization in microfibrils and structural properties.', *J Mol Biol,* 258(1), 104-16.

Reinhardt, D. P., Ono, R. N., Notbohm, H., Müller, P. K., Bächinger, H. P. and Sakai, L. Y. (2000b) 'Mutations in calcium-binding epidermal growth factor modules render fibrillin-1 susceptible to proteolysis. A potential disease-causing mechanism in Marfan syndrome', *J Biol Chem,* 275(16), 12339-45.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E. and Nilges, M. (2007) 'ARIA2: automated NOE assignment and data integration in NMR structure calculation', *Bioinformatics,* 23(3), 381-2.

Ritty, T. M., Broekelmann, T., Tisdale, C., Milewicz, D. M. and Mecham, R. P. (1999) 'Processing of the fibrillin-1 carboxyl-terminal domain.', *J Biol Chem,* 274(13), 8933-40.

Ritty, T. M., Broekelmann, T. J., Werneck, C. C. and Mecham, R. P. (2003) 'Fibrillin-1 and -2 contain heparin-binding sites important for matrix deposition and that support cell attachment.', *Biochem J,* 375(Pt 2), 425-32.

Robertson, I., Jensen, S. and Handford, P. (2011) 'TB domain proteins: evolutionary insights into the multifaceted roles of fibrillins and LTBPs.', *Biochem J,* 433(2), 263-76.

Robinson, P. N., Booms, P., Katzke, S., Ladewig, M., Neumann, L., Palz, M., Pregla, R., Tiecke, F. and Rosenberg, T. (2002) 'Mutations of FBN1 and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies', *Hum Mutat,* 20(3), 153-61.

Rock, M. J., Cain, S. A., Freeman, L. J., Morgan, A., Mellody, K., Marson, A., Shuttleworth, C. A., Weiss, A. S. and Kielty, C. M. (2004) 'Molecular basis of elastic fiber formation. Critical interactions and a tropoelastin-fibrillin-1 cross-link.', *J Biol Chem,* 279(22), 23748-58.

Rommel, K., Karck, M., Haverich, A., von Kodolitsch, Y., Rybczynski, M., Müller, G., Singh, K. K., Schmidtke, J. and Arslan-Kirchner, M. (2005) 'Identification of 29 novel and nine recurrent fibrillin-1 (FBN1) mutations and genotype-phenotype correlations in 76 patients with Marfan syndrome', *Hum Mutat,* 26(6), 529-39.

Sabatier, L., Chen, D., Fagotto-Kaufmann, C., Hubmacher, D., McKee, M., Annis, D., Mosher, D. and Reinhardt, D. (2009) 'Fibrillin assembly requires fibronectin.', *Mol Biol Cell,* 20(3), 846-58.

Saharinen, J. and Keski-Oja, J. (2000) 'Specific sequence motif of 8-Cys repeats of TGF-beta binding proteins, LTBPs, creates a hydrophobic interaction surface for binding of small latent TGF-beta.', *Mol Biol Cell,* 11(8), 2691-704.

Sakai, H., Visser, R., Ikegawa, S., Ito, E., Numabe, H., Watanabe, Y., Mikami, H., Kondoh, T., Kitoh, H., Sugiyama, R., Okamoto, N., Ogata, T., Fodde, R., Mizuno, S., Takamura, K., Egashira, M., Sasaki, N., Watanabe, S., Nishimaki, S., Takada, F., Nagai, T., Okada, Y., Aoka, Y., Yasuda, K., Iwasa, M., Kogaki, S., Harada, N., Mizuguchi, T. and Matsumoto, N. (2006) 'Comprehensive genetic analysis of relevant four genes in 49 patients with Marfan syndrome or Marfan-related phenotypes.', *Am J Med Genet A,* 140(16), 1719-25.

Sakai, L., Keene, D. and Engvall, E. (1986) 'Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils.', *J Cell Biol,* 103(6 Pt 1), 2499-509.

Sakai, L. Y., Keene, D. R., Glanville, R. W. and Bachinger, H. P. (1991) 'Purification and partial characterization of fibrillin, a cysteine-rich structural component of connective tissue microfibrils', *J Biol Chem,* 266(22), 14763.

Sampoli Benitez, B. A., Hunter, M. J., Meininger, D. P. and Komives, E. A. (1997) 'Structure of the fifth EGF-like domain of thrombomodulin: An EGF-like domain with a novel disulfide-bonding pattern.', *J Mol Biol,* 273(4), 913-26.

Sanders, C. R. and Schwonek, J. P. (1992) 'Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR', *Biochemistry,* 31(37), 8898-905.

Santoro, J. and King, G. (1992) 'A constant-time 2D Overbodenhausen experiment for inverse correlation of isotopically enriched species', *J Magn Reson,* 97(1), 202-207.

Schatz, P. J. (1993) 'Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli', *Biotechnology (N Y),* 11(10), 1138-43.

Schneider, R., Jensen, S. A., Whiteman, P., McCullagh, J. S., Redfield, C. and Handford, P. A. (2010) 'Biophysical characterisation of fibulin-5 proteins associated with disease', *J Mol Biol,* 401(4), 605-17.

Schrijver, I., Liu, W., Brenn, T., Furthmayr, H. and Francke, U. (1999) 'Cysteine substitutions in epidermal growth factor-like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes.', *Am J Hum Genet,* 65(4), 1007-20.

Schwieters, C. and Clore, G. (2001) 'Internal coordinates for molecular dynamics and minimization in structure determination and refinement.', *J Magn Reson,* 152(2), 288-302.

Schwieters, C., Kuszewski, J. and Clore, G. (2006) 'Using Xplor-NIH for NMR molecular structure determination', *Progress in Nuclear Magnetic Resonance Spectroscopy,* 48(1), 47-62.

Schwieters, C., Kuszewski, J., Tjandra, N. and Clore, G. (2003) 'The Xplor-NIH NMR molecular structure determination package.', *J Magn Reson,* 160(1), 65-73.

Sengle, G., Charbonneau, N. L., Ono, R. N., Sasaki, T., Alvarez, J., Keene, D. R., Bächinger, H. P. and Sakai, L. Y. (2008a) 'Targeting of bone morphogenetic protein growth factor complexes to fibrillin.', *J Biol Chem,* 283(20), 13874-88.

Sengle, G., Ono, R. N., Lyons, K. M., Bächinger, H. P. and Sakai, L. Y. (2008b) 'A new model for growth factor activation: type II receptors compete with the prodomain for BMP-7.', *J Mol Biol,* 381(4), 1025-39.

Sengle, G., Ono, R. N., Sasaki, T. and Sakai, L. Y. (2011) 'Prodomains of transforming growth factor beta (TGFbeta) superfamily members specify different functions: extracellular matrix interactions and growth factor bioavailability.', *J Biol Chem,* 286(7), 5087-99.

Sengle, G., Tsutsui, K., Keene, D. R., Tufa, S. F., Carlson, E. J., Charbonneau, N. L., Ono, R. N., Sasaki, T., Wirtz, M. K., Samples, J. R., Fessler, L. I., Fessler, J. H., Sekiguchi, K., Hayflick, S. J. and Sakai, L. Y. (2012) 'Microenvironmental regulation by fibrillin-1.', *PLoS Genet,* 8(1), e1002425.

Shattil, S. J., Kim, C. and Ginsberg, M. H. (2010) 'The final steps of integrin activation: the end game.', *Nat Rev Mol Cell Biol,* 11(4), 288-300.

Shen, Y. and Bax, A. (2010) 'Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts', *J Biomol NMR,* 46(3), 199-204.

Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) 'TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.', *J Biomol NMR,* 44(4), 213-23.

Shi, M., Zhu, J., Wang, R., Chen, X., Mi, L., Walz, T. and Springer, T. A. (2011) 'Latent TGF-β structure and activation', *Nature,* 474(7351), 343-9.

Smallridge, R., Whiteman, P., Werner, J., Campbell, I., Handford, P. and Downing, A. (2003) 'Solution structure and dynamics of a calcium binding epidermal growth factor-like domain pair from the neonatal region of human fibrillin-1.', *J Biol Chem,* 278(14), 12199-206.

Smallridge, R. S., Whiteman, P., Doering, K., Handford, P. A. and Downing, A. K. (1999) 'EGF-like domain calcium affinity modulated by N-terminal domain linkage in human fibrillin-1.', *J Mol Biol,* 286(3), 661-8.

Stenberg, E., Persson, B., Roos, H. and Urbaniczky, C. (1991) 'Quantitative-determination of surface concentration of protein with surface-plasmon resonance using radiolabeled proteins', *J Colloid Interf Sci,* 143(2), 513-526.

Stheneur, C., Collod-Béroud, G., Faivre, L., Buyck, J. F., Gouya, L., Le Parc, J. M., Moura, B., Muti, C., Grandchamp, B., Sultan, G., Claustres, M., Aegerter, P., Chevallier, B., Jondeau, G. and Boileau, C. (2009) 'Identification of the minimal combination of clinical features in probands for efficient mutation detection in the FBN1 gene', *Eur J Hum Genet,* 17(9), 1121-8.

Stheneur, C., Faivre, L., Collod-Béroud, G., Gautier, E., Binquet, C., Bonithon-Kopp, C., Claustres, M., Child, A. H., Arbustini, E., Adès, L. C., Francke, U., Mayer, K., Arslan-Kirchner, M., De Paepe, A., Chevallier, B., Bonnet, D., Jondeau, G. and Boileau, C. (2011) 'Prognosis factors in probands with an FBN1 mutation diagnosed before the age of 1 year.', *Pediatr Res,* 69(3), 265-70.

Ståhl-Hallengren, C., Ukkonen, T., Kainulainen, K., Kristofersson, U., Saxne, T., Tornqvist, K. and Peltonen, L. (1994) 'An extra cysteine in one of the non-calcium-binding epidermal growth factor-like motifs of the FBN1 polypeptide is connected to a novel variant of Marfan syndrome', *J Clin Invest,* 94(2), 709-13.

Suk, J.Y. (2003). 'Molecular consequences of protein-misfolding mutations in *FBN1'*. Biochemistry D.Phil. thesis, University of Oxford.

Suk, J. Y., Jensen, S., McGettrick, A., Willis, A. C., Whiteman, P., Redfield, C. and Handford, P. A. (2004) 'Structural consequences of cysteine substitutions C1977Y and C1977R in calcium-binding epidermal growth factor-like domain 30 of human fibrillin-1.', *J Biol Chem,* 279(49), 51258-65.

Thurmond, F. and Trotter, J. (1996) 'Morphology and biomechanics of the microfibrillar network of sea cucumber dermis', *J Exp Biol,* 199(Pt 8), 1817-28.

Tiedemann, K., Bätge, B., Müller, P. K. and Reinhardt, D. P. (2001) 'Interactions of fibrillin-1 with heparin/heparan sulfate, implications for microfibrillar assembly.', *J Biol Chem,* 276(38), 36035-42.

Tiedemann, K., Sasaki, T., Gustafsson, E., Göhring, W., Bätge, B., Notbohm, H., Timpl, R., Wedel, T., Schlötzer-Schrehardt, U. and Reinhardt, D. P. (2005) 'Microfibrils at basement membrane zones interact with perlecan via fibrillin-1.', *J Biol Chem,* 280(12), 11404-12.

Tjandra, N. and Bax, A. (1997) 'Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium', *Science,* 278(5340), 1111-4.

Todorovic, V. and Rifkin, D. B. (2012) 'LTBPs, more than just an escort service.', *J Cell Biochem,* 113(2), 410-8.

Tran, D. Q. (2012) 'TGF-β: the sword, the wand, and the shield of FOXP3(+) regulatory T cells', *J Mol Cell Biol,* 4(1), 29-37.

Trask, T., Ritty, T., Broekelmann, T., Tisdale, C. and Mecham, R. (1999) 'N-terminal domains of fibrillin 1 and fibrillin 2 direct the formation of homodimers: a possible first step in microfibril assembly.', *Biochem J,* 340 (Pt 3), 693-701.

Trask, T. M., Trask, B. C., Ritty, T. M., Abrams, W. R., Rosenbloom, J. and Mecham, R. P. (2000) 'Interaction of tropoelastin with the amino-terminal domains of fibrillin-1 and fibrillin-2 suggests a role for the fibrillins in elastic fiber assembly.', *J Biol Chem,* 275(32), 24400-6.

Tsuruga, E., Sato, A., Ueki, T., Nakashima, K., Nakatomi, Y., Ishikawa, H., Yajima, T. and Sawa, Y. (2009) 'Integrin alphavbeta3 regulates microfibril assembly in human periodontal ligament cells.', *Tissue Cell,* 41(2), 85-9.

Tsutsui, K., Manabe, R., Yamada, T., Nakano, I., Oguri, Y., Keene, D. R., Sengle, G., Sakai, L. Y. and Sekiguchi, K. (2010) 'ADAMTSL-6 is a novel extracellular matrix protein that binds to fibrillin-1 and promotes fibrillin-1 fibril formation.', *J Biol Chem,* 285(7), 4870-82.

Uyeda, T., Takahashi, T., Eto, S., Sato, T., Xu, G., Kanezaki, R., Toki, T., Yonesaka, S. and Ito, E. (2004) 'Three novel mutations of the fibrillin-1 gene and ten single nucleotide polymorphisms of the fibrillin-3 gene in Marfan syndrome patients', *J Hum Genet,* 49(8), 404-7.

Vranken, W., Boucher, W., Stevens, T., Fogh, R., Pajon, A., Llinas, M., Ulrich, E., Markley, J., Ionides, J. and Laue, E. (2005) 'The CCPN data model for NMR spectroscopy: development of a software pipeline.', *Proteins,* 59(4), 687-96.

Vuister, G. and Bax, A. (1992) 'Resolution enhancement by spectral editing of uniformly C-13-enriched proteins by homonuclear broad-band C-13 decoupling', *J Magn Reson,* 98(2), 428-435.

Wagenseil, J. E. and Mecham, R. P. (2007) 'New insights into elastic fiber assembly.', *Birth Defects Res C Embryo Today,* 81(4), 229-40.

Waldmüller, S., Müller, M., Warnecke, H., Rees, W., Schöls, W., Walterbusch, G., Ennker, J. and Scheffold, T. (2007) 'Genetic testing in patients with aortic aneurysms/dissections: a novel genotype/phenotype correlation?', *Eur J Cardiothorac Surg,* 31(6), 970-5.

Wallis, D. D., Putnam, E. A., Cretoiu, J. S., Carmical, S. G., Cao, S. N., Thomas, G. and Milewicz, D. M. (2003) 'Profibrillin-1 maturation by human dermal fibroblasts: proteolytic processing and molecular chaperones.', *J Cell Biochem,* 90(3), 641-52.

Wang, M. C., Lu, Y. and Baldock, C. (2009) 'Fibrillin microfibrils: a key role for the interbead region in elasticity.', *J Mol Biol,* 388(1), 168-79.

Wedemeyer, W. J., Welker, E. and Scheraga, H. A. (2002) 'Proline cis-trans isomerization and protein folding', *Biochemistry,* 41(50), 14637-44.

Weinbaum, J. S., Broekelmann, T. J., Pierce, R. A., Werneck, C. C., Segade, F., Craft, C. S., Knutsen, R. H. and Mecham, R. P. (2008) 'Deficiency in microfibril-associated glycoprotein-1 leads to complex phenotypes in multiple organ systems', *J Biol Chem,* 283(37), 25533-43.

Werner, J., Knott, V., Handford, P., Campbell, I. and Downing, A. (2000) 'Backbone dynamics of a cbEGF domain pair in the presence of calcium.', *J Mol Biol,* 296(4), 1065-78.

Wess, T. J., Purslow, P. P. and Kielty, C. M. (1998) 'X-Ray diffraction studies of fibrillin-rich microfibrils: effects of tissue extension on axial and lateral packing.', *J Struct Biol,* 122(1-2), 123-7.

Whiteman, P., Downing, A. K., Smallridge, R., Winship, P. R. and Handford, P. A. (1998b) 'A Gly --> Ser change causes defective folding in vitro of calcium-binding epidermal growth factor-like domains from factor IX and fibrillin-1.', *J Biol Chem,* 273(14), 7807-13.

Whiteman, P. and Handford, P. A. (2003) 'Defective secretion of recombinant fragments of fibrillin-1: implications of protein misfolding for the pathogenesis of Marfan syndrome and related disorders.', *Hum Mol Genet,* 12(7), 727-37.

Whiteman, P., Hutchinson, S. and Handford, P. A. (2006) 'Fibrillin-1 misfolding and disease', *Antioxid Redox Signal,* 8(3-4), 338-46.

Whiteman, P., Smallridge, R. S., Knott, V., Cordle, J. J., Downing, A. K. and Handford, P. A. (2001) 'A G1127S change in calcium-binding epidermal growth factor-like domain 13 of human fibrillin-1 causes short range conformational effects.', *J Biol Chem,* 276(20), 17156-62.

Whiteman, P., Willis, A. C., Warner, A., Brown, J., Redfield, C. and Handford, P. A. (2007) 'Cellular and molecular studies of Marfan syndrome mutations identify co-operative protein folding in the cbEGF12-13 region of fibrillin-1.', *Hum Mol Genet,* 16(8), 907-18.

Wight, T., Toole, B., Hascall, V. and Mecham, R. (2011) 'Hyaluronan and the Aggregating Proteoglycans', *Extracellular Matrix: an Overview*, 147-195.

Yanagisawa, H. and Davis, E. C. (2010) 'Unraveling the mechanism of elastic fiber assembly: The roles of short fibulins.', *Int J Biochem Cell Biol,* 42(7), 1084-93.

Yuan, X., Downing, A., Knott, V. and Handford, P. (1997) 'Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils.', *EMBO J,* 16(22), 6659-66.

Yuan, X., Werner, J., Lack, J., Knott, V., Handford, P., Campbell, I. and Downing, A. (2002) 'Effects of the N2144S mutation on backbone dynamics of a TB-cbEGF domain pair from human fibrillin-1.', *J Mol Biol,* 316(1), 113-25.

Appendix

Appendix A: references for disease-linked mutations in *FBN1* affecting the FUN-EGF3 and cbEGF41-43 regions of fibrillin-1

Table A1 Mutations affecting the FUN-EGF3 region (exons 2 to 5 of *FBN1*). Information for many of the mutations was obtained from the human gene mutation database (HGMD; http://www.hgmd.cf.ac.uk/ac/index.php) and *FBN1* universal mutation database (UMD; http://www.umd.be/FBN1/

Codon change	Amino acid change	Domain	Database	Reference
AAT > GAT	N57 > D	FUN	-	(Chung *et al.*, 2009)
CGT > TGT	R62 > C	FUN	HGMD/UMD	(Körkkö *et al.*, 2002)
TGC > AGC	C68 > S	FUN	-	(Stheneur *et al.*, 2009)
TGG > CGG	W71 > R	FUN	HGMD	(Sakai *et al.*, 2006)
TGT > CGT	C80 > R	FUN	-	(Stheneur *et al.*, 2009)
TGT > TAT	C80 > Y	FUN	-	(Howarth *et al.*, 2007)
TGT > TTT	C89 > F	EGF1	HGMD/UMD	(Loeys *et al.*, 2001)
TGT > TAT	C89 > Y	EGF1	-	(Stheneur *et al.*, 2009)
TGT > TGG	C89 > W	EGF1	-	(Stheneur *et al.*, 2009)
ACT > GCT	T101 > A	EGF1	UMD	(Stheneur *et al.*, 2009)
TGT > CGT	C111 > R	EGF1	HGMD/UMD	(Hayward *et al.*, 1997)
TCC > TGC	S115 > C	EGF1	HGMD/UMD	(Robinson *et al.*, 2002)
CGC > TGC	R122 > C	EGF2	HGMD/UMD	(Black *et al.*, 1998; Comeglio *et al.*, 2002; Loeys *et al.*, 2001; Ståhl-Hallengren *et al.*, 1994)
TGT > TAT	C123 > Y	EGF2	HGMD	(Arbustini *et al.*, 2005)
Information not available	G127 > D	EGF2	-	(Kan *et al.*, 2010)
TGC > TAC	C129 >Y	EGF2	HGMD/UMD	(Nijbroek *et al.*, 1995)
TGT > TAT	C134 > Y	EGF2	-	(Howarth *et al.*, 2007)
TGT > GGT	C134 > G	EGF2	HGMD	(Sakai *et al.*, 2006)
CCT > TCT	P148 > S	EGF3	-	(Stheneur *et al.*, 2009)
TGT > TCT	C154 > S	EGF3	HGMD	(Biggin *et al.*, 2004)
TGT > TGA	C154 > stop	EGF3	HGMD	(Rommel *et al.*, 2005)
TGT > GGT	C160 > G	EGF3	UMD	(Stheneur *et al.*, 2009)
TGT > TAT	C160 > Y	EGF3	-	(Stheneur *et al.*, 2009)
AAT > AGT	N164 > S	EGF3	HGMD/UMD	(Comeglio *et al.*, 2002)
CGT > TGA	R165 > stop	EGF3	HGMD	(Rommel *et al.*, 2005)
TGT > AGT	C166 > S	EGF3	HGMD/UMD	(Biggin *et al.*, 2004)
TGT > TCT	C166 > S	EGF3	HGMD/UMD	(Collod-Béroud *et al.*, 1997)
TGT > TTT	C166 > F	EGF3	HGMD/UMD	(Nijbroek *et al.*, 1995)
TAC > TAG	Y170 > stop	EGF3	HGMD	(Biggin *et al.*, 2004)
CAG > TAG	Q176 > stop	EGF3	HGMD/UMD	(Liu *et al.*, 1997)
CAG > CCG	Q176 > P	EGF3	-	(Waldmüller *et al.*, 2007)
TGT > CGT	C177 > R	EGF3	HGMD	(Arbustini *et al.*, 2005)

Table A2 Mutations affecting the cbEGF41-43 region (exons 62 to 64 of *FBN1*). Information for the mutations was obtained from the human gene mutation database (HGMD; http://www.hgmd.cf.ac.uk/ac/index.php) and *FBN1* universal mutation database (UMD; http://www.umd.be/FBN1/

Codon change	Amino acid change	Domain	Database	Reference
GAG > AAG	E2570 > K	cbEGF41	HGMD	(Arbustini *et al.*, 2005)
TGT > CGT	C2571 > R	cbEGF41	HGMD	(Arbustini *et al.*, 2005)
TGC > CGC	C2577 > R	cbEGF41	HGMD	(Pepe *et al.*, 1997)
TGC > TTC	C2581 > F	cbEGF41	HGMD	(Loeys *et al.*, 2001)
TGC > TGG	C2581 > W	cbEGF41	HGMD	(Körkkö *et al.*, 2002)
ATT > ACT	I2585 > T	cbEGF41	HGMD	(Liu *et al.*, 1997)
TGC > AGC	C2592 > S	cbEGF41	HGMD	Arbustini 2005
CAG > TAG	Q2601 > stop	cbEGF41	HGMD	(Arbustini *et al.*, 2005) (Baumgartner *et al.*, 2005)
TGT > TAT	C2605 > Y	cbEGF41	HGMD	(Rommel *et al.*, 2005)
GAT > GGT	D2607 > G	cbEGF42	HGMD	(Liu *et al.*, 1997)
GAA > AAA	E2610 > K	cbEGF42	HGMD	(Liu *et al.*, 1997)
GGA > AGA	G2618 > R	cbEGF42	HGMD	(Mátyás *et al.*, 2002)
TGT > AGT	C2622 > Y	cbEGF42	HGMD	(Loeys *et al.*, 2004)
CAC > CCC	H2623 > P	cbEGF42	HGMD	(Collod-Béroud *et al.*, 1997)
AAC > AGC	N2624 > S	cbEGF42	HGMD	(Palz *et al.*, 2000)
AAC > AAG	N2624 > K	cbEGF42	HGMD	(Loeys *et al.*, 2001)
GGG > AGG	G2627 > R	cbEGF42	HGMD	(Dietz and Pyeritz, 1995)
TAC > TGC	Y2629 > C	cbEGF42	HGMD	(Arbustini *et al.*, 2005)
TAT > AAT	Y2639 > N	cbEGF42	HGMD	(Mátyás *et al.*, 2002)
TAT > TGT	Y2639 > C	cbEGF42	HGMD	(Mátyás *et al.*, 2002)
TGT > GGT	C2652 > G	cbEGF43	HGMD	(Körkkö *et al.*, 2002)
AGC > CGC	S2660 > R	cbEGF43	HGMD	(Sakai *et al.*, 2006)
TGT > TCT	C2663 > S	cbEGF43	HGMD	(Uyeda *et al.*, 2004)
GGC > GAC	G2668 > D	cbEGF43	HGMD	(Mátyás *et al.*, 2002)
GGC > TGC	G2668 > C	cbEGF43	HGMD	(Loeys *et al.*, 2001)
CGC > TGC	R2680 > C	cbEGF43	HGMD/UMD	(Palz *et al.*, 2000)
TGT > TTT	C2686 > F	cbEGF43	HGMD	(Liu *et al.*, 1997)

Appendix B: oligonucleotide primer sequences

1) pQE-30 cbEGF32-34-BirA cloning

Forward primer

5'-TAGTAGGGATCCATAGAAGGACGATCAGCA GATATGGACGAATGCAAAGAACC-3'

Reverse primer

5'-TAGTAGGTCGACTTTGCACATCCTACGGTCTTC-3'

2) pQE-30 cbEGF41-43-BirA cloning

Forward primer

5'-TAGTAGGGATCCATAGAAGGACGATCAGCAGACGTGGACGAGTGTGAGGG-3'

Reverse primer

5'-TAGTAGGTCGACAACACAGTGCCCTTGGCCTATG-3'

3) pQE-30 EGF2-EGF3 cloning

Forward primer

5'-TAGTAGGGATCCATAGAAGGACGATCAGCATCCAGATCCATACAACACTGC-3'

Reverse primer

5'-TAGTAGAAGCTTCTA TTATTCACACTGGGGTCCAG-3'

4) pQE-30 EGF2-EGF3-BirA cloning

Forward primer

As for EGF2-EGF3

Reverse primer

5'-TAGTAGGTCGACTTCACACTGGGGTCCAGT-3'

5) N57D QuikChange mutagenesis

Forward primer

5'- GCGCTTAAAGGACCCGATGTCTGTGGATCACG-3'

Reverse primer

5'-ACGTGATCCACAGACATCGGGTCCTTTAAGCGC-3'

6) W71R QuikChange mutagenesis

Forward primer

5'-CTGTTGCCCTGGACGGAAAACCTTACCTGGC-3'

Reverse primer

5'-GCCAGGTAAGGTTTTCCGTCCAGGGCAACAG-3'

7) P148S QuikChange mutagenesis

Forward primer

5'- GACTCACTGTGGACAATCTGTTTGTGAAAGTGGC-3'

Reverse primer

5'-GCCACTTTCACAAACAGATTGTCCACAGTGAGTC-3'

8) pKG-52 sequencing primers

Forward primer

5'- GGTCCACTTCGCATATTAAGG -3'

Reverse primer

5'- CAGGGTCAAGGAAGGCACG -3'

Appendix C: protein extinction coefficients and molecular weights

Protein	Reduced		Oxidised	
	Molecular weight (Da)	Extinction coefficient at 280 nm (M^{-1} cm^{-1})	Molecular weight (Da)	Extinction coefficient at 280 nm (M^{-1} cm^{-1})
His-FUN-EGF3 WT	15957.93	11460	15935.93	12835
FUN-EGF3 WT	N/A	11460	14081.94	12835
FUN-EGF3 WT (^{15}N)	N/A	11460	14268.68	12835
FUN-EGF3 WT ($^{13}C/^{15}N$)	N/A	11460	14834.56	12835
His-FUN-EGF3 N57D	15958.91	11460	15936.91	12835
FUN-EGF3 N57D	N/A	11460	14082.93	12835
His-FUN-EGF3 W71R	15927.90	5960	15905.90	7335
FUN-EGF3 W71R	N/A	5960	14051.92	7335
His-FUN-EGF3 P148S	15947.89	11460	15925.89	12835
FUN-EGF3 P148S	N/A	11460	14071.91	12835
FUN-EGF3 P148S (^{15}N)	N/A	11460	14258.65	12835
His-FUN-EGF3 R62A	15872.82	11460	15850.82	12835
FUN-EGF3 R62A	N/A	11460	13996.84	12835
His-FUN-EGF3 S88A	15941.93	11460	15919.93	12835
FUN-EGF3 S88A	N/A	11460	14065.95	12835
His-FUN-EGF3 N98A	15914.90	11460	15892.90	12835
FUN-EGF3 N98A	N/A	11460	14038.92	12835
His-FUN-EGF3 D131A	15913.92	11460	15891.92	12835
FUN-EGF3 D131A	N/A	11460	14037.93	12835
His-FUN-EGF3 R159A	15872.82	11460	15850.82	12835
FUN-EGF3 R159A	N/A	11460	13996.84	12835
His-FUN-EGF1	8987.07	8480	8977.07	9105
FUN-EGF1	N/A	8480	7123.09	9105
FUN-EGF1 (^{15}N)	N/A	8480	7217.45	9105
His-EGF2-EGF3-BirA	11095.41	8480	11083.41	9230
EGF2-EGF3-BirA	N/A	8480	9229.43	9230
EGF2-EGF3-BirA (^{15}N)	N/A	8480	9353.59	9230
His-EGF2-EGF3	9159.17	2980	9147.17	3730
His-cbEGF32-34-BirA	17339.40	12950	17321.40	14075
cbEGF32-34-BirA	N/A	12950	15467.42	14075
cbEGF32-34-BirA-bio	N/A	12950	15693.73	14075
His-cbEGF41-43-BirA	17176.87	22920	17158.87	24045
cbEGF41-43-BirA	N/A	22920	15304.89	24045
cbEGF41-43-BirA-bio	N/A	22920	15531.20	24045
cbEGF41-43-BirA (^{15}N)	N/A	22920	15497.59	24045
His-cbEGF41-43-BirA Y2639N	17127.80	21430	17109.80	22555
cbEGF41-43-BirA Y2639N	N/A	21430	15273.82	22555
cbEGF41-43-BirA-bio Y2639N	N/A	21430	15482.13	22555

Appendix D: purification, refolding and characterisation of negative control fragment cbEGF32-34-BirA

The protocol was very similar to the method for purification of cbEGF41-43-BirA with several exceptions. The HPLC peak of reduced material was similarly broad (Figure A1A). Unlike cbEGF41-43-BirA, the protein was refolded in buffer at pH 8.3 and without glycerol. The HPLC profile of refolded material also showed a significant dependence on $CaCl_2$ − the EDTA-refolded material eluted as a much broader peak (Figure A1B). Thereafter, the His_6-tag was removed using factor Xa, the protein was purified further by FPLC (Figure A1C) and HPLC (Figure A1D) and biotinylated as for cbEGF41-43-BirA.

The final products were analysed by SDS-PAGE and mass spectrometry. The SDS-PAGE gel showed that the biotinylated and non-biotinylated proteins were both homogeneous under non-reducing and reducing sample conditions (Figure A2). Mass spectrometry confirmed the expected molecular weights (Table A4).

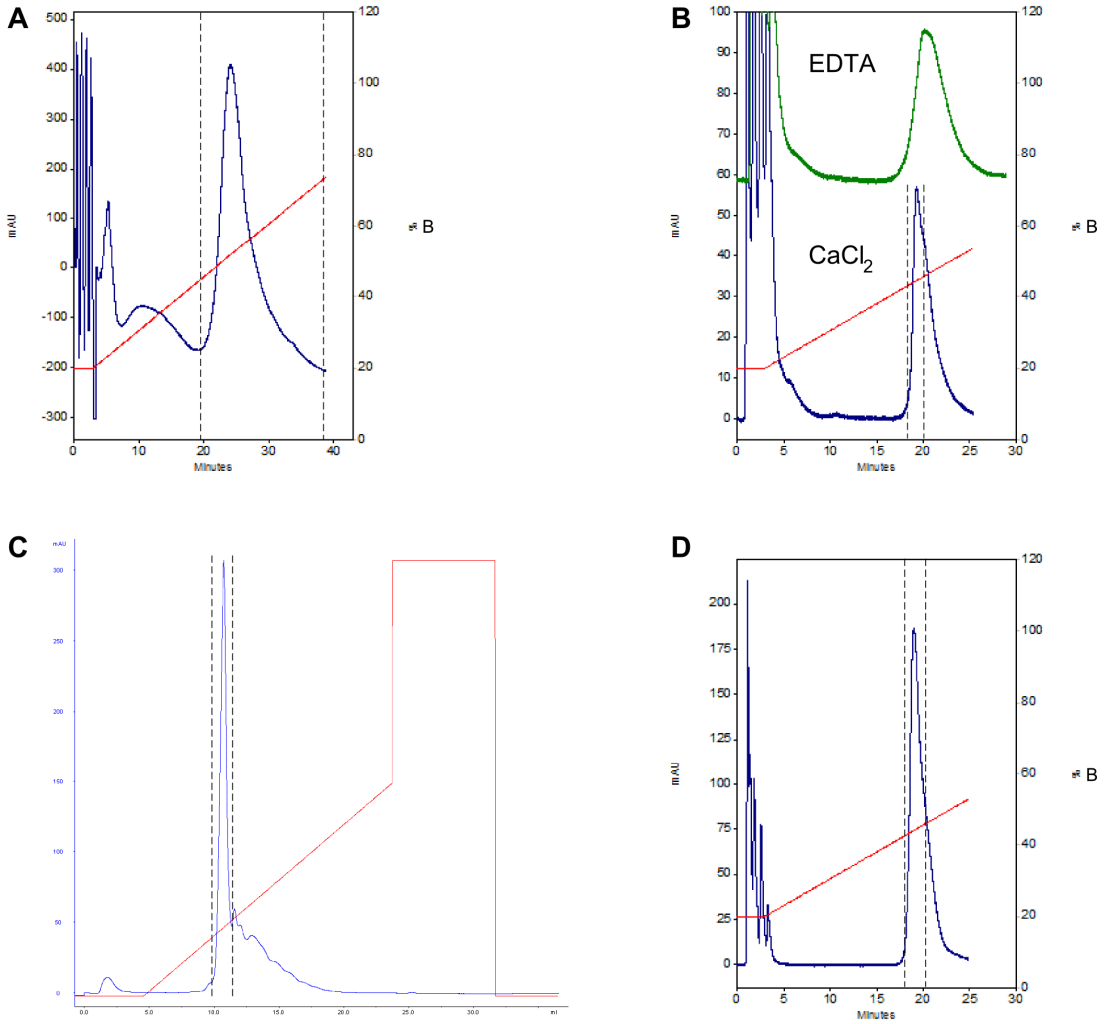


Figure A1 HPLC and FPLC chromatograms of cbEGF32-34-BirA. In this case, Buffer B for the MonoQ FPLC purification was 50 mM Tris pH 7.5 1 mM $CaCl_2$, 1M NaCl. A) HPLC chromatogram of reduced material. B) HPLC chromatogram of protein refolded in the presence of 50 mM $CaCl_2$ (blue trace) or 10 mM EDTA (green trace). C) FPLC chromatogram of Factor Xa-treated cbEGF32-34-BirA. D) Final HPLC chromatogram after FPLC purification. Dashed lines indicate collection boundaries for fractions.

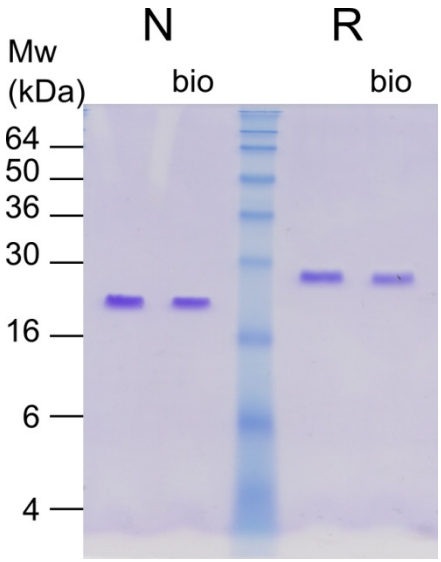


Figure A2 16% SDS-PAGE gel of cbEGF32-34-BirA and cbEGF32-34-BirA-bio. N and R indicated non-reducing and sample conditions. The biotinylated protein is marked 'bio'. Mw = molecular weight.

Table A4 Mass spectrometry results for cbEGF32-34-BirA and cbEGF32-34-BirA-bio.

Protein	**Expected molecular weight (Da)**	**Measured molecular weight (Da)**
cbEGF32-34-BirA	15467	15466
cbEGF32-34-BirA-bio	15994	15698

Appendix E: ^{1}H, ^{15}N and ^{13}C resonance assignments for FUN-EGF3

Assignments are given for every native residue in FUN-EGF3, starting at R45. Spectra were acquired in 5% D_2O/95% H_2O (v/v) or 100% D_2O at pH 5.4.

Residue	H^N	N	N^δ	N^ε	H^α	H^β	H^γ	H^δ	H^ε	H^ζ	H^η	C'	C^α	C^β	C^γ	C^δ	C^ε	C^ζ	C^η
R45	8.45	120.80	-	-	4.32	1.75,1.84	1.62,1.63	3.17,3.17	-	-	-	176.94	56.38	30.79	27.15	43.19	-	-	-
G46	8.43	110.42	-	-	3.98,3.98	-	-	-	-	-	-	174.67	45.34	-	-	-	-	-	-
G47	8.39	108.76	-	-	3.98,3.98	-	-	-	-	-	-	174.81	45.13	-	-	-	-	-	-
G48	8.33	108.55	-	-	3.91,3.91	-	-	-	-	-	-	174.82	45.14	-	-	-	-	-	-
G49	8.34	108.80	-	-	3.97,3.97	-	-	-	-	-	-	174.25	45.20	-	-	-	-	-	-
H50	8.35	118.25	-	-	4.69	3.25,3.13	-	-	-	-	-	174.35	55.54	29.52	-	-	-	-	-
D51	8.43	121.29	-	-	4.59	2.55,2.65	-	-	-	-	-	175.62	54.35	41.31	-	-	-	-	-
A52	8.14	124.17	-	-	4.28	1.34	-	-	-	-	-	177.42	52.41	19.51	-	-	-	-	-
L53	7.88	120.18	-	-	4.56	1.44,1.23	1.59	0.71,0.71	-	-	-	176.61	54.86	43.63	26.95	22.93,25.26	-	-	-
K54	8.24	120.05	-	-	4.54	1.61,1.74	1.23,1.30	1.57,1.58	2.88,2.88	-	-	175.69	54.78	35.74	23.80	29.21	42.19	-	-
G55	8.33	108.81	-	-	4.38,3.64	-	-	-	-	-	-	171.55	44.25	-	-	-	-	-	-
P56	-	-	-	-	4.27	1.49,2.00	1.55,1.78	3.37,3.59	-	-	-	176.32	62.49	33.02	26.56	49.92	-	-	-
N57	8.47	111.35	113.52	-	3.76	2.87,1.74	-	5.05,6.09	-	-	-	173.07	53.82	36.12	-	-	-	-	-
V58	7.57	117.48	-	-	4.05	1.70	0.47,0.69	-	-	-	-	175.62	63.50	32.34	22.72,22.69	-	-	-	-
C59	9.19	130.73	-	-	4.87	2.33,2.73	-	-	-	-	-	172.84	59.54	41.44	-	-	-	-	-
G60	8.24	106.32	-	-	3.67,4.38	-	-	-	-	-	-	173.95	44.76	-	-	-	-	-	-
S61	8.10	116.48	-	-	4.62	3.86,4.08	-	-	-	-	-	175.03	56.96	65.82	-	-	-	-	-
R62	8.51	119.22	-	-	3.87	1.65,1.41	1.08,1.08	2.95,2.95	-	-	-	177.10	58.71	29.70	26.29	43.25	-	-	-
Y63	7.83	115.57	-	-	4.59	2.78,3.26	-	7.12	6.83	-	-	175.30	57.83	38.15	-	133.00	118.14	-	-
N64	7.60	120.06	111.66	-	4.85	2.68,2.82	-	6.89,7.55	-	-	-	172.62	52.97	38.93	-	-	-	-	-
A65	8.19	126.47	-	-	4.81	1.27	-	-	-	-	-	177.02	51.80	20.71	-	-	-	-	-
Y66	7.80	117.86	-	-	4.70	2.97,3.11	-	6.99	6.70	-	-	174.16	55.53	39.50	-	134.82	117.27	-	-
C67	9.07	118.25	-	-	5.22	3.13,2.77	-	-	-	-	-	175.71	53.02	35.63	-	-	-	-	-
C68	9.06	121.79	-	-	4.65	3.46,2.35	-	-	-	-	-	170.37	53.74	39.70	-	-	-	-	-
P69	-	-	-	-	4.38	1.90,2.33	2.25,2.18	3.85,4.00	-	-	-	177.25	64.65	31.80	28.11	50.46	-	-	-
G70	8.72	112.67	-	-	3.43,4.09	-	-	-	-	-	-	171.29	44.69	-	-	-	-	-	-
W71	7.97	118.38	-	128.84	5.38	2.90,3.10	-	6.70 (Hδ1)	9.85 (Hε1) 6.97 (Hε3)	7.35 (Hζ2) 6.73 (Hζ3)	7.20 (Hη2)	175.14	55.31	32.95	-	126.93	118.11 (Cε3)	114.29 (Cζ2)	123.45 (Cη2)
K72	9.58	118.23	-	-	5.00	2.11,1.86	1.32,1.25	1.54,1.63	2.80,2.98	-	-	174.83	55.26	36.21	23.44	29.98	42.15	-	-
T73	7.86	107.72	-	-	5.29	4.43	1.00	-	-	-	-	177.96	60.26	71.62	20.35	-	-	-	-
L74	8.76	121.65	-	-	4.57	1.54,1.55	1.71	1.00,1.00	-	-	-	176.68	54.81	41.18	27.49	25.49,25.55	-	-	-
P75	-	-	-	-	4.44	2.29,1.95	2.20,2.02	3.91,3.60	-	-	-	178.60	64.59	30.86	28.15	50.27	-	-	-
G76	9.03	115.64	-	-	4.09,3.80	-	-	-	-	-	-	174.65	45.92	-	-	-	-	-	-
G77	8.46	108.74	-	-	4.41,3.90	-	-	-	-	-	-	173.63	44.25	-	-	-	-	-	-
N78	8.23	114.57	110.20	-	4.91	3.13,2.70	-	7.47,6.70	-	-	-	173.59	52.05	39.01	-	-	-	-	-
Q79	7.57	113.61	-	111.51	4.99	1.82,2.11	2.28,2.38	-	6.77,7.37	-	-	177.18	54.48	33.73	33.75	-	-	-	-
C80	8.44	120.78	-	-	4.41	2.62,2.99	-	-	-	-	-	173.83	55.10	38.50	-	-	-	-	-
I81	7.39	110.78	-	-	4.53	2.24	1.11,1.28 (Hγ1) 0.75 (Hγ2*)	0.91	-	-	-	176.21	60.52	40.10	26.22 (Cγ1) 18.58 (Cγ2)	14.08	-	-	-
V82	9.07	126.76	-	-	4.69	2.04	0.91,0.90	-	-	-	-	173.11	59.83	32.28	20.73,20.77	-	-	-	-
P83	-	-	-	-	4.16	0.94,0.63	1.51,1.51	4.02,3.91	-	-	-	175.40	61.52	30.35	26.38	50.35	-	-	-
I84	7.81	118.68	-	-	3.91	1.25	1.26,0.63 (Hγ1) 0.63 (Hγ2*)	0.63	-	-	-	176.27	60.99	39.53	27.24 (Cγ1) 17.10 (Cγ2)	14.03	-	-	-
C85	8.79	127.95	-	-	4.94	3.03,3.04	-	-	-	-	-	173.88	59.17	39.10	-	-	-	-	-
R86	8.59	123.62	-	-	3.93	1.57,1.58	1.36,1.35	3.06,3.05	-	-	-	175.94	57.57	30.11	27.06	43.09	-	-	-

Residue	H^N	N	N^δ	N^ε	H^α	H^β	H^γ	H^δ	H^ε	H^ζ	H^η	C'	C^α	C^β	C^γ	C^δ	C^ε	C^ζ	C^η
H87	8.25	116.55	-	-	4.80	3.16,3.17	-	7.13	-	-	-	173.88	53.41	30.74	-	120.01	-	-	-
S88	8.38	112.98	-	-	4.25	4.08,3.76	-	-	-	-	-	177.15	58.62	63.90	-	-	-	-	-
C89	8.86	124.41	-	-	4.90	2.38,2.91	-	-	-	-	-	175.50	53.51	36.66	-	-	-	-	-
G90	8.58	111.17	-	-	3.54,3.79	-	-	-	-	-	-	174.91	47.48	-	-	-	-	-	-
D91	8.84	126.59	-	-	4.68	2.86,2.58	-	-	-	-	-	177.26	54.94	40.65	-	-	-	-	-
G92	8.40	108.00	-	-	4.61,4.19	-	-	-	-	-	-	171.30	44.74	-	-	-	-	-	-
F93	7.95	113.62	-	-	4.95	3.04,2.98	-	7.08	7.31	7.35	-	172.96	55.17	41.79	-	133.29	131.06	129.64	-
C94	8.92	123.54	-	-	4.79	2.97,3.06	-	-	-	-	-	174.24	57.11	40.22	-	-	-	-	-
S95	9.14	124.91	-	-	4.79	3.50,3.89	-	-	-	-	-	174.30	58.44	65.40	-	-	-	-	-
R96	8.51	124.80	-	-	4.05	1.99,1.48	1.72,1.72	3.23,3.22	-	-	-	171.18	54.47	31.75	28.12	43.57	-	-	-
P97	-	-	-	-	3.95	2.00,1.70	1.70,0.76	3.32,2.77	-	-	-	176.88	65.06	31.05	27.48	49.99	-	-	-
N98	8.73	120.18	113.54	-	4.44	2.83,3.12	-	7.57,6.69	-	-	-	173.05	55.10	38.53	-	-	-	-	-
M99	8.40	119.72	-	-	4.69	1.50,1.50	2.27,2.36	-	1.96	-	-	172.82	54.69	35.90	32.59	-	17.26	-	-
C100	8.55	121.40	-	-	5.20	2.56,2.66	-	-	-	-	-	174.38	52.76	41.73	-	-	-	-	-
T101	8.84	117.32	-	-	4.42	2.76	0.97	-	-	-	-	175.10	62.69	68.64	21.40	-	-	-	-
C102	8.83	126.90	-	-	5.16	2.58,3.54	-	-	-	-	-	175.31	52.40	35.89	-	-	-	-	-
P103	-	-	-	-	4.38	2.43,1.94	2.14,1.93	3.85,3.53	-	-	-	175.49	65.83	31.56	28.26	50.47	-	-	-
S104	7.73	109.87	-	-	4.27	3.93,4.08	-	-	-	-	-	176.81	58.73	63.84	-	-	-	-	-
G105	8.57	111.39	-	-	3.55,4.43	-	-	-	-	-	-	173.89	44.70	-	-	-	-	-	-
Q106	7.51	120.03	-	112.80	4.19	1.74,1.85	2.22,2.29	-	7.55,6.88	-	-	173.85	55.81	29.90	33.85	-	-	-	-
I107	8.15	123.93	-	-	5.15	1.67	1.18,1.50 (Hγ1) 0.72 (Hγ2*)	0.78	-	-	-	176.32	59.81	38.28	27.49 (Cγ1) 17.57 (Cγ2)	13.24	-	-	-
A108	9.16	131.26	-	-	4.81	1.45	-	-	-	-	-	174.98	50.97	21.46	-	-	-	-	-
P109	-	-	-	-	4.55	2.46,1.88	2.07,2.03	3.73,3.75	-	-	-	174.81	64.76	32.25	27.84	50.76	-	-	-
S110	7.07	105.72	-	-	4.53	3.86,3.60	-	-	-	-	-	173.02	56.68	65.02	-	-	-	-	-
C111	8.54	117.52	-	-	5.04	2.90,3.35	-	-	-	-	-	175.32	53.08	39.59	-	-	-	-	-
G112	8.33	110.38	-	-	3.99,3.99	-	-	-	-	-	-	173.74	45.05	-	-	-	-	-	-
S113	8.33	115.69	-	-	4.48	3.80,3.80	-	-	-	-	-	174.68	58.06	64.26	-	-	-	-	-
R114	8.64	123.67	-	-	4.32	1.75,1.85	1.63,1.63	3.17,3.17	-	-	-	176.44	56.39	30.82	27.12	43.41	-	-	-
S115	8.57	118.01	-	-	4.43	3.82,3.82	-	-	-	-	-	174.46	58.42	63.81	-	-	-	-	-
I116	8.17	122.72	-	-	4.15	1.82	1.14,1.40 (Hγ1) 0.81 (Hγ2*)	0.82	-	-	-	176.02	61.17	38.77	27.24 (Cγ1) 17.50 (Cγ2)	12.98	-	-	-
Q117	8.37	124.25	-	112.63	4.31	1.92,2.04	2.32,2.32	-	7.48,6.85	-	-	175.59	55.63	29.62	33.83	-	-	-	-
H118	8.56	120.70	-	-	4.70	3.22,3.12	-	-	-	-	-	173.84	55.57	29.43	-	-	-	-	-
C119	-	-	-	-	4.69	3.33,3.08	-	-	-	-	-	174.52	56.45	41.69	-	-	-	-	-
N120	9.15	122.57	113.04	-	4.80	2.68,2.84	-	7.68,6.84	-	-	-	174.55	53.46	39.23	-	-	-	-	-
I121	7.23	118.54	-	-	4.17	1.59	1.38,1.00 (Hγ1) 0.72 (Hγ2*)	0.63	-	-	-	173.91	60.79	40.30	27.73 (Cγ1) 17.08 (Cγ2)	13.74	-	-	-
R122	8.40	124.61	-	-	4.29	1.66,1.66	1.58,1.43	3.10,3.10	-	-	-	175.72	55.60	30.30	27.14	43.34	-	-	-
C123	8.55	123.95	-	-	4.68	2.40,2.40	-	-	-	-	-	173.66	53.18	39.57	-	-	-	-	-
M124	8.34	121.25	-	-	4.39	1.98,1.80	2.50,2.49	-	1.99	-	-	176.37	55.15	34.89	32.64	-	17.21	-	-
N125	9.05	114.54	118.35	-	4.08	1.44,1.91	-	7.74,7.27	-	-	-	176.54	54.84	36.58	-	-	-	-	-
G126	8.75	107.05	-	-	3.69,4.00	-	-	-	-	-	-	175.46	45.66	-	-	-	-	-	-
G127	7.36	107.06	-	-	3.44,4.39	-	-	-	-	-	-	171.70	45.43	-	-	-	-	-	-
S128	8.40	114.63	-	-	4.70	3.78,3.79	-	-	-	-	-	172.43	57.07	65.38	-	-	-	-	-

Residue	H^N	N	N^δ	N^ε	H^α	H^β	H^γ	H^δ	H^ε	H^ζ	H^η	C'	C^α	C^β	C^γ	C^δ	C^ε	C^ζ	C^η
C129	8.89	123.74	-	-	4.86	3.10,3.16	-	-	-	-	-	174.44	56.25	41.95	-	-	-	-	-
S130	8.74	122.80	-	-	4.49	3.52,3.53	-	-	-	-	-	173.94	57.94	64.13	-	-	-	-	-
D131	9.07	126.04	-	-	4.22	2.91,2.67	-	-	-	-	-	174.73	56.66	39.64	-	-	-	-	-
D132	8.15	115.27	-	-	4.36	2.79,2.80	-	-	-	-	-	174.25	55.09	40.13	-	-	-	-	-
H133	7.84	114.61	-	-	4.87	3.24,3.24	-	7.21	-	-	-	172.61	54.52	31.02	-	120.30	-	-	-
C134	9.05	118.20	-	-	4.90	2.67,2.77	-	-	-	-	-	174.09	55.03	41.93	-	-	-	-	-
L135	9.27	126.90	-	-	4.63	1.66,1.65	1.55	0.89,0.89	-	-	-	176.76	54.07	41.20	27.22	24.32,24.87	-	-	-
C136	8.83	123.35	-	-	4.79	3.35,2.67	-	-	-	-	-	176.72	53.16	37.41	-	-	-	-	-
Q137	8.63	121.03	-	112.55	4.33	2.36,1.86	2.53,2.32	-	7.38,6.93	-	-	175.14	55.71	29.39	33.79	-	-	-	-
K138	8.39	121.21	-	-	4.04	1.79,1.81	1.40,1.48	1.68,1.69	2.99,2.99	-	-	177.17	58.46	32.24	24.64	29.29	42.24	-	-
G139	8.69	113.13	-	-	3.46,4.06	-	-	-	-	-	-	171.67	44.62	-	-	-	-	-	-
Y140	7.75	116.42	-	-	5.24	2.77,2.77	-	6.69	6.67	-	-	175.00	57.36	41.84	-	133.70	117.65	-	-
I141	9.15	116.47	-	-	4.75	1.95	1.25,0.98 (Hγ1) 0.77 (Hγ2*)	0.77	-	-	-	175.05	60.11	42.89	25.62 (Cγ1) 17.46 (Cγ2)	13.95	-	-	-
G142	7.88	106.14	-	-	3.86,4.93	-	-	-	-	-	-	177.37	43.61	-	-	-	-	-	-
T143	8.72	118.44	-	-	3.81	3.81	0.97	-	-	-	-	174.81	65.96	69.01	21.86	-	-	-	-
H144	7.98	-	-	-	5.31	2.90,3.55	-	7.25	-	-	-	174.99	54.13	28.86	-	120.07	-	-	-
C145	7.88	113.12	-	-	4.05	3.13,3.43	-	-	-	-	-	174.98	54.92	36.36	-	-	-	-	-
G146	8.64	104.52	-	-	3.89,3.77	-	-	-	-	-	-	174.01	45.98	-	-	-	-	-	-
Q147	8.34	121.58	-	111.19	5.08	1.92,1.99	2.25,2.25	-	7.29,6.68	-	-	172.57	52.16	30.59	33.47	-	-	-	-
P148	-	-	-	-	4.70	1.29,1.53	1.64,1.63	3.91,3.90	-	-	-	175.81	61.64	31.45	27.01	50.66	-	-	-
V149	8.09	120.20	-	-	3.94	1.64	0.73,0.74	-	-	-	-	176.07	62.23	33.45	20.90,20.93	-	-	-	-
C150	8.83	128.50	-	-	5.09	3.01,2.75	-	-	-	-	-	175.24	58.96	37.18	-	-	-	-	-
E151	8.97	125.39	-	-	3.82	2.00,2.00	2.36,2.22	-	-	-	-	177.65	59.85	29.81	36.96	-	-	-	-
S152	8.65	113.76	-	-	4.48	3.89,3.88	-	-	-	-	-	174.60	58.45	63.39	-	-	-	-	-
G153	7.67	108.65	-	-	3.71,4.04	-	-	-	-	-	-	172.34	44.64	-	-	-	-	-	-
C154	8.16	116.30	-	-	4.73	2.47,2.46	-	-	-	-	-	174.21	53.15	38.94	-	-	-	-	-
L155	8.38	123.20	-	-	4.29	1.37,1.78	1.65	0.80,0.79	-	-	-	177.66	54.93	43.08	27.28	22.36,25.17	-	-	-
N156	8.83	114.23	118.45	-	4.09	2.16,2.17	-	7.98,7.03	-	-	-	175.43	58.24	37.05	-	-	-	-	-
G157	8.48	105.39	-	-	4.07,3.64	-	-	-	-	-	-	175.40	45.36	-	-	-	-	-	-
G158	7.63	107.33	-	-	4.30,3.66	-	-	-	-	-	-	171.61	45.33	-	-	-	-	-	-
R159	8.25	117.74	-	-	4.63	1.69,1.69	1.54,1.54	3.16,3.17	-	-	-	174.57	54.25	33.63	27.17	43.34	-	-	-
C160	9.04	125.98	-	-	4.86	2.97,2.97	-	-	-	-	-	175.25	56.62	39.13	-	-	-	-	-
V161	8.62	122.24	-	-	4.53	2.33	0.93,0.78	-	-	-	-	175.07	62.00	33.47	22.15,18.10	-	-	-	-
A162	8.68	126.73	-	-	4.31	1.41	-	-	-	-	-	171.98	50.86	18.49	-	-	-	-	-
P163	-	-	-	-	4.24	1.82,2.15	1.23,1.91	3.49,3.45	-	-	-	176.92	64.89	31.16	28.23	50.00	-	-	-
N164	9.08	121.38	113.12	-	4.47	2.83,3.16	-	7.48,7.09	-	-	-	173.18	54.85	37.88	-	-	-	-	-
R165	8.37	121.34	-	-	4.60	1.41,1.55	1.48,1.41	3.14,3.13	-	-	-	173.65	55.95	32.78	27.06	43.20	-	-	-
C166	8.77	124.15	-	-	5.30	2.77,2.62	-	-	-	-	-	174.80	54.14	43.03	-	-	-	-	-
A167	9.57	126.46	-	-	4.72	1.40	-	-	-	-	-	177.31	51.43	20.38	-	-	-	-	-
C168	8.77	121.99	-	-	4.91	2.73,3.36	-	-	-	-	-	175.78	53.46	39.15	-	-	-	-	-
T169	8.60	115.22	-	-	4.35	4.35	1.18	-	-	-	-	174.06	62.08	69.83	22.15	-	-	-	-
Y170	8.24	119.53	-	-	4.23	2.94,3.10	-	7.10	6.82	-	-	176.47	60.01	38.11	-	132.79	-	-	-
G171	8.12	112.61	-	-	3.24,3.80	-	-	-	-	-	-	173.54	44.55	-	-	-	-	-	-
G172	7.64	120.36	-	-	4.72	2.92,2.92	-	7.20	7.26	7.21	-	175.61	58.05	40.22	-	132.77	131.16	129.49	-

Residue	H^N	N	N^δ	N^ε	H^α	H^β	H^γ	H^δ	H^ε	H^ζ	H^η	C'	C^α	C^β	C^γ	C^δ	C^ε	C^ζ	C^η
T173	-	-	-	-	4.77	4.18	1.06	-	-	-	-	174.63	61.14	71.34	20.84	-	-	-	-
G174	8.30	109.55	-	-	4.37,4.06	-	-	-	-	-	-	172.86	45.07	-	-	-	-	-	-
P175	-	-	-	-	4.39	2.00,2.35	2.10,2.08	3.62,3.82	-	-	-	177.52	64.93	31.95	27.44	49.72	-	-	-
Q176	8.69	115.28	-	111.95	4.68	2.22,2.00	2.21,2.33	-	6.74,7.46	-	-	175.02	55.40	29.96	34.90	-	-	-	-
C177	7.85	-	-	-	4.45	-	-	-	-	-	-	172.80	54.07	-	-	-	-	-	-
E178	8.00	124.78	-	-	4.07	1.88,2.04	2.20,2.19	-	-	-	-	181.04	58.38	31.60	37.04	-	-	-	-

Appendix F: fibrillin protein sequences used in alignments

Protein sequences were obtained from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) and Ensembl databases (http://www.ensembl.org/index.html).

Organism	Accession number	Database
Homo sapiens (human)	NP_000129.3 (fibrillin-1) NP_001990.2 (fibrillin-2) NP_115823.3 (fibrillin-3) NP_996826.2 (LTBP-1L)	NCBI NCBI NCBI NCBI
Canis familiaris (dog)	XP_535468.2 (fibrillin-1)	NCBI
Mus musculus (mouse)	NP_032019.2 (fibrillin-1)	NCBI
Gallus gallus (chicken)	XP_413815.3 (fibrillin-1)	NCBI
Xenopus tropicalis (clawed frog)	XP_002936615.1	NCBI
Danio rerio (zebrafish)	NP_001129262.1 (fibrillin-2)	NCBI
Petromyzon marinus (lamprey)	ENSPMAP00000002691	Ensembl
Ciona intestinalis (sea squirt)	XP_002119988.1	NCBI
Tribolium castaneum (red flour beetle)	XP_974344.2	NCBI
Nematostella vectensis (starlet sea anenome)	A7SBY7_NEMVE (N-terminal fragment)	Ensembl