

An Investigation Into The Molecular Mechanism Of The Fibrillin1 - LTBP1 Interaction

Ian Robertson

Department of Biochemistry and
Somerville College

University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2012

An Investigation Into The Molecular Mechanism Of The Fibrillin1 - LTBP1 Interaction

Ian Robertson - Department of Biochemistry and Somerville College

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

Abstract

Many studies have demonstrated a connection between the fibrillin matrix and TGF β signalling, but at present the mechanistic basis for this link is unclear. An interaction between the C-terminus of Latent TGF β Binding Protein 1 (LTBP1) and the N-terminus of fibrillin1 has previously been identified, and may have the potential to directly link the fibrillin matrix to TGF β signalling.

To investigate the structural basis for this interaction, several multi-domain fragments of fibrillin1 and LTBP1 were expressed prokaryotically and refolded *in vitro*. After initial characterisation to confirm folding, the structure, dynamics, and interdomain interactions of these fragments were investigated in more detail using NMR techniques. Domains in both LTBP1 and fibrillin1 appear to demonstrate folds consistent with homologous structures, and while the LTBP1 C-terminal cbEGF14-TB3-EGF3-cbEGF15 region contains many flexible linkers and few interdomain interactions, the fibrillin1 EGF2-EGF3-hyb1-cbEGF1 region appears rigid, with interfaces forming between all domains present.

SPR studies were used to demonstrate binding between distinct LTBP1 and fibrillin fragments, suggesting interactions between multiple domains are involved in the LTBP1-fibrillin1 interaction. The binding sites involved were then mapped to specific residues using HSQC titration studies, and structural models for the LTBP1-fibrillin1 interaction were generated based on these data. Predictions from these models were used to target residues for site-directed mutagenesis, based on their potential involvement in salt bridges, and when certain residues were replaced with those of opposite charge, reductions in binding could be seen in the SPR assay. These key residues were consistent with a particular model of the LTBP1-fibrillin1 interaction, as derived from the HSQC titration data. The conservation of potential binding site residues through deuterostome evolution also supports an important biological role for the LTBP-fibrillin interaction.

Table of Contents

Title Page	i
Abstract	ii
Table of Contents	iii
Acknowledgements	vii
Abbreviations	ix
Chapter 1 – Introduction	1
1.1 Overview	1
1.2 The structure of the fibrillins and LTBP.....	4
1.2.1 Multi-domain organisation	4
1.2.2 TB Domain motif	6
1.2.3 Hybrid domain motif	8
1.2.4 cbEGF domain motifs	10
1.3 Biological role of the fibrillins	11
1.3.1 Microfibrils	11
1.3.2 Marfan syndrome	19
1.3.3 Mouse models of MFS	25
1.3.4 Fibrillin and growth factor signalling	27
1.3.5 Other genetic diseases affecting the fibrillins	30
1.3.6 Conclusions on the biological roles of fibrillins	33
1.4 TGF β signalling and LTBP.....	34
1.4.1 The TGF β signalling pathway	34
1.4.2 Interaction of LTBP.....	38
1.4.3 Interaction of LTBP.....	41
1.4.4 Interaction of LTBP.....	44
1.4.5 Role of LTBP.....	45
1.5 In vivo role of LTBP.....	46
1.5.1 LTBP.....	48
1.5.2 LTBP.....	51
1.5.3 The in vivo role of LTBP.....	54
1.6 Role of fibrillin-LTBP interaction in TGF β signalling	57
1.7 Thesis aims	59
Chapter 2 - Evolution of fibrillins and LTBP.....	69
2.1 Introduction	69

2.2 Evolutionary history of the TB domain protein family	69
2.3 Phylogenetic analysis of fibrillins and LTBP1s.....	78
2.4 Specific insights into the LTBP-Fibrillin interaction	85
2.6 Conclusions	93
Chapter 3 - Structure and dynamics of the fibrillin N-terminus	96
3.1 Introduction	96
3.2 Production of recombinant fibrillin	96
3.3 Initial characterisation and assessment of fold	98
3.4 Backbone and side chain resonance assignment	103
3.5 Fast timescale Dynamics	109
3.6 Calcium binding	109
3.7 Backbone solvent accessibility	113
3.8 Structure determination of FBN ^{e2cb1} – General strategy	115
3.9 Chemical shift information	117
3.10 NOE assignment of FBN ^{e2cb1}	119
3.11 Residual Dipolar Couplings	123
3.12 Structure calculations	127
3.13 Structural features of FBN ^{e2cb1}	132
3.14 Conclusions	139
Chapter 4 - Structural studies of the LTBP1 C-terminus	143
4.1 Introduction	143
4.2 Production of recombinant LTBP1 proteins	143
4.3 Initial assessment of fold	148
4.4 Backbone resonance assignment	151
4.5 Modelling the LTBP1 C-terminus	155
4.6 Calcium-binding	157
4.7 Solvent accessibility and hydrogen bonding	162
4.8 Fast time scale dynamics	165
4.9 Conclusions	171
Chapter 5 - Studies of the interactions between the Fibrillin1 N-terminus and LTBP1 C-terminus	173
5.1 Introduction	173
5.2 LTBP1 and fibrillin1 protein fragments	173
5.3 Pull down studies	175
5.4 Surface plasmon resonance studies	177
5.5 NMR titrations to define binding sites	184
5.6 Docking simulations between Fibrillin1 and LTBP1 fragments	196

5.7 Conclusions	201
Chapter 6 - Disruption of the LTBP1 - Fibrillin1 interaction by site directed mutagenesis	206
6.1 Introduction	206
6.2 Production and characterisation of Fibrillin1 ^{e2cb1} mutants	208
6.3 Production and characterisation of LTBP1 ^{TB3E3} mutants	212
6.4 Characterisation of mutant interactions by SPR	214
6.4.1 Substitutions disrupting the LTBP1 EGF3 - fibrillin1 hyb1 interaction	221
6.4.2 Substitutions disrupting the LTBP1 TB3 – fibrillin1 interaction	222
6.5 Interaction of fibrillin mutants with overlapping LTBP1 fragments	224
6.6 Effect of R232E substitution in NMR titration experiments.	224
6.7 Conclusions	227
Chapter 7 – Discussion	230
7.1 Summary	230
7.2 Structure of the fibrillin N-terminus	230
7.2.1 Domain interfaces	230
7.2.2 Position of cysteine 204 in hyb1	234
7.2.3 Context of FBN ^{e2cb1} within the fibrillin N-terminus	235
7.3 Structure of the LTBP1 C-terminus	235
7.4 Structure and interfaces of the fibrillin1- LTBP1 complex	237
7.4.1 Multi-domain binding	237
7.4.2 Confirmation of interfaces by mutagenesis	239
7.4.3 Conserved interface residues	239
7.4.4 Context of the fibrillin1-LTBP1 interaction within the matrix	246
7.5 Biological implications of the fibrillin1- LTBP1 interaction	248
7.6 Future directions	252
7.7 Concluding remarks	255
Chapter 8 - Materials and Methods	258
Bioinformatics and other In Silico techniques	
8.1 Sequence discovery	258
8.2 Domain analyses	258
8.3 Alignments	259
8.4 Phylogenetic trees	260
8.5 Summary tree	260
8.6 Homology modelling	260
Molecular Biology	
8.7 Cloning	260

8.8 Site-directed mutagenesis	264
Protein production	
8.9 Protein expression- rich media	265
8.10 Protein expression- minimal media	267
8.11 Guanidine-HCl – Nickel column purification	268
8.12 Inclusion body purification	269
8.13 High performance liquid chromatography (HPLC)	270
8.14 Re-dissolving proteins precipitated in dialysis	271
8.15 Protein refolding	271
8.16 Factor Xa cleavage	272
8.17 Fast protein liquid chromatography (FPLC)	273
8.18 Biotinylation	273
Biochemical and biophysical assays	
8.19 SDS-PAGE	274
8.20 Mass-Spectrometry	275
8.21 UV-vis spectroscopy - protein concentration determination	275
8.22 Calcium Chelation	276
8.23 Fluorimetry	276
8.24 Pull down experiments	278
8.25 Western blotting	278
8.26 Surface plasmon resonance (SPR)	280
NMR experiments	
8.27 General experiment information	283
8.28 NMR – 2D experiments	288
8.29 NMR – 3D spectra and assignment	289
8.30 NMR - Residual dipolar couplings	290
8.31 NMR – Structure determination	291
8.32 NMR – D ₂ O exchange	293
8.33 NMR – Heteronuclear NOE	293
8.34 NMR – HSQC protein titrations	294
8.35 NMR- HADDOCK	295
Appendix 1 – NMR assignments and other supplementary NMR data ..	298
Appendix 2 – Representing protein evolution	314
Appendix 3 – Extensive tables of relevant mouse models	319
Appendix 4 – Sequences used in evolutionary analyses	329
Appendix 5 – Parameters used to generate phylogenetic trees	331

Funding

I would like to thank the BBSRC for funding my PhD studentship, and the Wellcome Trust for project funding used to support the work carried out in this Thesis. I would also like to thank Somerville College for the scholarship and travel grant support I have received.

Acknowledgements

I owe a great deal of thanks to Professor Penny Handford, who has been a great supervisor, and invaluable in providing advice, resources, and vital moderation of my frequently over-ambitious ideas.

I am also heavily indebted to Professor Christina Redfield for her frequent help and near unending patience with the NMR experiments, which were vital to the success of this project, and whose super-human time management abilities are always a source of inspiration.

Dr Sacha Jensen deserves many, many thanks, for his patient mentoring in all aspects of protein production and other lab techniques, not to mention generous proof reading of much of this document.

Isabelle Osuch was also a great help in getting this project up and running, and her perseverance with pull downs in her part II project displayed an admirable strength of character! Sarah Jackson also contributed to this project, and, although her data is not used directly in this thesis, she helped produce a number of proteins used in the work presented here. She also helped establish a side project on the LTBP2 and 4 isoforms that is not discussed here. Sophie Gilbert also made a great contribution to furthering this LTBP2/4 project, and cloned the isolated LTBP1 linker region mentioned in Chapter 5.

Thanks are also owed to David Yadin, for frequent help and advice with NMR and NMR structure determination, as well as providing useful data discussed in Chapter 3, and also just for being David.

Thanks also go to the rest of the Handford lab past and present, particularly Sarah Iqbal, who provided a number of control proteins used in Chapter 5, and also to Mouli, Pat, Paul, and Philip for generally keeping the lab running, as well as frequently providing help, advice and good conversation.

Dr David Staunton also helped with my training for SPR and other biophysics instruments, and Dr Ed Lowe helped with our sadly ill-fated attempts at protein crystallisation. Professor Chris Ponting also gave helpful comments on our review of TB protein evolution.

Finally I would like to thank my parents, my family, my friends, and past tutors, teachers, and mentors, who have all collectively got me to the place I am in today. Particular thanks go to the friends I've got to know over the last eight years in this little town; you've all made it a very good place to be.

Abbreviations

AD- Acromicric dysplasia
BLAST – Basic local alignment search tool
BMP- Bone morphogenic protein
cbEGF- Calcium binding epidermal growth factor
CCA- Congenital contractural arachnodactyly
E- Embryonic day
ECM- Extracellular matrix
EGF- Epidermal growth factor
ER- Endoplasmic reticulum
EST- Expressed sequence tag
FBN- Fibrillin
GD- Geleophysic dysplasia
GDF- Growth and differentiation factor
HGMD- Human Gene Mutation Database (www.hgmd.org)
LAP- Latency Associated Peptide (TGF β pro-peptide)
LDS- Loeys-Dietz syndrome
LLC- Large latent complex
LOX- Lysyl oxidase
LTBP- Latent transforming growth factor beta binding proteins
MAGP- Microfibril associated glycoprotein
MASS- Mitral valve prolapse, Aortic enlargement, Skin and Skeletal findings
MFS- Marfan syndrome
MMP- Matrix metalloprotease
NMD- Nonsense mediated decay
NMR- Nuclear magnetic resonance
NOE – Nuclear Overhauser effect
P- Postnatal day
pSMAD- Phosphorylated SMAD
RMSD – Root mean square deviation
ROS- Reactive oxygen species
R-SMAD –Receptor SMAD
SAXS – Small angle X-ray scattering
SLC- Small latent complex

SMAD- C elegans “SMA” and Drosophila “Mothers Against Decapentaplegic”
homologue
SSS- Stiff skin syndrome
TB- Transforming growth factor beta binding
TGase - Transglutaminase
TGF β - Transforming growth factor beta
TSP- Thrombospondin
UMD- Ultimate Mutation database (www.umd.be/FBN)
URDS- Urban-Rifkin-Davies syndrome
WMS- Weill-Marchesani syndrome
WT- Wild type
y2h – yeast 2 hybrid screen

Introduction

1.1 Overview

The fibrillins and latent TGF β binding proteins (LTBPs) comprise a distinctive group of large, multi-domain, extracellular matrix (ECM) proteins, which perform diverse and important roles in the extracellular matrix. The fibrillins and LTBPs can be clearly distinguished from other protein groups by the presence of a unique domain type, defined by a specific pattern of eight cysteine residues, and commonly referred to as “8-cysteine” or “TGF β Binding” (TB) domains. Both fibrillins and LTBPs contain several TB domains, often interspersed between arrays of calcium binding epidermal growth factor (cbEGF) domains, a more widely distributed module type. Humans possess three fibrillins and four LTBPs, and their domain structure is summarised in Figure 1.1.

Although fibrillins and LTBPs have some similarities in domain structure, they can be clearly distinguished, as LTBPs are shorter, and contain several large regions that lack any cysteine residues or other recognisable structural motifs. Fibrillins were the first member of the TB domain family to evolve, emerging before the divergence of cnidarians¹ and bilateria² [1] (examined further in chapter 2), while LTBPs emerged later with the appearance of deuterostomes³.

Figure 1.1 also shows some of the published alternatively spliced forms of the LTBPs, which have varying tissue expression patterns [2, 3]. No significant alternatively spliced forms have been reported for the fibrillins, and while alternatively spliced

¹ Simple organisms such as Jelly fish, corals, and anemones.

² Organisms with bilateral symmetry including flatworms, insects and humans

³ Superphylum of animals distinguished by the anus developing prior to the mouth, includes sea urchins, tunicates, lancelets and humans.

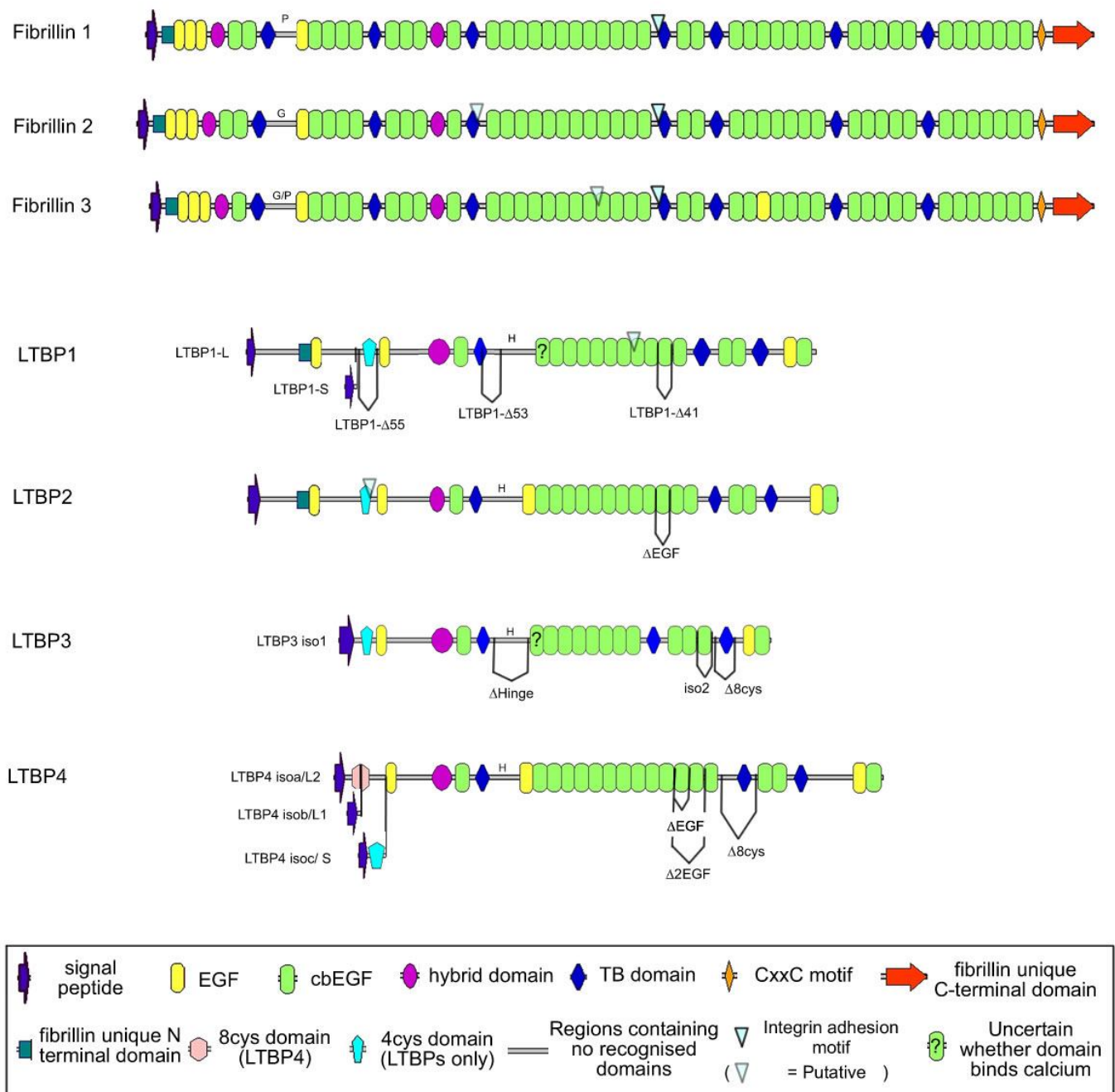


Figure 1.1- The domain organisation of the human TB domain proteins

All known human TB proteins are shown, with domain symbols summarised in the box at the top of the Figure. Note the high conservation of fibrillin domain organisation compared with LTBPs. Some of the alternatively spliced forms of the LTBPs in humans are also illustrated [2, 162]. The letter H denotes the "hinge" region in LTBPs, the letter P denotes the proline rich region in fibrillin1. The letter G denotes the glycine rich region in fibrillin2, and the letters G/P denotes the glycine and proline rich region of fibrillin3.

fibrillin mRNAs can be found in some databases, they mostly lack coding signal peptides, and so are unlikely to be physiologically relevant.

The most clearly established role of fibrillin is structural, being the principal component of beaded 10-12 nm microfibrils, which are structures found in both elastic and inelastic tissues [4, 5]. As well as this structural role, several *in vivo* studies have identified fibrillin as playing an important role in growth factor regulation, and in particular fibrillin seems to help control TGF β signalling [6]. Many questions remain to be answered as to the mechanism of TGF β control by fibrillin, but most hypotheses involve the sequestration of TGF β to fibrillin microfibrils via LTBP1 [6].

LTBP1, as their name suggests, were first identified as proteins that bind TGF β . LTBP1 maintain this growth factor in a latent state through interactions with the TGF β propeptide [7], and are important for both the proper secretion of TGF β from the cell [8] and its activation by integrins [9]. Additionally some LTBP1 may play TGF β independent roles; LTBP2 does not bind TGF β [10, 11], and LTBP4 has been implicated as playing a role in elastic fibre assembly [12]. Critically for this thesis, several LTBP1 have been demonstrated to interact with both fibrillin and TGF β (discussed in more detail later), and it is hypothesised that this interaction may play a key role in the regulation of this growth factor, potentially as part of a complex extracellular “sensor” involving numerous components [13-15].

The focus of this thesis is to determine the molecular basis for the LTBP1 – fibrillin1 interaction, using a range of techniques, including bioinformatics studies of fibrillin and LTBP1 evolution, and biochemical and biophysical assays of the structure, dynamics, and interactions of the relevant regions of fibrillin1 and LTBP1.

1.2 The structure of the fibrillins and LTBP

1.2.1 - Multi-domain organisation

Fibrillins are primarily composed of multiple cbEGF and TB domain modules (Figure 1.1). Fibrillins and LTBP_s also contain non-calcium binding EGF domains, a 4 cysteine N-terminal domain (referred to as the Fibrillin Unique N-terminal (FUN) domain), and “hybrid” domains, (named as such because they share features with both TB domains and cbEGF domains [16]). Fibrillins also have a unique C-terminal domain, which is not found in LTBP_s, and is cysteine-free except for a CXXC motif shortly after the last cbEGF domain. Fibrillins also have a cysteine-free variable region after TB1, often referred to as the proline rich region in fibrillin1, and the glycine rich region in fibrillin2. LTBP_s on the other hand have much larger variable regions that are cysteine-free, especially at the N-terminus, and the variable region after the first TB domain and before the central EGF array, found in all LTBP_s, is often referred to as the “hinge region”. Additionally LTBP_s can possess another 4-cysteine type module, exclusive to the LTBP_s, and for which no structure is currently known (blue pentagon Figure 1.1).

The extremely large size of fibrillins and LTBP, and their challenging solubility properties, make high-resolution structure determination of the full-length proteins effectively impossible. However a great deal of success has been obtained using a dissection approach to look at single or small groups of domains using X-ray crystallography and nuclear magnetic resonance (NMR) studies [16-21]. Fibrillin and LTBP structures determined in this way are summarised in Figure 1.2.

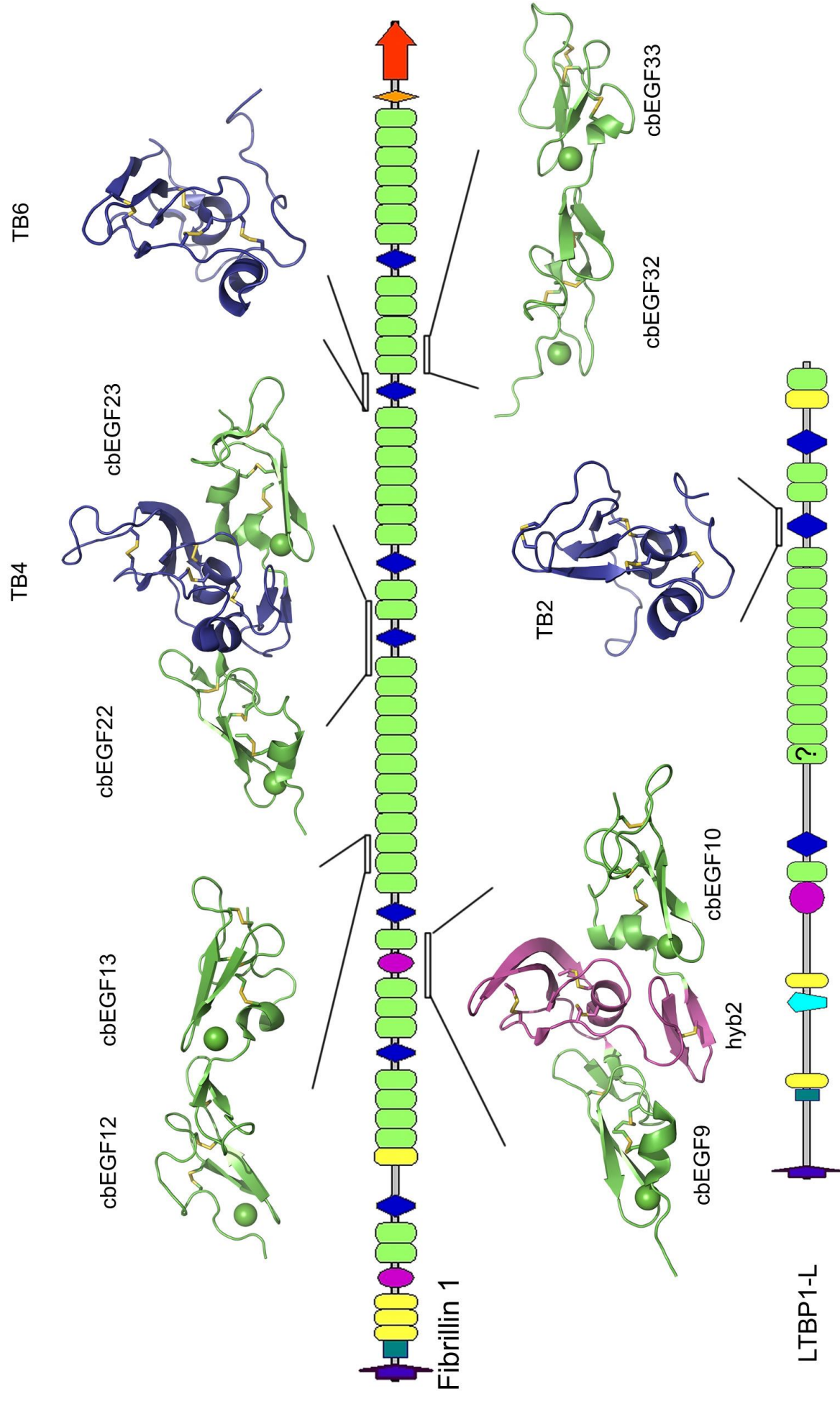


Figure 1.2- Known domain structures of fibrillin1 and LTBP1, Domains are coloured according to the same scheme as fig 1.1, calcium atoms are included as green spheres. References [16-21].

1.2.2 - TB Domain motif

The TB domain is defined by an 8-cysteine motif that forms four disulphide bonds, which have a 1-3, 2-6, 4-7, 5-8 arrangement. This stabilises the fold, along with a small hydrophobic core, focused around a conserved tryptophan residue. Several high-resolution structures of these domains have been determined, from fibrillin-1 and LTBP-1 [17-19], both in isolation, and covalently linked to cbEGF domains (Figure 1.2). In all cases they demonstrate a mostly globular fold, with a 4-stranded anti-parallel beta-sheet and two small alpha helices comprising most of the domain, and another beta hairpin at the C-terminus (Figure 1.3).

The TB4 domain of fibrillin was crystallised as a three-domain construct (cbEGF22-TB4-cbEGF23) with both flanking cbEGF domains present [17]. This structure demonstrates extensive hydrophobic and salt bridge interdomain interactions between both cbEGF22-TB4 and TB4-cbEGF23 domain pairs (Figure 1.4 A). It has also been demonstrated by NMR that the TB4-cbEGF23 packing interaction occurs in solution, and creates a particularly high affinity calcium-binding site [22]. These data, together with studies of calcium affinity in other TB-cbEGF domain pairs [22], support a structural model for fibrillin involving rigid interfaces between TB and cbEGF domains.

TB domains have also been shown to mediate specific protein-protein interactions, which may be important for the regulatory functions associated with the LTBP/fibrillin family. One such interaction is the binding of fibrillin TB4 to α_v and α_5 integrins at the cell surface, which is mediated primarily by an “RGD” sequence on the first beta-hairpin of this domain [17, 23-27]. This interaction appears to be relevant in fibroblast migration and the formation of focal adhesions, and the disruption of this domain in stiff skin syndrome leads to various phenotypic abnormalities (section 1.3.5) [28].

Another protein-protein interaction that TB domains have been implicated in is mediated by the second TB domain of LTBP. This TB domain has been shown to bind covalently, via disulphide linkage, to the pro-peptide of TGF β [29], which is involved in regulation of the TGF β growth factor (discussed in more detail in Section 1.4.2).

1.2.3 - Hybrid domain motif

The hybrid (hyb) type domain is closely related to the TB domain, and recently, the structure of the fibrillin1 hyb2 domain, along with both adjacent cbEGFs, has been solved by crystallography [16]. This has demonstrated that, as well as being similar in sequence to the TB domains, the hyb domains also adopt a similar fold. Although the hyb domains differ from TB domains in that their central core lacks one disulphide, and an additional disulphide is present in the C-terminal beta hairpin (Figure 1.3), (which is also reminiscent of the C-terminus of a cbEGF), giving a 1-3, 2-5, 4-6, 7-8 disulphide arrangement [16].

The hyb2 structure also demonstrates extensive interdomain hydrophobic interactions with the C-terminal cbEGF10, and many of the amino acids involved are highlighted in Figure 1.4 B. This interdomain interaction has also been confirmed in solution by NMR and shown to create a high affinity calcium binding site in cbEGF10 [22]. The N-terminal domain, cbEGF9, on the other hand shows relatively few side chain contacts with the hyb2 domain, and, although there is some limited interdomain packing in the crystal structure, it is not clear whether this interface is stable in solution.

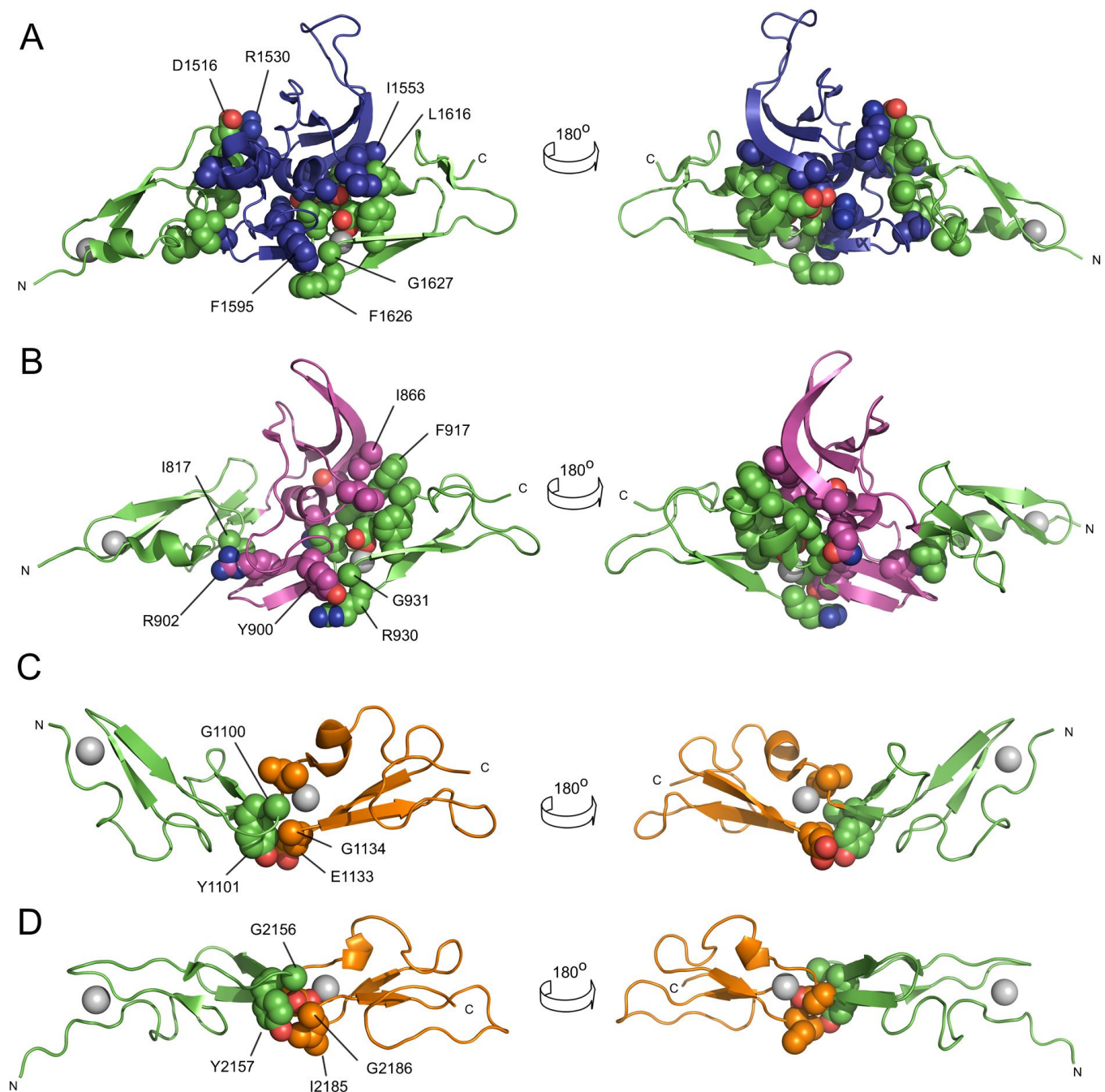


Figure 1.4- Interdomain interactions of TB, hybrid and cbEGF domains

Domain colouration: blue - TB, magenta - hybrid, cbEGF - green (in the cbEGF pairs the downstream cbEGF is coloured orange). Amino acid side chains at domain interfaces are highlighted as spheres, some of which are labelled. Calcium ions are shown as light grey spheres. **A)** The crystal structure of cbEGF22-TB4-cbEGF23 fragment (PDB 1UZJ), **B)** cbEGF9-Hyb2-cbEGF10 (PDB 2W86) **C)** cbEGF12-13 (PDB 1LMJ) **D)** cbEGF32-33 (PDB 1EMN).

No detailed protein binding sites within a hyb domain have yet been published, but the hyb1 domain of fibrillin has been implicated in a wide variety of interactions, especially binding to LTBP [15] (discussed later).

1.2.4 - cbEGF domain motifs

cbEGF domains are found in a wide variety of ECM proteins, and can often be found in multiple tandem repeats. These domains are relatively small, rarely much more than 40 amino acids long, and generally lack a clear hydrophobic core, making their folding reliant on the three disulphide bonds present. These bonds form between cysteines 1-3, 2-4, and 5-6 (Figure 1.3). Calcium binding EGF's also contain specific conserved residues important for binding calcium (coloured green in fig 1.3), with the following consensus sequence: (D/N)X(D/N)(E/Q)X_m(D/N)*X_n(Y/F) where *m* and *n* are variable, and * indicates a potential β -hydroxylation site [20]. If a domain lacks one or more of these residues it is generally predicted to be an EGF domain rather than a cbEGF.

cbEGF and EGF structures have been determined from various proteins, including the structures of cbEGF12-13 and cbEGF32-33 domain pairs from fibrillin [20, 21]. Interdomain interfaces are also seen between these domain pairs, resulting in calcium-stabilised rod-like structures in solution, (both structures were determined by NMR) [20, 21]. Packing residue side chains are highlighted in Figure 1.4 C and D, where the interdomain interaction is focused around a generally conserved G(Y/F) motif, found three amino acids after the 5th cysteine in the N-terminal cbEGF. This packs against the turn of the beta hairpin in the C-terminal cbEGF domain, where a glycine can often be found in an "XG" motif, (X may often be hydrophobic).

Some protein-protein interaction sites have been defined in detail for EGF and cbEGF domains, such as the interaction of EGF with its receptor [30]. However at present no detailed binding sites have been defined in fibrillin or LTBP EGF/cbEGF domains.

1.3 - Biological role of the fibrillins

The LTBP1 – fibrillin1 interaction was chosen as the subject of this thesis, as it may have important implications for the biological roles of both proteins. However before the role of this interaction is looked at in detail, the biological roles of fibrillin and LTBPs that have been firmly established are discussed.

1.3.1 - Microfibrils

Fibrillins were initially identified as the principal component of higher order structures, known as 10-12 nm microfibrils [4, 5], which have been shown by rotary shadowing electron microscopy to form beaded string-like structures, with an average periodicity of 55 nm [31, 32]. These microfibrils are found in a wide variety of tissues, are often associated with elastin, where they are thought to direct the assembly of elastic fibres [32]. These fibres are vital for tissue elasticity, for example in the dermis, lungs and in the elastic lamellae surrounding blood vessels [32]. Microfibrils can also be found without elastin, and can form other fibrous structures such as the ciliary zonules, (which anchor the eye's lens to the ciliary muscles) [31], the mesangium of the kidney [33], and in the periodontal ligament [34, 35]. General microfibril niches are summarised in Figure 1.5 A.

Before assembly into microfibrils, fibrillin undergoes a number of post-translational modifications, including glycosylation, and cleavage of N and C-terminal propeptides

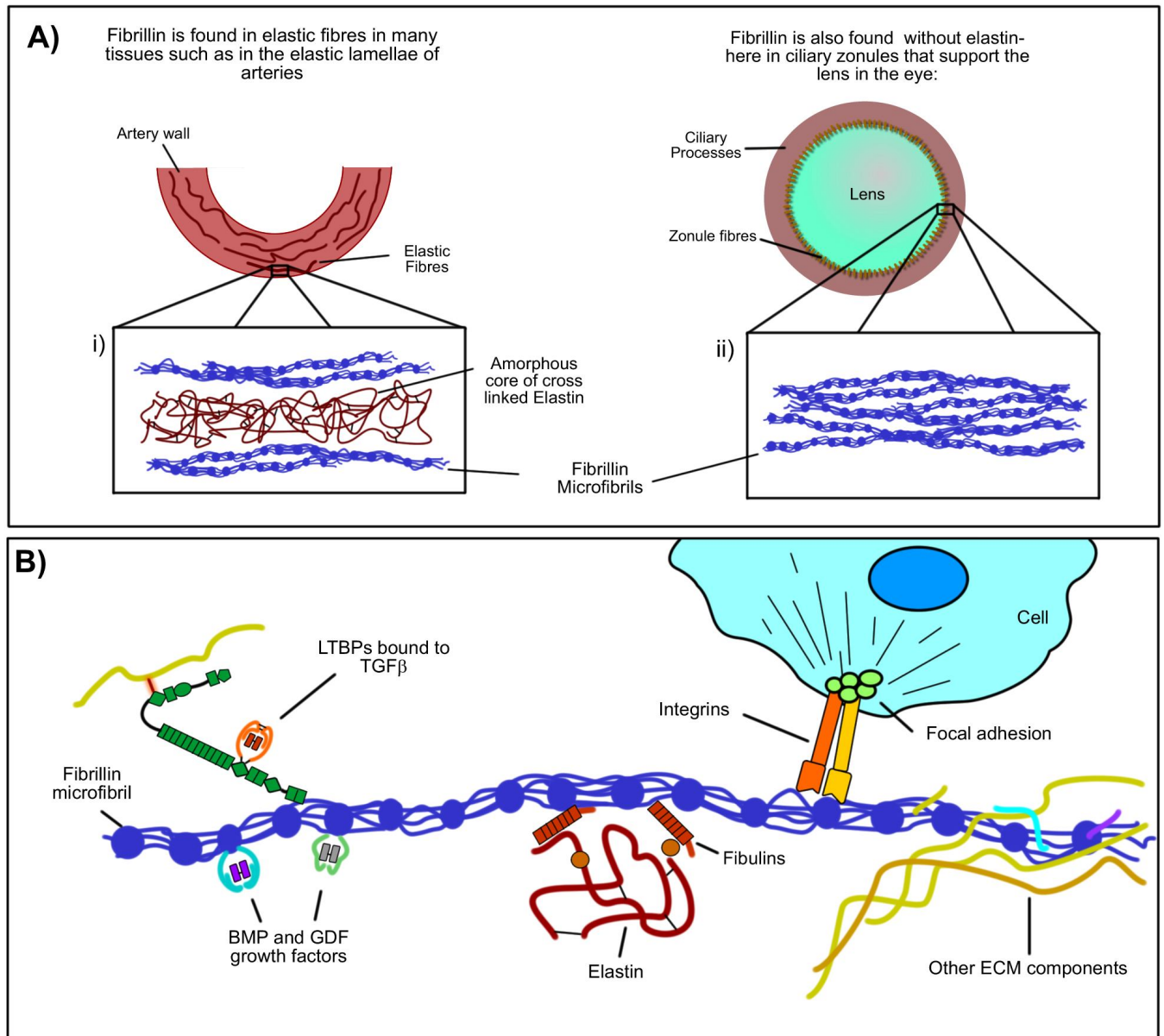


Figure 1.5- The roles of Fibrillin in the ECM.

A) Fibrillin-microfibrils can be associated with i) elastin to form elastic fibres in tissues such as the blood vessel wall, lung and skin, or ii) occur independently, for example in the suspensory ligaments of the eye, periodontal ligament, and mesangium of the kidney glomerulus. **B)** Some of the key cell matrix interactions mediated by fibrillin-microfibrils: LTBPs bind microfibrils and sequester TGF β in a latent state. A variety of BMP and GDF growth factors bind fibrillin directly and are sequestered to the ECM. Microfibrils direct the formation of elastic fibres by interacting with various proteins including fibulins. Integrins bind microfibrils, contributing to cell adhesion. Microfibrils have also been shown to interact with a variety of other proteins and ECM components (Figure 1.7) [6, 58, 59].

by furin [36-39]. How exactly fibrillin monomers assemble to form oligomeric microfibrils is still poorly understood, and the organisation of fibrillin within the microfibril that results is also a matter for debate, with two general models put forward summarised in Figure 1.6. Early efforts at homology modelling of fibrillin based on the rod like NMR structures of tandem cbEGF domains gave an estimated molecule length of around ~150 nm [40]. This is considerably longer than the ~55 nm periodicity of the microfibril beads, and so a staggered model was proposed, in which one fibrillin monomer spans two or more beads [40]. An alternative model has also been put forward where fibrillin molecules are “pleated” in order to fit within one inter-bead distance [41, 42]. The most recent evidence for this model comes from SAXS and single molecule EM studies, in which various fibrillin fragments spanning much of the molecule have been studied, and structures fitted to the molecular shapes obtained [41]. However many of the sharp turns featuring in the various forms of the pleated model, are inconsistent with the canonical inter-domain interactions, as discussed in Section 1.2 [43].

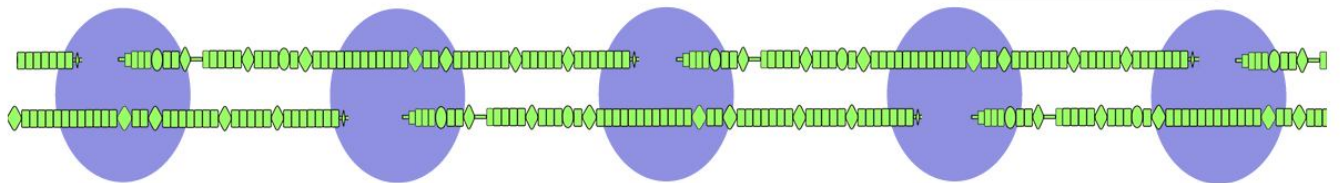
Key questions relevant to both models are: 1) what fibrillin-fibrillin interactions drive assembly? and 2) are other molecules involved in this process? Various fibrillin-fibrillin interactions have been defined using a number of techniques, including ELISA and SPR [44-46], but the challenging solubility properties of fibrillin fragments often cast doubt on the biological significance of some of these results. There is evidence to suggest that fibrillin oligomerisation involves an initial multimerisation stage [46], interaction of the N and C termini [38, 46-48], and covalent disulphide exchange [49], although trans-glutaminase cross links have also been detected in microfibrils [50].

Estimated length of extended Fibrillin monomer: ~150 nm



A - Staggered model

Average interbead distance ~56 nm
Observed by rotary shadowing EM



B - Pleated model

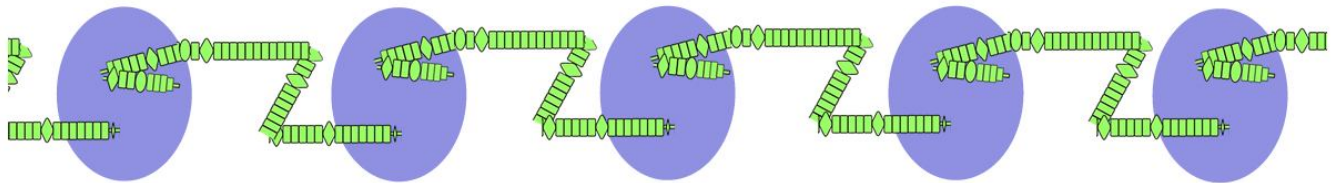


Figure 1.6- Models of microfibril structure

A) Staggered model, where individual fibrillin monomers span two inter-bead distances and adopt a more linear conformation. **B)** Pleated model, where each fibrillin monomer only spans a single inter-bead distance, and is bent / pleated in order to fit within the space.

Fibronectin and certain integrins have been demonstrated to be essential for formation of a fibrillin network in mammalian cell culture [51, 52]. However canonical “beads on a string” microfibrils cannot be extracted from cell cultures until considerably later than an initial fibrillin network is observable [53], and so it is not clear how closely the fibrillin network in these cell culture experiments resembles the microfibril network *in vivo*. Several factors have also been demonstrated to stimulate increased fibrillin deposition in culture, including heparin [54], ADAMTSL6 [55, 56], and ADAMTSL4 [57], but a consensus model for the key steps in microfibril biogenesis has yet to emerge.

Fibrillin has been implicated in a large number of protein-protein interactions, which are summarised in Figure 1.7. Some of these interactions are thought to allow fibrillin to perform a variety of roles in addition to being the structural component of microfibrils, such as directing growth factor signalling, encouraging cell adhesion, and acting as an organiser for the assembly of higher order ECM structures, such as elastic fibres [6, 58, 59]. While the biological relevance of some of these interactions is not clear, a number have had their physiological importance highlighted by human pathologies and mouse models.

Fibrillin in human disease

The physiological importance of the fibrillins has been demonstrated by a number of pathologies resulting from mutations in the fibrillin gene, often with quite contrasting phenotypes, as summarised in Table 1.1. Mutations in genes for different fibrillin isoforms cause distinct conditions, demonstrating their non-redundant roles, but FBN1 mutations in particular have been shown to result in a diverse range of pathologies that can have severe outcomes.

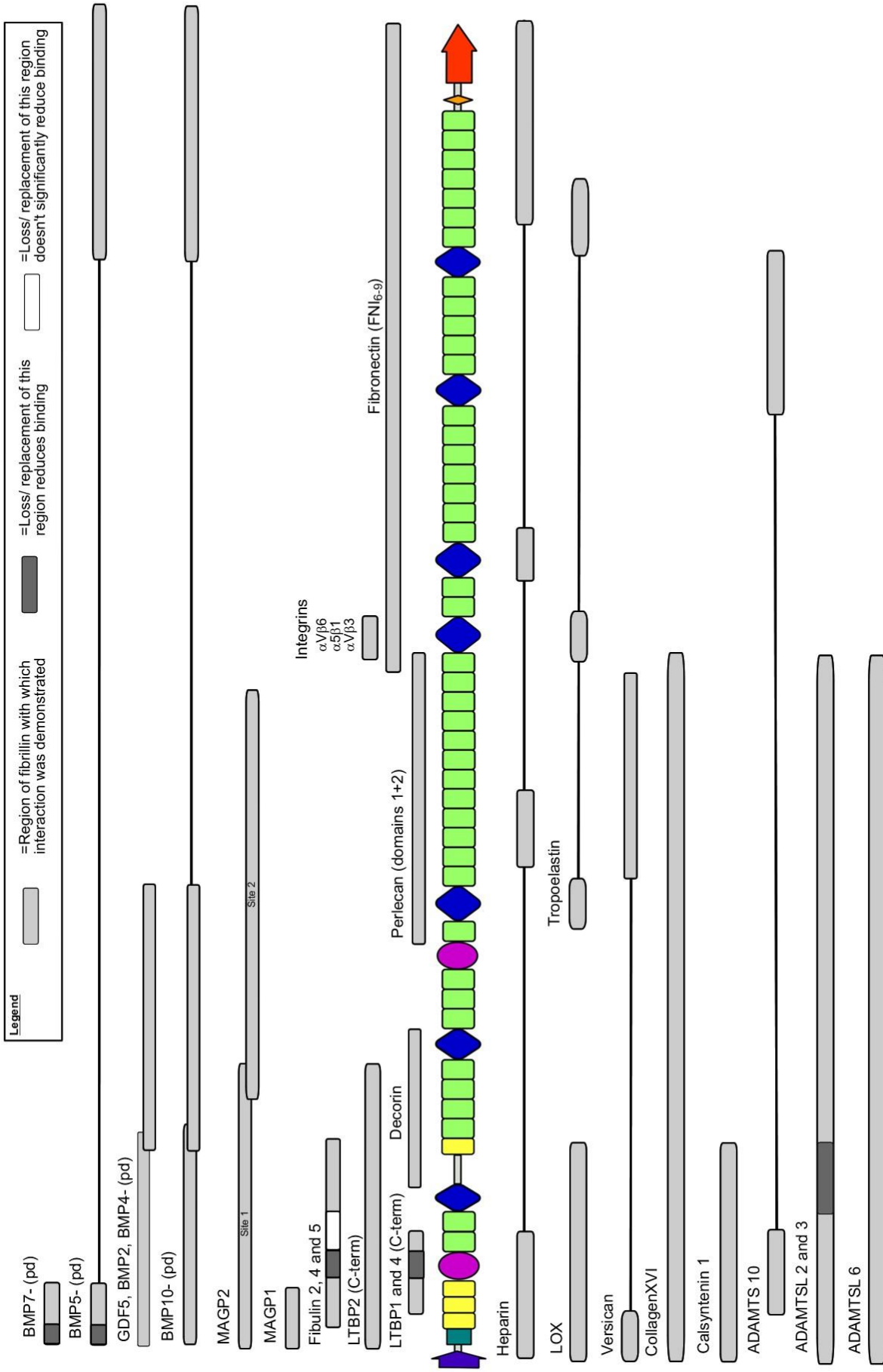


Figure 1.7. Known protein and proteoglycan interactions of fibrillin1 with various other ECM components.

Fibrillin domains coloured as in Figure 1.1. References are as follows: BMP7 prodomain [180], GDF5, BMP2, 4, 7 and 10 [181], BMP5 [182], MAGP2 [183], MAGP1 [184], Fibulin 2, 4 and 5 [15, 185, 186] (microfibril association of fibulin-5 lost on proteolytic removal of fibulin N-terminal cbEGF [187]), LTBP2 [144], LTBP1 and 4 [13, 15], Decorin [188], Perlecan domains 1+2 [189], Versican and Aggrecan [190], Integrins [17, 23-27], Fibronectin [51], ADAMTS 10 [191], ADAMTSL6 [55, 56], ADAMTSL2 and 3 [76], Calsynenin 1 [192], CollagenXVI [193, 194], Tropoelastin [195] (196), LOX [186], Heparin [195], Integrins [17, 23-27].

Table 1.1 – Fibrillinopathies and related genetic diseases

Disease	Symptoms	Genes involved	Notes and role of TGF β	Refs
Marfan syndrome	- Elongation of the long bones - Arachnodactyly (long fingers) - Pectus carinatum/ pectus excavatum - Abnormal joint flexibility - High arched palate - Ectopia lentis and early onset glaucoma - Aortic aneurysm, and dilatation - Mitral valve prolapse	<i>FBN1</i> (dominant)	Excessive active TGF β seen in mouse models, and in patient sera [60]. Increased levels of phosphorylated SMAD 2 seen in MFS aorta [61]	[62-64],
Neonatal Marfan syndrome	-Severe MFS symptoms present at birth -Also exhibit joint contractures, crumpled ears and loose skin	<i>FBN1</i> (dominant)	May represent the severe end of the MFS spectrum rather than a separate disorder	[65-67]
Shprintzen-Goldberg Syndrome	- Craniosynostosis (early fusion of skull bones) - Arachnodactyly (long fingers) and other Marfan like skeletal features - Mitral valve prolapse may occur but aortic dilatation rare - Mental retardation	<i>FBN1</i> (dominant)	(Only two mutations linked to fibrillin, both affecting cysteine residues in cbEGF15)	[68]
Progressive kyphoscoliosis	- Skeletal defects, particularly in the spine, no ocular or cardiac involvement	<i>FBN1</i> (dominant)	(Only one fibrillin mutation described so far)	[69]
Isolated ectopia lentis	- Ectopia lentis	<i>FBN1</i> (dominant) <i>ADAMTSL4</i> (recessive) [70]	Various mutations of a similar nature to MFS mutations, scattered throughout FBN gene	[71]
Familial thoracic aneurysm and dissection	-Thoracic aneurysm and dissection without other marfanoid features	<i>FBN1</i> (dominant)	Only a few mutations, scattered through FBN1 gene	[72]
MASS phenotype	-Mitral valve prolapse, -Aortic root dilatation without dissection, -Skeletal and skin abnormalities	<i>FBN1</i> (dominant)	May represent a milder form of MFS	[73]
Stiff Skin syndrome	- Tight and thick skin - Limited joint mobility - Short stature	<i>FBN1</i> (dominant)	Increase in phosphorylated SMAD 2 in dermis, indicative of TGF β signalling.	[28]
Weill-Marchesani syndrome	- Short stature - Brachydactyly (short fingers) - Thick skin - Myopia, glaucoma, ectopia lentis - Hypermuscularity	<i>FBN1</i> (dominant) <i>ADAMTS10</i> (recessive) (<i>ADAMTS17</i> (recessive) WMS-like syndrome [74]) <i>LTBP2</i> (recessive [75])	ELISA assays of unable to measure a significant difference in active or total TGF β . [76] Normal secretion of mutant FBN, + normal fibril formation in skin	[77]

Geleophysic dysplasia	-Severe short stature and extremities -Stiff joints -Skin thickening -Cardiac valvular thickening -lysosomal storage like vacuoles	<i>FBN1</i> (dominant) <i>ADAMTSL2</i> (recessive GD)	Increased TGF β signalling in cell culture assays where either FBN mutated [78] or ADAMTSL2 [79]	[78]
Acromicric dysplasia	-Severe short stature and extremities -Stiff joints -Skin thickening -Distinctive "round" face -Pseudomuscular build -Distinctive skeletal "notches"	<i>FBN1</i> (dominant)	Increased TGF β signalling in cell culture assays	[78]
Progeria +lipodystrophy	-Lipodystrophy (degeneration of adipose tissue) -Ectopia lentis -Marfanoid habitus (tall stature)	<i>FBN1</i> (dominant)	(only a limited number of patients reported so far)	[80-82]
Beals-Hecht syndrome or congenital contractural arachnodactyly (CCA)	- Contractural arachnodactyly (long fingers bent in at the joints) and contractures of various other joints, (improve with age) - "Crumpled" ears - Tall stature and Marfan like appearance	<i>FBN2</i> (dominant)		[62, 63]
Polycystic ovary syndrome (PCOS)	-Enlarged polycystic ovaries -Menstrual abnormalities -Infertility -Other metabolic consequences	<i>FBN3</i> ?? (precise mutation unknown)	-Loss of FBN 3 staining seen in transitional follicles, but FBN 1 + 2 staining normal	[83] [84-86]
<i>LTBP2</i> mutation	-Primary congenital glaucoma -Secondary glaucoma with megalocornea and spherophakia -Isolated microspherophakia [87]	<i>LTBP2</i> (recessive)	LTBP2 mutation also associated with WMS [75]	[88-90] [91] [92] [87]
<i>LTBP3</i> mutation	-Oligodontia, failure of 6 or more teeth to develop -Short stature	<i>LTBP3</i> (recessive)		[93]
Urban-Rifkin-Davis syndrome (URDS)	Severe developmental defects in the pulmonary, gastrointestinal, genitourinary, musculoskeletal and dermal systems	<i>LTBP4</i> (recessive)	TGF β signalling shown to be significantly higher when patient fibroblasts co-cultured with reporter cells	[94]
Camurati-Engelmann disease (CED)	-Thickening of bones especially shafts of long bones and skull -Sensory impairment may result from smaller passages for nerves and blood vessels -Chronic bone pain in the limbs -Muscle weakness -Waddling gait	<i>TGFB1</i> (dominant)	Missense mutations in propeptide reduce stability of the latent TGF β complex [95] [96]	[97] [98, 99]
Arrhythmogenic right ventricular cardiomyopathy type 1 (ARVC)	Replacement of myocytes in the right ventricular myocardium with Fibrotic and fatty tissue	<i>TGFB3</i> (dominant)	Mutations in 5' or 3' UTR which increase TGF β 3 expression	[100]
Loeys-Dietz syndrome (LDS)	- Cleft palate or bifid uvula - Abnormally long limbs and fingers - Pectus excavatum - Contractures of fingers and toes	<i>TGFBR1</i> <i>TGFBR2</i> (dominant)	-Only missense mutations detected so far -Increased levels of phosphorylated	[61, 102]

LDS continued	- Abnormal joint flexibility, - Widely spaced eyes - Aortic aneurysm, and aortic dissection with tortuosity		SMAD 2 in aorta, but mutations have also been shown to be inactivating [101]	
Multiple self-healing squamous epithelioma (MSSE)	-Multiple skin tumours appear, grow rapidly, and then regress leaving scars	<i>TGFBR1</i> (dominant)	Includes frame shift and nonsense mutations	[103]
Arterial tortuosity syndrome (ATS)	- Tortuosity of large and medium arteries - Aneurysms of large arteries and stenosis of the pulmonary artery - Fragmentation of elastic lamellae	<i>GLUT10</i> (recessive)	-Increased levels of nuclear pSMAD2 and ctgf expression in arterial tissue -Decorin expression severely reduced in ATS but not in LDS VSMCs	[104]

1.3.2 - Marfan syndrome

Marfan syndrome (MFS) is an autosomal dominant genetic disorder, first described in 1896 by the French paediatrician Antoine Marfan. In 1991 it was linked to mutations in the fibrillin gene [105]. Marfan patients exhibit a variety of defects in the skeletal, ocular and cardiovascular systems. Skeletal defects include abnormally long bones, pectus excavatum, high arched palate and joint hyper-mobility. Sufferers may also develop ectopia lentis, and have a high risk of developing aortic dilatation and dissection, which can often be fatal [62].

Over 1000 different mutations in fibrillin 1 have been associated with Marfan syndrome, including nonsense mutations, exon skipping and cryptic splice site mutations, base insertions, base deletions, and a large number of missense mutations resulting in single amino acid substitutions [62-64, 106]. While the full mechanism of Marfan syndrome is still unclear, Figure 1.8 summaries some of the points at which Marfan mutations may interfere with the biosynthesis of fibrillin.

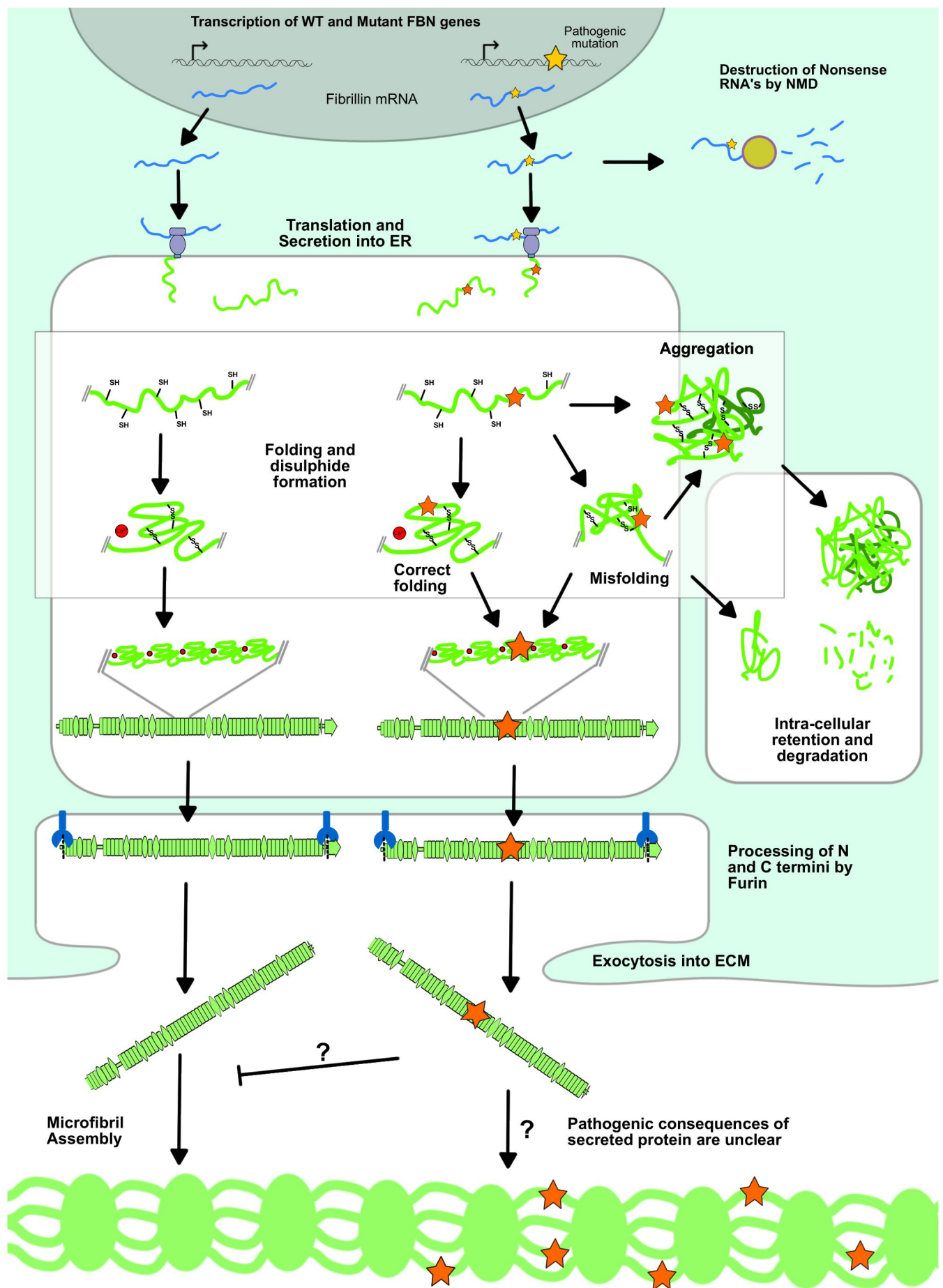


Figure 1.8 - Mechanisms for the disruption of Fibrillin production in MFS

The diagram follows production of fibrillin, starting at the top, with transcription of the fibrillin mRNA, and its translation and secretion into the Endoplasmic reticulum. The fibrillin protein then folds, with a significant contribution from disulphide bond formation, to form a large multi domain protein, which is then secreted from the cell. Cleavage of N and C Pro-peptides occurs at the cell surface [39]. Fibrillin then assembles into extracellular microfibrils via an unknown mechanism. Marfan mutations may disrupt this process at various points and lead to haploinsufficiency.

As MFS is dominantly inherited there was initially uncertainty over whether MFS mutations acted in a dominant negative fashion, in which the mutant allele interferes with the function of the normal gene [73], or whether the disease was caused by haploinsufficiency, where despite having one normal allele, cells are unable to produce enough fibrillin overall to maintain a minimum level of function. Today there is a large amount of evidence demonstrating that fibrillin haploinsufficiency is capable of causing MFS on its own, (although dominant negative effects have not been ruled out in some situations).

The haploinsufficiency mechanism is particularly well illustrated by a small group of Marfan patients who have a complete deletion of a single copy of the fibrillin gene [107]. These patients display the classical Marfan phenotype, and as there is no mutant fibrillin gene present in their genomes, this rules out any dominant negative effect from a mutant allele. This observation is also reinforced by numerous nonsense MFS mutations, which if they occur prior to the last exon junction, will target the mRNA for degradation by NMD [73]. The severity of the disease suffered by patients with nonsense mutations is variable [73, 108-111], but no nonsense mutations have been associated with the most severe presentations of the disease.

The majority of Marfan mutations however are missense mutations that do not inhibit the translation of mRNA into protein, or the transport of protein into the endoplasmic reticulum [62]. In this case the next hurdle in protein production becomes whether the mutant protein can pass successfully through the intracellular processing pathways of the endoplasmic reticulum (ER) and Golgi apparatus, and escape the surveillance mechanisms of the cell. Many missense MFS mutations have been shown to interfere with fibrillin production at this stage, and cause intracellular retention [112]. Frequently

MFS missense mutations result in either the removal of a single cysteine, or the introduction of an additional cysteine into the protein sequence. As almost all the domains in fibrillin rely on disulphide bond formation to stabilise their fold, an odd number of cysteines in a domain may seriously disrupt folding, through formation of incorrect intra or inter-molecular disulphides. The quality control machinery of the ER (including chaperones such as calreticulin [113]) may then detect the resulting misfolded protein and protein aggregates, and retain them inside the cell for degradation.

However not all MFS mutations cause the fibrillin protein to be retained in the cell, and several have been shown to be secreted into the ECM at WT or close to WT levels [112, 113]. These missense substitutions may allow normal domain folding in the ER, or disrupt folding in a more subtle fashion, so that the protein can escape the proof reading ability of the cell, and be secreted into the ECM. The mechanisms by which these MFS substitutions affect the fibrillin network, and cause MFS, are less clear, although several possibilities can be hypothesised;

- 1) They are unable to assemble into functional microfibrils and so still result in haploinsufficiency
- 2) They interfere with the assembly of WT fibrillin protein into microfibrils, and so exert a dominant negative effect on microfibril biogenesis
- 3) They are able to assemble into microfibrils but exert a dominant negative effect on microfibril function, for example by:
 - a. Perturbing the overall structural properties of the microfibril.
 - b. Increasing susceptibility to proteolysis and microfibril degradation

- c. Disrupting the normal interactions of microfibrils with other ECM proteins important for microfibril function.
- 4) Exert microfibril independent pathogenic effects.
- 5) Some combination of the above.

As well as being caused by a wide variety of mutations, the symptoms of MFS can also vary greatly from patient to patient, both in nature and severity. In mild familial cases some sufferers may develop to maturity without any major symptoms warranting diagnosis. While in particularly severe cases, the symptoms of MFS can be present at birth, and death from congestive heart failure can occur before the patient reaches 2 years old [65, 66]. This severe form of the disease has sometimes been described as “neonatal Marfan syndrome”, although due to the broad spectrum of MFS phenotypes and severity, it can be difficult to define precise boundaries between classes of MFS.

Regardless of the neonatal classification, mutations causing severe early onset MFS have been consistently localised to a particular region of the fibrillin gene, between exons 24 and 32. This corresponds to the fibrillin TB3 domain through to cbEGF18, in the middle of the protein. Certain specific mutations in this region consistently result in severe early onset MFS, and have not been found in patients with milder forms of the disease. This phenotypic consistency, and specific localisation of mutations, suggests that background genetic factors are unlikely to be solely responsible for the increase in disease severity in these neonatal, or atypically severe patients [65-67]. It should also be noted that a particularly high concentration of classical MFS mutations has been found in this “neonatal” region of the fibrillin gene, suggesting this region of fibrillin is particularly sensitive to disruption.

In summary the large amount of genetic and biochemical evidence collected from studying human Marfan patients, points towards multiple pathological mechanisms affecting fibrillin production, all of which lead to a deficiency in the fibrillin network produced by cells. With the characterisation of null and nonsense FBN1 mutations causing classical MFS, we can conclude that haploinsufficiency alone is able to cause the disease, but in some cases unknown dominant negative mechanisms may be at work, especially in the case of severe neonatal MFS.

Questions arising from studies of MFS patients

As well as uncertainty over the downstream consequences of some MFS mutations for the fibrillin network, it is also unclear how a deficiency in the fibrillin network translates to the developmental and pathological changes seen in MFS. The structural role of fibrillin microfibrils intuitively suggests that a lack of fibrillin may alter the structural properties of the ECM, and a resultant weakening of specific ECM structures might seem sufficient to explain symptoms such as ectopia lentis and aortic dissection. However it is harder to relate tissue weakening to the developmental differences seen in MFS, such as elongation of the long bones, and high arched palate. It is also unclear what percentage of normal fibrillin and microfibrils are lost from the ECM due to haploinsufficiency in MFS patients. Fibroblast cultures and tissue samples from patients frequently show a reduction in fibrillin staining and increased fibrillin network “disorganisation”, but these results are often variable, and may reflect a significantly greater than 50% reduction in fibrillin levels, suggesting some positive feedback mechanism may be at work [114-116]. This would require further downstream feedback mechanisms to be set in motion by an initial fibrillin shortage that could exacerbate the fibrillin deficit, and the disease outcome.

1.3.3 - Mouse models of MFS

Potential clues as to the nature of the feedback mechanisms that further erode the fibrillin network, comes from the large number of mouse models, which have provided valuable insights into the fibrillin field. An extensive summary of these mice, their phenotypes, and significant biochemical observations, is given in Appendix 3.

Generally these mouse models support and enhance the conclusions drawn from studies of Marfan patients (although in many cases the insights from mouse models pre-date their observation in patients). Several mouse models were critical for demonstrating that FBN1 haploinsufficiency alone is capable of re-capitulating the phenotypic abnormalities seen in humans, especially those relating to aortic dissection. The first clear evidence for haploinsufficiency in a mouse model came from the mgR mice [117]. These mice do not have a mutation in the FBN1 gene, but exhibit reduced expression due to inclusion of a neomycin resistance cassette (produced in the process of generating the mg Δ mice [118]). Each mgR mouse allele displays an approximately five fold reduction in fibrillin1 expression compared to WT, and so homozygotes express ~20% of WT levels of fibrillin1. About 3-4 weeks after birth homozygous mgR mice display fatal aortic dissection, while heterozygotes on the other hand express ~70% of normal fibrillin levels and are phenotypically normal. This suggests a “threshold” level of fibrillin somewhere between these two amounts, below which an MFS like pathology is manifest.

This threshold level was further investigated with studies on mice with a complete knockout of the fibrillin gene, the mgN mice [119]. Homozygotes (with 0% fbn1) show a more severe pathology than the mgR/mgR mice, dying of aortic dissection within the

first 2 weeks of birth, while heterozygotes (50% fbn1) are phenotypically normal (although may display sub clinical abnormalities in the aortic wall [119, 120])). Additionally almost all fbn1 ^{-/-} and fbn2 ^{-/-} mice, and half of fbn1^{+/-} fbn2^{-/-} mice die *in utero*, demonstrating critical compensation for fbn1 by fbn2 during embryonic development. These studies [121] showed that when considering fibrillin threshold levels, fibrillin2 should also be factored in, despite the fact that fbn2^{-/-} mice do not display the aortic abnormalities seen with fbn1 ^{-/-} mice (Appendix 3).

The observation that heterozygous fbn1 null mice do not show any serious pathology, while humans with a single null allele display classic MFS, suggests that the haploinsufficiency threshold is different in mice and humans. This could be for a variety of reasons, such as:

1. Increased compensation by fibrillin2 in mice
2. The shorter mouse life cycle
3. Developmental processes that are less sensitive to fibrillin levels in mice
4. Reduced biomechanical stress imposed on the fibrillin matrix in these smaller organisms.

Generally though, the close phenotypic similarities and concordant biochemical observations suggest that many of the mouse models are good representations of human MFS.

Potential dominant negative mechanisms for MFS have also been investigated with mouse models. For example, a mouse was generated using a transgenic yeast artificial chromosome system to express human FBN1 with a MFS causing C1663R

substitution [108]. These mice were phenotypically normal, and showed incorporation of the human C1663R FBN into microfibrils. In the same study another MFS mouse was generated with a C1039G substitution in its native fibrillin1 protein. Mice heterozygous for this C1039G mutation display some non-fatal aortic abnormalities, which were rescued by crosses with transgenic mice expressing WT human fibrillin, and some rescue was also observed with the transgenic human C1663R FBN, further demonstrating that this mutation doesn't act in a dominant negative fashion.

Unfortunately limited data is published on the heterozygous mgN FBN1 null mice, but the suggestion from the literature is that C1039G heterozygotes display a more severe, although non-fatal, aortic phenotype than the mgN heterozygotes [119]. This in itself might suggest some dominant negative effect for this C1039G substitution in mice, if the heterozygous phenotype is consistently worse than that of mgN heterozygotes. It is also worth noting that this C1039G substitution is in the “neonatal” region of fibrillin, although the equivalent human mutation (C1039Y) results in a classical MFS phenotype [207, 208]. At present no mouse models of neonatal MFS have been published.

1.3.4 - Fibrillin and growth factor signalling

As well as providing insight into haploinsufficiency, mouse models have also provided important clues into downstream events that may occur as a result of a deficiency in the fibrillin network, and their potential role in MFS pathology.

One key insight came from studies on mg Δ mice, in which several core fibrillin exons are deleted, corresponding to domains cbEGF8-TB3 [118]. This was the first mouse model of MFS to be constructed, and displays a similar phenotype to the mgN, and

C1039G mice, with homozygotes dying a few weeks after birth from aortic dissection. As well as aortic abnormalities, these mice show defects in lung development [209], particularly in alveolar septation, a process by which growth and involution of the alveoli epithelium generates a large surface area for gas exchange. In $mg\Delta$ mice, (and other mouse models of MFS), histology of the lungs shows larger alveoli airspaces during postnatal lung development. Further analysis by immunohistochemistry, and crosses with transgenic reporter mice, showed increased signalling by a specific growth factor; TGF β . Furthermore, treatment of $mg\Delta$ mice with a TGF β neutralising antibody was shown to restore normal alveolar septation, implicating TGF β as a key mediator of developmental responses to fibrillin deficiency.

Numerous other studies since this, many of which were conducted in the alternative C1039G mouse model, have revealed dysregulation of TGF β in multiple tissues affected by MFS, including the mitral valves [122], muscle [123], and aorta [124]. Inhibition of TGF β signalling, either directly with antibodies, or indirectly through the angiotensin II type 1 receptor antagonist losartan, (which indirectly inhibits TGF β signalling via several pathways [125]), shows significant restoration of normal development in all of these tissues.

These mouse studies revealed a critical role for TGF β signalling in the disruption of development in murine tissues caused by fibrillin1 deficiency. This also provided a significant breakthrough in the treatment prospects for patients with Marfan syndrome, as TGF β signalling is a good target for pharmacological intervention, and losartan in particular has already been clinically approved for treating hypertension [124]. Until the discovery of the TGF β link to MFS, therapeutic prospects for MFS patients appeared dependent on advances in gene therapy.

However a key question, when considering therapeutic inhibition of TGF β signalling, is whether the growth factor dysregulation seen in mouse models also occurs in humans. Numerous pieces of evidence seem to support this, including observation of increased pSMAD2/3 in some MFS tissues [61], and direct measurement of increased active TGF β in MFS patient sera [60], although this does not in itself demonstrate a causal role for the growth factor. Clinical trials of losartan are currently underway [210], and may provide the necessary evidence to verify an important role for TGF β signalling in the pathology of human MFS.

Additional genetic evidence also exists supporting a role for TGF β in human MFS, which comes from the genetic disorder “Loeys-Dietz syndrome” (LDS). LDS patients have various phenotypic similarities to MFS patients, with elongation of the long bones, and a high-risk of aortic dissection, but also have some differences, such as aortic tortuosity, and presence of a “bifid uvula”, due to incomplete palate fusion [61, 102]. This disease has been linked to mutations in TGF β receptor genes (TGFB1 and TGFB2), and so the overlapping phenotype with MFS strongly suggests a role for TGF β signalling in the same processes of tissue development and homeostasis that are disrupted in MFS.

Some contradiction exists however, in that LDS missense mutations in the TGF β receptor genes were initially predicted to cause loss of receptor function, which is supported by some recent cell culture assays [101]. This loss of receptor function would be expected to reduce TGF β signalling, in contrast to the excessive TGF β activation seen in MFS. However increased staining of nuclear pSMAD2 and CTGF in the aortic media of LDS patients is suggestive of an increase in TGF β signalling in these tissues. Additionally while LDS has only been linked to miss-sense mutations in TGFB1, TGFB2, and TGFB3, MFS has been linked to mutations in all three genes.

nonsense and frame-shift mutations in the *TGFBR1* gene have been linked to a distinct disorder; “Multiple self-healing squamous epithelioma” (MSSE) [103]. Further experiments are required then to elucidate the detailed mechanisms of LDS pathology.

1.3.5 - Other genetic diseases affecting the fibrillins

Specific mutations in *FBN1* have also been associated with a variety of other genetic disorders, (also summarised in Table 1.1), including stiff skin syndrome [28], Shprintzen-Goldberg syndrome [68], autosomal dominant Weill-Marchesani syndrome [77], geleophysic and acromicric dysplasias [78], and a recently described progeria and lipodystrophy phenotype [80-82].

Stiff skin syndrome

In many cases the symptoms of these diseases contrast significantly with those of MFS. Patients with stiff skin syndrome (SSS), for example, exhibit shortened stature, thickened skin, and in-flexible joints, in marked contrast to the tall stature, and joint hypermobility of MFS patients [28]. MFS patients may also sometimes develop unexplained “stretch marks” in the skin, but do not have the excessive skin thickening seen in SSS. The mutations linked to SSS are clustered within the TB4 domain of fibrillin1 [28], which also contains an RGD sequence known to bind integrins. Fibroblasts exhibit defective spreading and migration on a substrate of SSS fibrillin compared to WT fibrillin [28], and so one potential disease mechanism is that in SSS the mutant fibrillin allele may exert a dominant negative affect through disruption of integrin mediated microfibril-cell interactions.

A mouse model that is potentially relevant to SSS exists in the form of the tight skin (TSK) mouse, where heterozygotes exhibit a similar skin thickening to SSS sufferers

[126] (Appendix 3). However, these mice do not have a missense disruption in the TB4 domain, but instead have a large duplication within the fibrillin gene, giving rise to a much longer protein, which contains two RGD containing TB4 like domains instead of one.

Whether the pathological outcome in TSK mice is due to an excess of integrin sites, or another factor, such as altered RGD presentation within the microfibril, has not been determined. Interestingly though, SSS patients also show an increase in pSMAD2 in the dermis, suggesting increased TGF β signalling in this tissue, and also TGF β 1 -/+ TSK heterozygous mice do not develop skin fibrosis, suggesting a critical role for increased TGF β signalling in the tight skin phenotype.

Weill-Marchesani syndrome

Weill-Marchesani syndrome (WMS) is another contrasting pathology that has been linked to autosomal dominant mutations in fibrillin1. Here short stature is also observed, but there are a number of differences from stiff skin syndrome, such as frequent presentation of ectopia lentis and glaucoma [77].

Only a small number of mutations in fibrillin have been linked to this disease, but unlike SSS they do not show localisation to a single region of fibrillin. They include a substitution of arginine 1596 to proline in TB4 [127], an eight amino acid deletion in TB5 [77], a glycine 214 to serine substitution in hyb1 [127], and a complete deletion of the TB1 to EGF4 exons [76]. The latter mutation has been best studied, with a mouse model constructed with the same three-domain deletion [76]. This mouse displays some of the characteristics of WMS patients, with increased skin fibrosis and reduced bone growth. Both patients and mice show normal incorporation of fibrillin into

microfibrils, although EM studies of mice tissues show formation of large microfibril aggregates in the skin and “moth eaten” elastic fibres, and so it appears that this WMS deletion is having a dominant negative effect on fibrillin function within the matrix.

The precise mechanism by which fibrillin mutations cause WMS is still unclear, but the presence of ADAMTS10 mutations in patients with recessive WMS [74], suggest that a functional linkage between this protease and fibrillin may be involved. Several ADAMTS-like proteins (ADAMTSL 2 and 3) have been shown to interact with the TB1-EGF4 region of fibrillin [76], and these may then recruit ADAMTS10 to form a large complex of unknown function. The observation of interactions between some ADAMTS-like proteins and LTBP [76] has also led to the suggestion that TGF β regulation may be involved [76], but cultured WMS fibroblasts showed no increase in gross active TGF β levels when measured by ELISA [76]. Although this does not rule out more localised effects on TGF β signalling.

Geleophysic and Acromicric dysplasias

As well as the WMS-associated deletion in the TB5 domain, recently a number of other mutations in this domain have been linked to the geleophysic and acromicric dysplasias, where patients also exhibit short stature, and thick skin [78]. Fibroblast cultures from these patients demonstrate a disorganised fibrillin network and increased TGF β signalling [78].

Recessive mutations in ADAMTSL2 have also been shown to cause geleophysic dysplasia [79], suggesting a functional connection between this protein and the TB5 region of fibrillin. As these dysplasias are phenotypically similar to WMS it is also possible that they share common disease mechanisms. It has been speculated that

the staggered model for microfibril assembly could potentially support a close positioning of the fibrillin TB5 and TB1-EGF3, and both regions could then be involved in formation of an ADAMTS-fibrillin complex that utilises multiple binding sites [76].

Progeria and lipodystrophy

Very recently a number of patients have been discovered with a marfanoid syndrome that also presents with progeroid features at birth and a severe loss of adipose tissue [80-82]. In the limited number of patients characterised so far, all have had mutations affecting the very last exon of fibrillin1. Several of these mutations cause frame shifts, but as they are in the last exon, the mRNA should not be degraded by NMD. Interestingly, these mutations are also found after the C-terminal furin site, and so suggest an important biological role for the unique fibrillin C-terminal region, which is not incorporated into microfibrils itself.

1.3.6 - Conclusions on the biological roles of fibrillins

The data from genetic diseases and animal models clearly demonstrates important and multi-faceted biological roles for the fibrillin family. There are still many unanswered questions however, especially when it comes to relating specific defects in the fibrillin matrix to the diverse phenotypic outcomes observed.

In many of these pathologies aberrant TGF β signalling appears to play a role, and in MFS mice it has been shown to be a critical factor in disease progression. A key question for further understanding the fibrillinopathies is; how do alterations to the fibrillin matrix disrupt TGF β signalling and lead to abnormal tissue development and homeostasis?

1. 4 - TGF β signalling and LTBP

1.4.1 - The TGF β signalling pathway

TGF β is an extremely important growth factor, which plays a significant role in the development and homeostasis of most human tissues. With over fifty thousand papers in the pubmed database related to TGF β , its biological significance has been extensively studied, and it has been shown to perform crucial roles in wound healing, immunity, muscle differentiation, palate closure, bone growth, control of cellular proliferation, and many more developmental and homeostatic processes [6].

The TGF β signalling pathway itself is complex, with many different levels of regulation, both outside and inside the cell. The main features of the pathway are summarised in Figure 1.9. Crucially TGF β is “packaged”, prior to secretion from the cell, in a complex with its pro-peptide and an LTBP protein [3], this assemblage is referred to as the “large latent complex” (LLC). In this form TGF β is inactive and cannot bind to its receptors. LTBPs also provide an adaptor function that allow the large latent complex to bind ECM fibres, the questions surrounding the functional significance of this will be discussed in more detail later. Before TGF β signalling can occur, the growth factor must be “activated”, and released from the LLC, A large variety of factors have been implicated in this activation process, although they may not all be physiologically relevant [128]. Once activated TGF β can then bind to its receptors and initiate intra-cellular signalling cascades, this involves the phosphorylation of the SMAD2 and SMAD3 transcription factors, which then transmit the TGF β signal to the nucleus. This signalling by free TGF β may also be modulated by other factors in the ECM binding TGF β , such as collagen and decorin [3]. To attenuate the growth factor signal, TGF β is

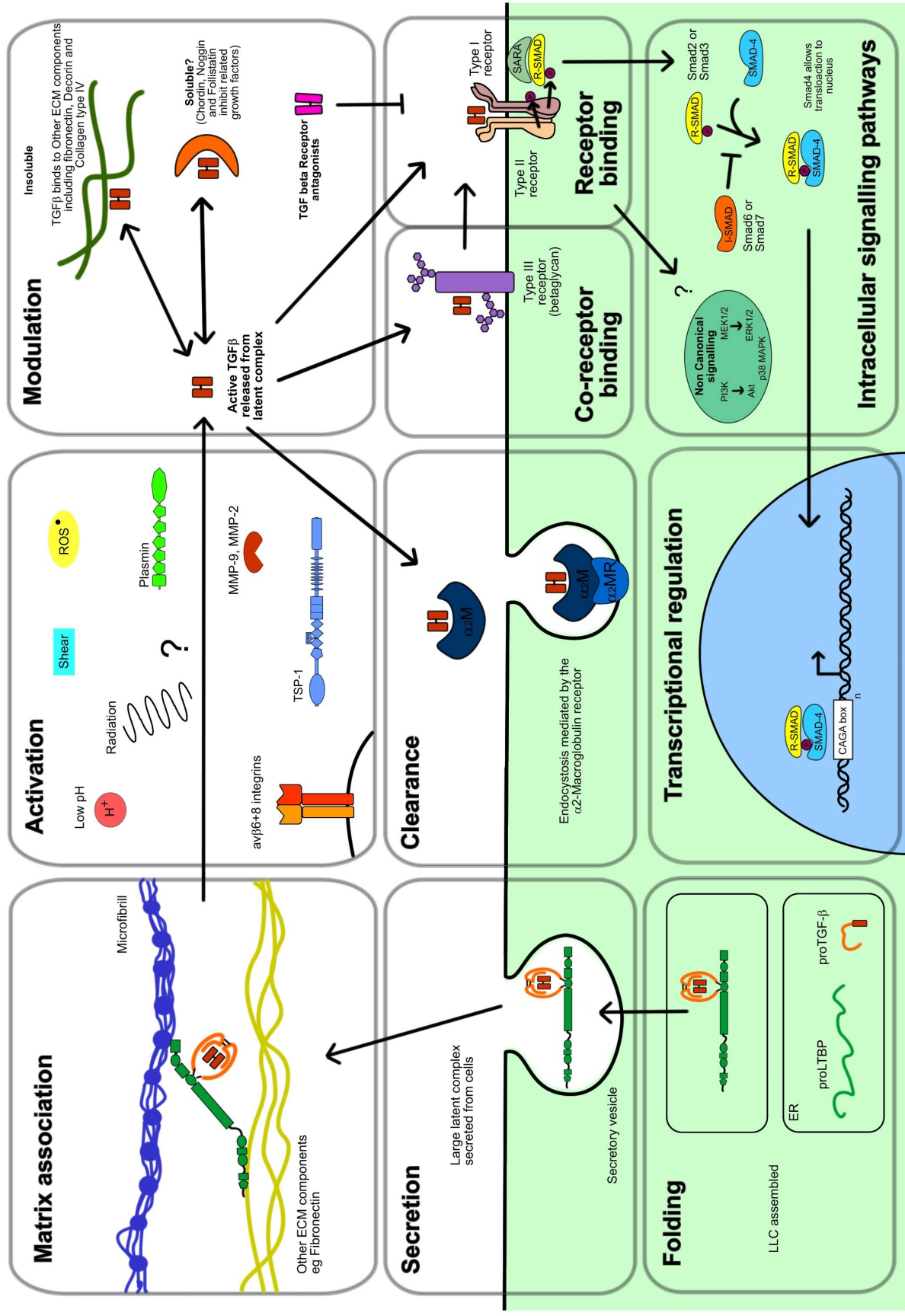


Figure 1.9- TGFβ signalling

A summary of the main stages of TGFβ signalling, starting at the bottom left, where TGFβ associates with LTBP inside the cell to form the large latent complex (LLC). After secretion it is localised to ECM fibres via LTBP mediated interactions. Active TGFβ can then be released by a number of factors. Once active TGFβ is released from the LLC it may also bind various other matrix components which may modulate TGFβ signalling [2]. TGFβ, once released is cleared from the extracellular space by α2-macroglobulin and endocytosis [2]. TGFβ can induce down stream signalling by binding its type 1 and type 2 receptors and forming a hetero-tetrameric complex [197], which may also be assisted by interaction with beta-glycan [198]. Formation of the receptor complex then initiates intracellular signalling. The canonical signalling pathway involves phosphorylation of R-SMAD, which then interacts with co-SMADs and is translocated to the nucleus, where it regulates transcription through interactions with CAGA repeat elements [199]. Non-canonical signalling pathways are also activated involving Erk 1 and 2 transcription factors, although the details of their activation are not well understood [200].

eventually cleared from the extracellular space by $\alpha 2$ macroglobulin sequestration followed by endocytosis [2].

There are three TGF β isoforms encoded by the human genome (TGF β 1, 2 and 3), which are also part of a much larger group of ~45 human growth factors, which include many of the bone morphogenic proteins (BMPs), growth and differentiation factors (GDFs), and several other important cytokines [129, 130]. This group of growth factors is described as the “TGF β superfamily”, which is in turn part of a large, diverse, group proteins that share a “cysteine knot” type motif. These cysteine knot proteins can be quite different in sequence, and may also contain additional domains not found in the TGF β superfamily. However the shared presence of the cysteine knot domain suggests a common evolutionary origin for all of these proteins, even though they now perform highly diverse biological functions [131].

The domain structure of TGF β 1 (Figure 1.10 A) consists of an N-terminal pro-peptide region, which is cleaved from the C-terminal growth factor region by a furin protease. When cleaved by furin the propeptide remains associated with the dimeric growth factor via non-covalent interactions, and when bound to its propeptide in this way TGF β is unable to bind to its receptors, and so is inactive. This complex is referred to as the small latent complex (SLC), the structure of which has recently been determined by X-ray crystallography [96], and shows that the propeptide forms a large folded structure around the TGF β growth factor (Figure 1.10 B).

This crystal structure was obtained with a mutant form of TGF β 1, in which cysteine 33, near the N-terminus of the propeptide, is mutated to a serine to remove a free sulphydryl. It has been demonstrated that TGF β 1 and its propeptide are secreted in a

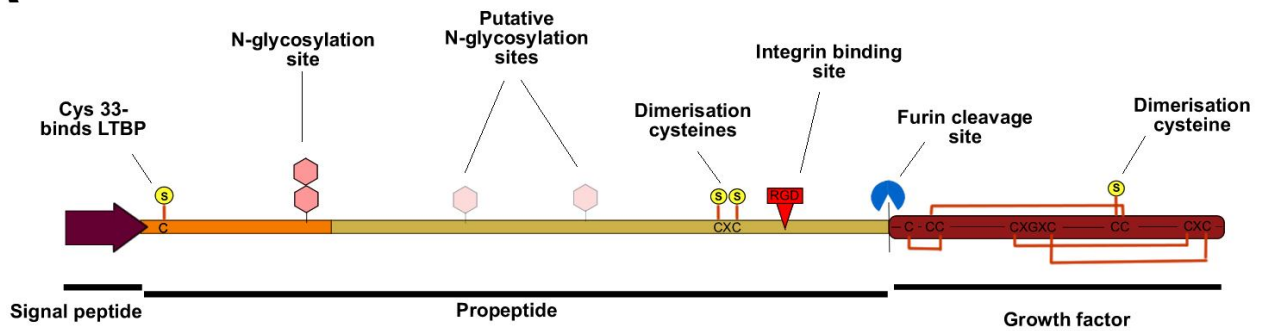
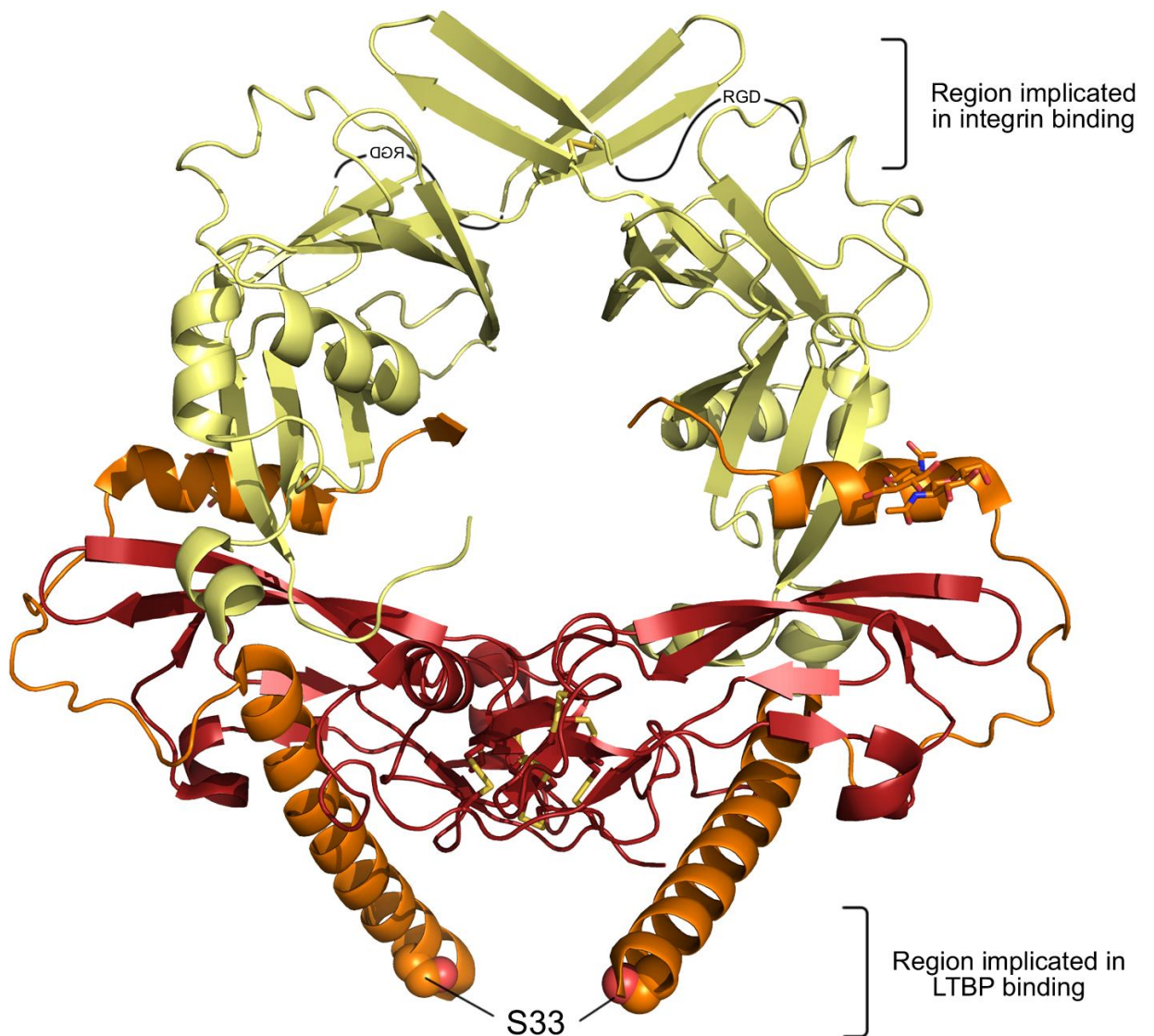
A**B**

Figure 1.10- Structure of the TGFβ1 Small Latent Complex.

A) Shows the key features of the TGFβ1 pro-protein. **B)** Demonstrates the recently determined crystal structure of TGFβ in complex with its pro-peptide, which is coloured according to A, with the growth factor in dark red, the N-terminal region of the pro-peptide (which binds the growth factor) is coloured in orange, and the rest of the pro-peptide is coloured a straw yellow. The disulphides are represented as sticks, while residue 33, a serine in this structure which corresponds to cysteine 33 in WT protein, is highlighted by sphere representation. The position of the loop containing the integrin binding RGD sequence is also drawn in, in black, although it was not resolved in this crystal structure.

covalent complex with LTBP, bound via disulphide linkage of the propeptide of TGF β , with cystine residues in the second TB domain of LTBP (via cysteine 33) [29], forming the LLC. It has been shown that the cells expressing LTBP1 and forming the LLC, significantly enhances TGF β secretion, and ensures correct disulphide bond formation in the SLC [8].

As well as LTBP expression enhancing secretion of TGF β , it has also been shown that LTBP3 is dependant on binding to TGF β for its own release from the cell [132]. Although LTBP1 and 4 on the other hand can be secreted without binding TGF β , sometimes in significant amounts [3].

1.4.2 - Interaction of LTBPs and TGF β

The interaction between the second TB domain of LTBP1 and the propeptide of TGF β has been shown by site-directed mutagenesis to be dependent on a number of specific sequence features within this TB domain (Figure 1.11 and 1.12). These features include the insertion of two amino acids (often FP) into the β sheet region prior to the 7th cysteine, and the presence of several acidic residues [10, 11]. The solution structure of this domain shows that the two amino acid insertion exposes the 2-6 disulphide bond, making it available to react with the propeptide of TGF β [19]. These inserted residues are not found in the equivalent domain in LTBP2, which does not bind TGF β [10, 11], and are also absent from the fibrillin TB domains. However it has been shown that LTBP2 can be engineered to bind TGF β covalently, just by inclusion of the FP insert from LTBP1 into its sequence [10].

Several acidic residues also surround the exposed 2-6 disulphide bond in LTBP1 TB2 (highlighted in Figure 1.12), and substitution of these residues with alanine inhibits the

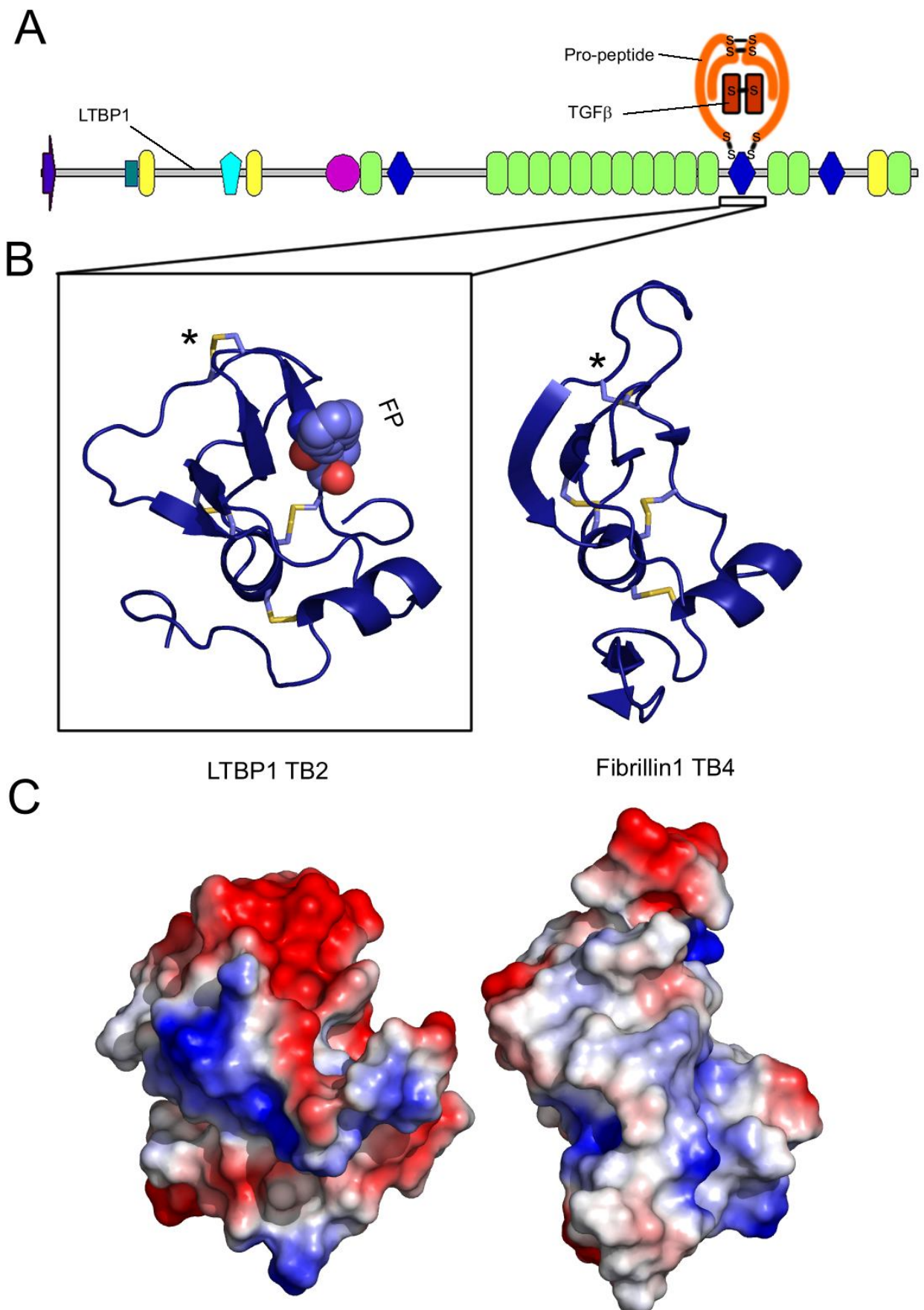


Figure 1.11 - TGFβ interacting region of LTBP1

A) TGFβ specifically binds to the second TB domain of LTBP, via disulphide linkage. Domains are coloured as in Figure 1.1. **B)** 3D structure of LTBP1 TB2 (PDB 1KSQ), compared to fibrillin 1 TB4 (PDB 1UZJ). The FP amino acid insertion is highlighted as spheres, and an asterisk highlights the 2-6 disulphide bond in both TB domains, buried in Fibrillin TB4, but surface exposed in LTBP TB2. **C)** A pymol based rendering of surface electrostatics of the above TB domains. Areas of negative polarity are highlighted in red, positive areas are shown in blue, and neutral areas are colourless.

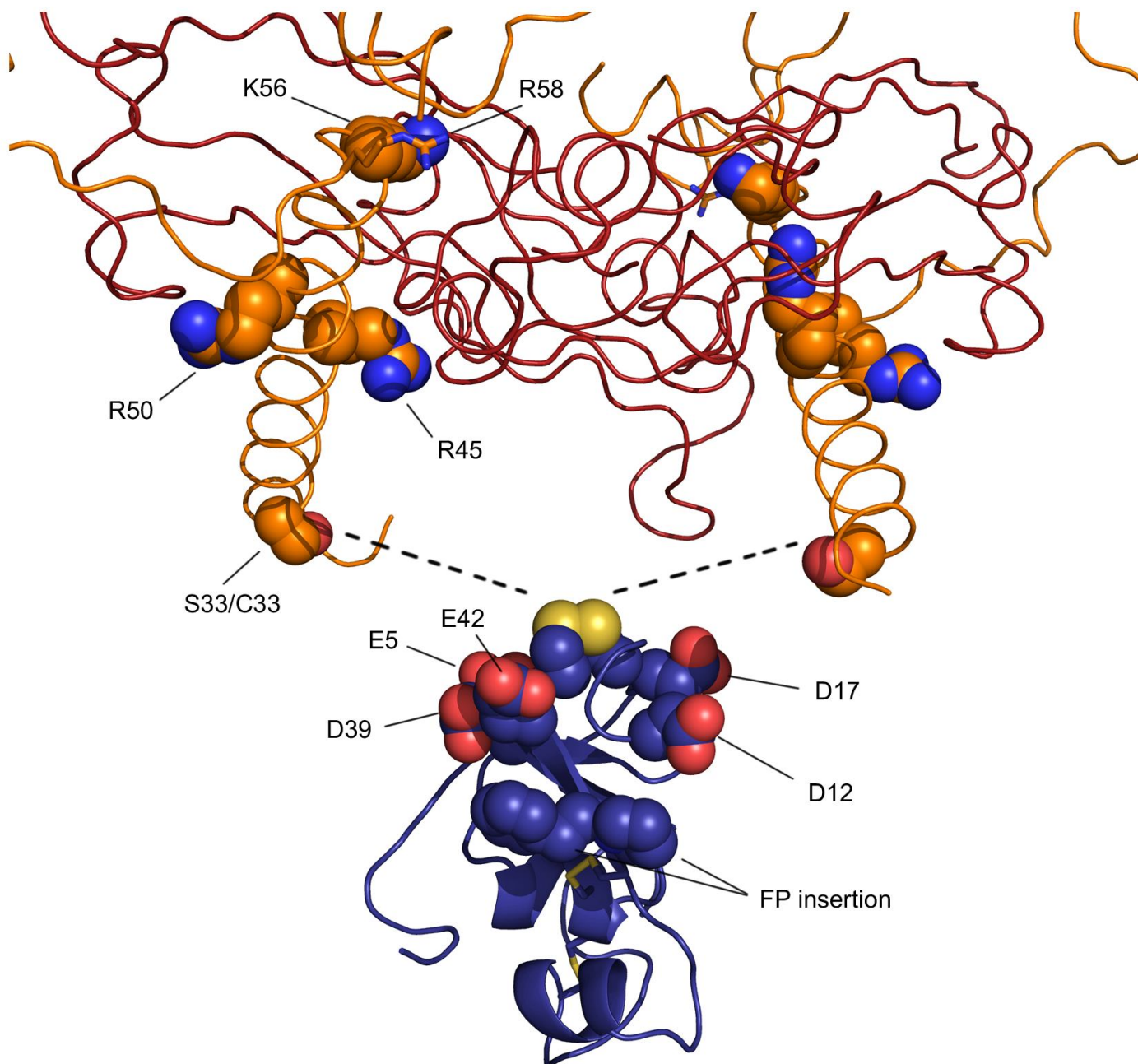


Figure 1.12 - Key residues in the LTBP1-TGFβ1 interaction

Structures of the SLC (propeptide- orange, growth factor- dark red), and LTBP1 TB2 (navy blue) placed beside each other. Residues implicated as important for the TGFβ1 - LTBP1 interaction are highlighted as spheres according to [10, 11] and [95] (LTBP residues numbered as in [11]), TGFβ1 R58, which did not affect LTBP binding, is also highlighted in stick representation. Dotted lines connect the residues between which disulphide bonds are proposed to form, although the two N-terminal α-helices of TGFβ are spaced quite widely apart, and significant structural rearrangements would be necessary for disulphide formation between TGFβ C33 the LTBP1 TB2 2-6 cysteine pair.

interaction of LTBP1 TB2 with the TGF β propeptide [11]. A mutagenesis study has also implicated a number of basic amino acids in the propeptide of TGF β 1 as being important for complex formation with LTBP1 [95], which would be consistent with the involvement of acidic residues in the TB domain. However, in the recent crystal structure of TGF β 1 and its propeptide, these residues are instead implicated in interactions between the pro-peptide and growth factor region of TGF β (Figure 1.12). Although there are also a number of other basic residues in the TGF β propeptide that are closer to cysteine 33, and their importance in the LTBP interaction has not yet been tested.

Studies also suggested that the TB2 domain of LTBP4 has a lower affinity for binding to the TGF β 1 pro-peptide than LTBP 1 and 3, and LTBP4 appears not to bind the pro-peptides of TGF β 2 and 3 [11]. This is potentially due to the absence of several key acidic residues from this domain, and a “QQ” instead of an “FP” insert between cysteine 6 and 7. Interestingly it has also been shown that in cell culture the short isoform of LTBP4 (LTBP4S) is also mostly secreted without TGF β 1 bound, while the long isoform (LTBP4L) is secreted with TGF β 1 bound, but less readily associates with the ECM [133]. This presents the possibility that the LTBP4L N-terminus somehow acts as a long range enhancer of TGF β binding.

1.4.3 - Interaction of LTBPs with the ECM

LTBPs do more than just act as a chaperones for TGF β secretion, LTBPs are large ECM proteins (Figure 1.1), and only their second TB domain has been shown to interact with the SLC. Several other regions of the LTBPs have been demonstrated to interact with insoluble components of the ECM [134], specifically the C-termini of LTBP1, 2 and 4 have been shown to bind the N-terminus of fibrillin-1 non-covalently

[13-15], and the N-terminus of LTBP1, 2 and 4 has been shown to associate with fibronectin [135, 136]. Figure 1.13 summarises the known interactions of the LTBPs, and the proposed organisation of the LTBP1-LLC in the ECM.

The interaction with fibronectin has been described for the N-terminus of LTBP1 [135] and LTBP4 [136]. Co-localisation of LTBP1 with fibronectin is observed in immature fibroblast cultures [134], and has been shown to be essential for initial matrix incorporation of LTBP1. However in more established cultures, localisation of LTBP1 and Fibronectin diverge, suggesting fibronectin may act as a chaperone for the initial incorporation of LTBP1 into immature matrix [137]. LTBP4 and LTBP3 also exhibit different localisation patterns to fibronectin, and show slower matrix incorporation than LTBP1 [138]. The precise fibronectin binding site on LTBP1 has not been mapped, and some data suggest that heparan sulphate proteoglycans may be necessary as accessory factors to mediate the interaction [139], and established heparin-binding site is located in the hinge region of LTBP1 [139, 140]. LTBP2 also appears to bind fibronectin, and inhibits fibroblast adhesion and spreading on fibronectin substrates, but not on collagen type I substrates [141].

A section of sequence near the hinge region of LTBP1 has been shown to be essential for formation of transglutaminase cross links with the ECM [142]. Specifically cross-linking activity is lost when residues 293-441 are removed from the N-terminus, corresponding to the cbEGF1 and TB1 domains and the first part of the hinge region (Figure 1.13). However which ECM component is cross-linked to LTBP1 by transglutaminase has not been determined. In by pulse chase experiments it has also been observed that the LTBP1L isoform, with an extended N-terminus, binds to the ECM more efficiently [143], but the molecular basis for this is unknown.

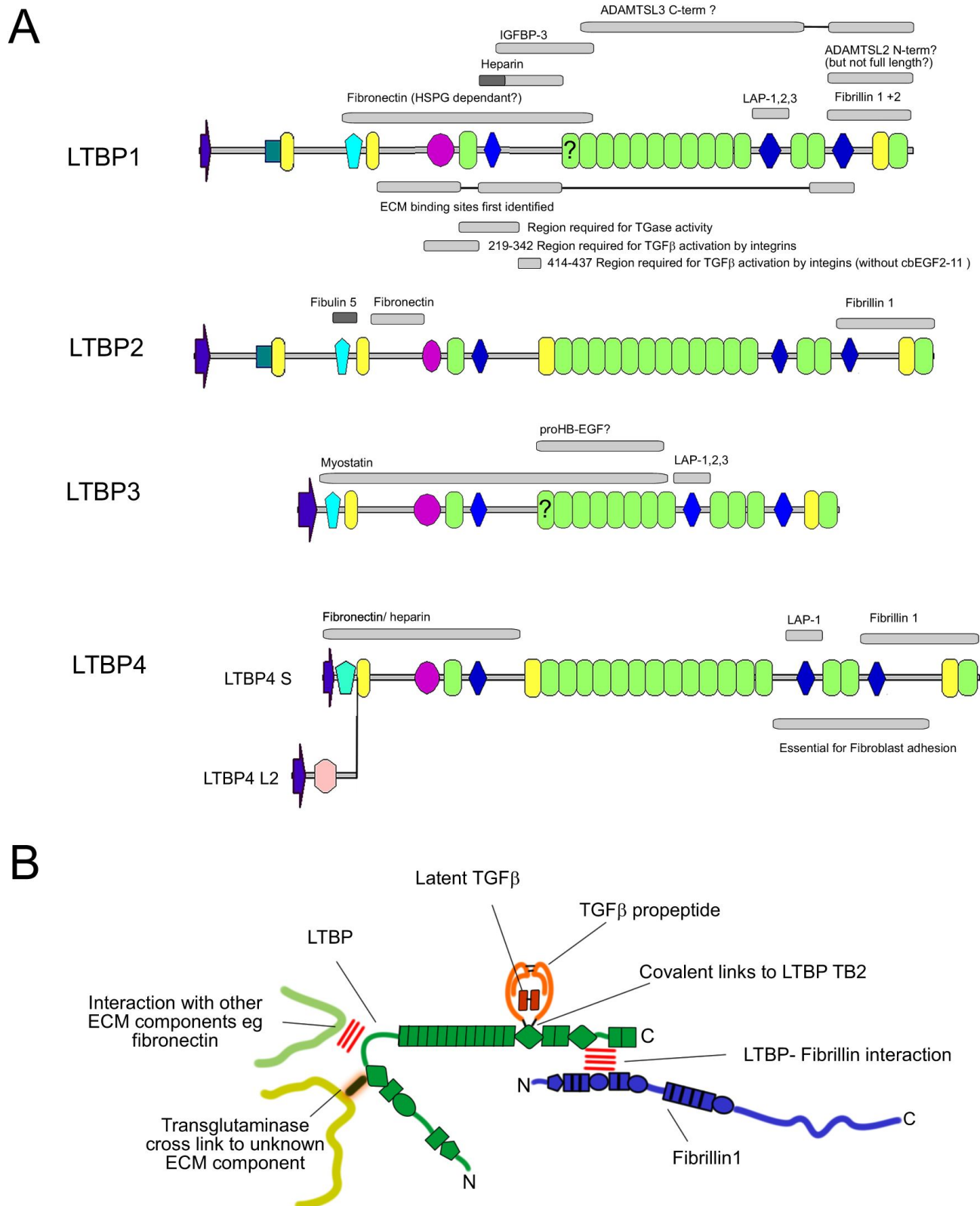


Figure 1.13 - Interactions of LTBP and the proposed organisation of LTBP1 within the ECM

A) Grey boxes denote LTBP interactions (coloured as in Figure 1.7). References: LTBP1-ADAMTSL3 C-term and ADAMTSL2 N-term [76] LTBP1-IGFBP3 [201], LTBP1-heparin+fibronectin [139], LTBP1-LAP [10, 11], LTBP1-Fibrillin1+2 [13-15], general ECM binding regions [134], TGase region [142], TGF β activating regions [9], LTBP2 - fibronectin [141], LTBP2 - fibulin 5 [202] LTBP2 - fibrillin 1 [144], LTBP3 - proHB-EGF (y2h) [203], LTBP3- Myostatin [204], LTBP3-LAP [10, 11], LTBP4-fibronectin and heparin, and fibroblasts [136], LTBP4- LAP [10, 11], LTBP4-Fibrillin1+2 [13-15]. **B)** The current model of LTBP organisation in the ECM. The LLC, consists of dimeric TGF β non-covalently bound to its propeptide, and the propeptide is covalently bound to LTBP via its second TB domain [2, 128, 205, 206]. The LLC itself is sequestered to the extracellular matrix, via interaction of its C-terminus with the N-terminus of fibrillin [13, 15], and interactions with other matrix components at its N-terminus [134], which may involve transglutaminase cross links [142].

1.4.4 - Interaction of LTBP3 with fibrillin

The interaction between LTBP3 and fibrillin is particularly interesting because it may also provide a mechanistic link between the fibrillin network and TGF β signalling. Initial studies of LTBP3-fibrillin interactions localized the binding regions to the N-terminal domains of fibrillin1, and C-terminal domains of LTBP1, 2 and 4 [13] [144]. The C-terminus of LTBP3 on the other hand did not interact with an N-terminal fragment of fibrillin1. Further studies have identified specific fibrillin domains that contribute to the interaction with LTBP1 and 4 [14, 15], and the fibrillin hyb1 domain in particular seems to be important for the interaction of fibrillin with both LTBP1 and 4, as a significant loss of binding is seen when this domain is deleted⁴. Additionally an amino acid substitution in the fibrillin1 EGF3 domain, of asparagine 164 to serine, (also associated with ectopia lentis), was shown to reduce binding to LTBP4 and the fibulins, but not to LTBP1. This suggests the involvement of more than one domain in the LTBP4 – fibrillin1 interaction, and that the precise binding sites of LTBP1 and 4 may be subtly different.

A recent study examining the matrix incorporation of various LTBPs in fibrillin null mice cells, has shown that LTBP3 and LTBP4 are both dependant on the presence of fibrillin1 for deposition in the ECM [145], as is LTBP2 [146] . LTBP1 on the other hand is still incorporated into the ECM in the absence of both fibrillin 1 and 2, and is instead dependant on a functional fibronectin network for its ECM deposition in cell culture.

It has also been shown that the assembly of a fibrillin network in culture is dependant on fibronectin, and so these results do not rule the possibility that fibrillin plays an important role in localising LTBP1 to ECM fibres, but it does demonstrate that for

⁴ The hyb1 domain also appears to be important for the interaction of fibrillin with fibulin 2, 4 and 5.

LTBP1 fibronectin is able to compensate for its absence in culture. It is also interesting that LTBP3 deposition is dependant on the presence of fibrillin1, as its C-terminus was demonstrated not to interact with the N-terminus of fibrillin. This could be due to interactions between alternative regions of either protein, or an indirect interaction of LTBP3 and fibrillin through an “adaptor” molecule, or that the LTBP3 fragment originally tested might have been misfolded.

LTBP 1, 2, and 4 all appear to bind to a particularly “busy” region of the fibrillin molecule (Figure 1.8), with many other proteins also interacting with this region. In some cases competitive binding has been demonstrated, [14, 147], which may have implications for TGF β regulation.

1.4.5 - Role of LTBP β s in TGF β activation

In addition to aiding secretion of TGF β , LTBP1 has also been demonstrated to play a role in the activation of TGF β 1 by specific integrins. A large number of different factors have been shown to activate TGF β (Figure 1.10), but activation of TGF β 1 by integrin α v β 6 is one of the best studied mechanisms [148], and its physiological relevance is backed up by *in vivo* studies [149, 150] (Section 1.5.3). In co-cultures of integrin α v β 6 and TGF β expressing cells it was shown that covalent attachment of TGF β 1 to LTBP1 was necessary for the release of active TGF β 1 in these cultures [9]. Additionally specific N-terminal regions of LTBP1 were also essential for this activation, corresponding to the region after the first hyb domain and including the “hinge” region, which has been shown to bind to the ECM (Figure 1.13). Fibronectin is believed to bind LTBP in this region, and it was demonstrated that assembly of a fibronectin network by these cells was also necessary for TGF β activation by α v β 6 [135]. The conclusion of these studies was that the key role of LTBP1, in this integrin mediated

TGF β activation assay, was anchoring TGF β to fibronectin fibres. Interestingly, a smaller region of the N-terminus of LTBP1 was needed for activation when the central array of cbEGF domains was not present (Figure 1.13), an observation that has not yet been explained.

The importance of LTBP1 in anchoring TGF β to an insoluble matrix for activation by integrins, led to a hypothesis where mechanical strain or traction between the matrix and the cell surface could be responsible for driving TGF β release (Figure 1.14). This is supported by observations of myofibroblast contraction in culture activating TGF β in an integrin dependant manner [151], and shear stress has also been shown to release active TGF β from platelets [152]. This traction mechanism may rely on TGF β being anchored to the ECM via LTBP, in order to exert the necessary force on the TGF β pro-peptide, distort its structure, and release active TGF β . The detailed unfolding process of the TGF β pro-peptide has been studied using AFM, and demonstrates discrete unfolding steps, which require forces that integrins should be capable of transmitting [153]. It is not clear if this is relevant to all TGF β activation, as TGF β 2 lacks the RGD sequence needed in TGF β 1 and TGF β 3 for integrin binding. Also it is not clear whether fibrillin may play a role in this integrin-mediated activation, perhaps by providing alternative ECM contacts to fibronectin.

1.5 - *In vivo* role of LTBP1 and TGF β

From the studies discussed in the previous section, it appears that LTBP1 can play multiple roles in TGF β signalling. However to assess the true physiological importance of these roles, the importance of LTBP1 *in vivo* must be examined.

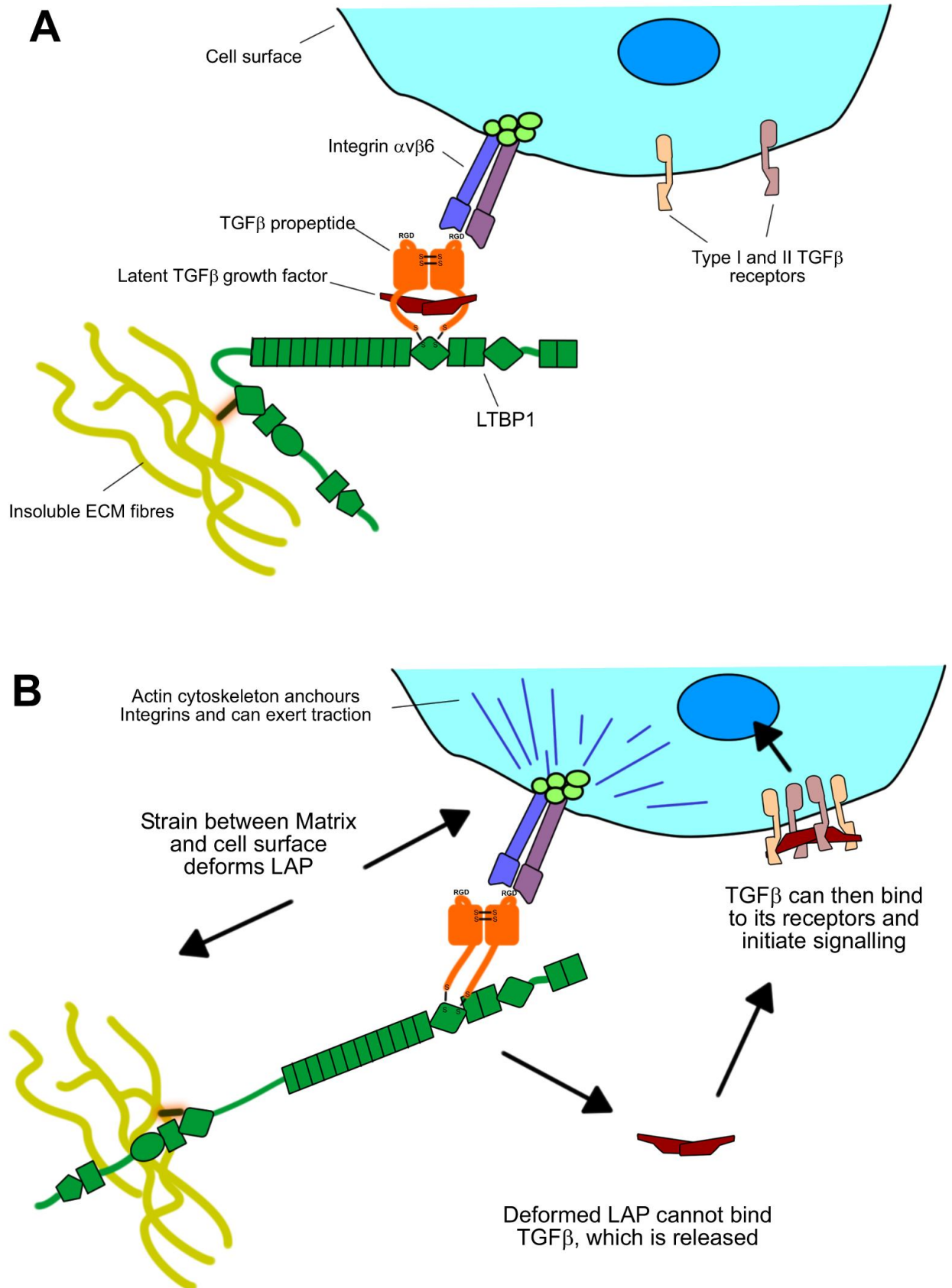


Figure 1.14 - Proposed mechanism for traction mediated activation of TGF β by integrin $\alpha\text{v}\beta 6$.

A) The proposed resting state for the LLC prior to activation by traction. LTBP is bound to the ECM and SLC, which can bind cell surface $\alpha\text{v}\beta 6$ integrins via its RGD motif. (It is possible that regulation of LAP proximity to the cell surface may represent an additional control point in this process.) **B)** Traction based activation; tensile force is applied to the LLC via movement of cell surface integrins, (bound to the actin cytoskeleton), relative to the insoluble ECM. Tensile force transferred to LAP, which is anchored to the ECM via LTBP, and is distorted into a conformation that is unable to bind TGF β . Active TGF β is consequently released to bind its cell surface receptors.

1.5.1 - LTBP2 in human disease

Unlike fibrillin there are only a few limited examples of LTBP2 mutations in human genetic disease. However the limited studies that have been carried out underscore the important physiological roles of these proteins, as well as giving insights into their function in human development (Table 1.1).

At present no *LTBP1* mutations have been associated with any human genetic disorder. Although up-regulation of LTBP1 has been associated with a number of acquired pathologies; such as idiopathic pulmonary fibrosis [154], and various cancers [155].

Recessive, loss of function, mutations in *LTBP2* have been associated with ocular defects, such as glaucoma [88-90]. A number of cases have been described but there seems to be some variety in the symptoms. Some papers report an association with primary congenital glaucoma [88, 89, 91], others with secondary glaucoma, and various other ocular features, such as megalocornea and spherophakia [90, 92], and one mutation has been linked to microspherophakia alone [87]. Although all these patients may represent a single disease spectrum.

Many mutations linked to both primary congenital glaucoma and secondary glaucoma introduce premature stop codons into the *LTBP2* sequence, via nonsense or frame shift changes, and so it is likely that they represent *LTBP2* null alleles, as the mRNA may be degraded by NMD. In the case of isolated microspherophakia, the associated *LTBP2* mutation causes a frame shift, replacing the last 6 amino acids, including the last cysteine of cbEGF16, with a stretch of 27 other amino acids, (including two new

cysteines). This would be predicted to disrupt the folding of the C-terminal cbEGF domain, but the functional consequences of this have not been investigated.

A recent report has also linked a novel *LTBP2* mutation to WMS in an Iranian family. Many of the affected patients have ocular defects, such as ectopia lentis and microspherophakia, but also shortened stature, brachydactyly (shortness of fingers and toes), joint stiffness, and pulmonary and aortic stenosis [156]. A shared mutation in this family causes a substitution of valine 1177 to methionine, which corresponds to the second residue of cbEGF9 in *LTBP2*. It is surprising that this substitution is deleterious, as this residue is not always conserved in other *LTBP2* sequences (replaced with a methionine in zebra finch *LTBP2*), and it is not a key structural residue, as many other cbEGFs possess a methionine in an identical position. Although this mutation was not found in 400 control individuals it remains possible that it represents a rare polymorphism in the Iranian population.

A recessive mutation in *LTBP3* has been associated with oligodontia; a condition where six or more teeth fail to develop [93]. Some scoliosis and short stature was also observed in the affected family (the tallest affected individual was 4ft 11), but limb ratios were normal, and bones generally looked normal, with some increased bone density restricted to the spine and base of skull. The mutation itself was identified as a nonsense mutation in the middle of the *LTBP3* gene (cbEGF5), and so likely represents a null allele, as with many of those described for *LTBP2*.

Recessive mutations in *LTBP4* have been shown to cause Urban-Rifkin-Davis syndrome (URDS), a severe disorder in which development of the pulmonary, gastrointestinal, genitourinary, musculoskeletal, craniofacial, and dermal systems is

disrupted [94]. Only four independent patients have been identified so far, of which three died from respiratory complications in the first two years of life. All patients displayed cutis laxa (loose hanging skin), microretrognathia (underdeveloped lower jaw bone), respiratory distress, (due to a number of complications in pulmonary development, including emphysema), joint laxity, low muscle tone, growth delay, along with gastrointestinal and genitourinary complications. Three of the four patients carried frame shift or nonsense mutations in LTBP4 (P264fsX, C857X, Q1185fsX, P1376fsX), while one had a missense mutation causing a substitution of a conserved cysteine 274 with glycine in the first hybrid domain of LTBP4, which may have the potential to cause domain misfolding and retention within the cell.

In fibroblast cultures from selected patients with frame shift and nonsense mutations, severely reduced expression of LTBP4 mRNA was observed, and LTBP4 protein was absent from the matrix of cultured fibroblasts. mRNA expression was improved with cyclohexamide, consistent with a nonsense mediated decay mechanism.

Studies on patient tissues showed reduced and fragmented elastic fibres in the dermis, and significantly enlarged saccular air spaces due to impaired alveolar septation, with a limited presence of elastic fibres. Electron Microscopy of these elastic fibres revealed a disruption in their ultrastructure, with the microfibril network supporting a diminished “moth eaten” elastin core, and globular deposits of elastin seen on the periphery of the elastic fibre, where normally it is contained entirely within the fibre core, and intermeshed with microfibrils. Additionally, increased levels of active TGF β were seen when patient fibroblasts were co-cultured with reporter cells, while TGF β 1 expression levels were similar to those seen in control fibroblasts. This suggests LTBP4 exerts an effect on TGF β signalling that is independent of TGF β gene

expression. Generally the severe phenotype of these patients highlights the crucial role of LTBP4 in development, which is consistent with its broad pattern of tissue expression [157, 158].

1.5.2 - LTBP mouse models

Further insights into the biological role of LTBPs comes from mouse models which have also been used to examine the roles of LTBP1 [159, 160], LTBP2 [161] [162], LTBP3 [163, 164], and LTBP4 [165, 166] (Appendix 3, Table A3.2).

No complete LTBP1 null mouse has been generated, but deletion of exons specific to the long isoform of LTBP1 is lethal due to defects in cardiac development. Specifically hypoplasia of the endocardial cushions early in embryonic valvulogenesis leads to valve hyper plasticity, and embryonic death for ~40% of homozygotes. Those LTBP1L mice which survive gestation die shortly after birth from defective development of the aortic arch and other cardiac defects [160, 167]. Lower levels of pSMAD are seen in specific tissues, such as the heart valves (but not the myocardium), suggesting that LTBP3 and 4 are unable to compensate for the role of LTBP1 in TGF β secretion and/or activation in these tissues.

Another study has deleted the first exon of the LTBP1 gene that is shared by both short and long isoforms (Figure 1.1), corresponding to the 4-cysteine domain unique to LTBPs [159]. However these mice are grossly normal, and only display a few mild defects, such as some shortening of the mandible and long bones. While both LTBP1 isoforms use this exon, it is possible that the LTBP1L isoform is able to splice across this missing exon and produce a functional product [162]. This mouse may then represent a specific knockout of this 4cys domain, or LTBP1S.

An early mouse model knocking out LTBP2 was reported not to be viable due to failure of embryo's to implant [161], but a new LTBP2 knock out mouse model has recently been generated that is viable, and displays ectopia lentis, but not the glaucoma observed in human patients [162] (full details of this model have yet to be published).

LTBP3 knock out mice show a variety of skeletal defects, including kyphosis and an osteopetrosis-like phenotype, whereby osteoclast function appears to be impaired, inhibiting bone resorption, and leading to increased bone density [163, 164, 168-170]. Reduced TGF β signalling is seen in these tissues, suggesting that LTBP3 has a non-redundant role in TGF β signalling in bone. When bred into an inbred background LTBP3 nulls also show spleen and thymus involution, and half of these mice die between weeks three and four, due to respiratory distress caused by defective alveolar septation. It is interesting that the LTBP3 null mice display contrasting phenotypes to the human LTBP3 patients with oligodontia, which may reflect different functional niches for the protein between the two species. Studies in zebra fish have also implicated an LTBP3 like protein as important in cardiac development [171], suggesting further evolutionary variation in the precise role of this protein.

Mice have also been generated with disrupted LTBP4 expression [166]. These mice were not specifically designed to knock out LTBP4, but were instead a product of a retroviral "gene trap" screen intended to target genes expressed early in development. The integration of the gene trap was shown to be intronic, and so the LTBP4 gene exons are still present. However northern blot and western blot analysis showed LTBP4 levels in homozygotes to be severely reduced, and so represent a near null

model⁵. These mice display multi organ defects, including colorectal cancers, rectal prolapse, and pulmonary emphysema all evident at birth, with cardiomyopathy developing later in life. They also show significant reductions in extracellular TGF β 1 deposition in the colon lung and heart, but not in the kidney where LTBP4 is only weakly expressed. During initial characterisation pSMAD levels were also shown to be decreased in lung and colon tissue, demonstrating LTBP4 is important for secretion and/or activation of TGF β 1, as well as ECM sequestration, in these tissues

However detailed study of the lungs of these LTBP4 disrupted mice suggest a more complex role for LTBP4 than just acting as a chaperone of TGF β secretion and/or activation [12]. The lungs of these mice display defective alveolar septation from embryonic day 18, and also show abnormal elastin deposition, with elastin aggregates evident by embryonic day 14, (before elastic fibres are even visible in normal lungs). This contrasts with LTBP3 $-/-$ mice, which display normal alveolar elastic fibres, but still display defective alveolar septation, albeit not as severely as LTBP4 $\sim/-$ mice. The ultra structure of LTBP4 $\sim/-$ fibres was also disrupted, and are identical to those seen in URDS patients, with dense globular deposits of elastin sitting on the periphery of microfibril bundles, rather than intermeshed within them. While initial characterisation of these mice had shown several indicators of decreased TGF β signalling, when the alveoli of 7 day old mice were examined, an increase in pSMAD2 was observed. More significantly, crosses of LTBP4 $\sim/+$ and TGF β 2 $-/+$ mice, showed normal embryonic lung septation was restored in LTBP4 $\sim/-$ TGF β 2 $-/-$ animals, suggesting a role for LTBP4 in suppressing specific TGF β signalling during septation. However these mice still display abnormal elastic fibres, and this result has lead to the suggestion that LTBP4 may have an additional, TGF β independent, role in elastic fibre assembly.

⁵ These will be referred to here as LTBP4 $\sim/-$ mice to distinguish them from complete knockout ($-/-$) mice.

These *in vivo* results all demonstrate significant and non-redundant physiological roles for all four of the LTBP. However it is possible that additional redundant roles exist, which are not demonstrated by single LTBP knock outs. Only redundancy between LTBP3 and LTBP4 has been investigated so far [172], and LTBP3 ^{-/-} LTBP4^{~/-} mice do not display any novel defects, aside from earlier lethality and exacerbated disruption of alveolar septation.

1.5.3 - The *in vivo* role of LTBPs in TGF β signalling

LTBP1L ^{-/-}, LTBP3 ^{-/-}, and LTBP4 ^{~/-} mice all, for the most part, display signs of reduced TGF β signalling, except in alveolar septation in LTBP4 ^{~/-} mice, where it appears increased TGF β 2 signalling may be responsible for the defect. URDS patient fibroblast cultures also demonstrate increased levels of active TGF β . However it is not clear from LTBP knock out studies alone precisely which roles of LTBP in TGF β signalling are physiologically significant. Cell culture studies have shown that LTBPs are important for correct folding and secretion of TGF β , and also for activation by specific integrins. However many factors beside integrins have been implicated as activating TGF β in culture (fig 1.9), and very recently “Glycoprotein-A repetitions predominant protein” (GARP) has been shown to covalently bind TGF β , aid its secretion from cells, and also enhance its activation by integrins⁶ [173]. With all the factors that could be involved in TGF β activation *in vivo*, it is important to specifically assess the physiological significance of LTBPs in TGF β signalling.

⁶ Unlike LTBPs GARP is a relatively small membrane bound protein, and its membrane association was also shown to be essential for GARP dependant TGF β activation by integrins.

Mice knockouts of individual TGF β isoforms demonstrate severe and distinct developmental defects; TGF β 1 knock out mice develop multi-organ inflammation [174], TGF β 2 mice demonstrate a range of skeletal, cardio-vascular, and other defects [175], while TGF β 3 knockouts display major defects in lung development, such as defective alveolar septation, and also cleft palate [176]. It is difficult to directly relate many of these phenotypes to those seen with individual LTBP null animals, and the fact that multiple TGF β isoforms can bind to multiple LTBP isoforms may provide a significant amount of redundancy.

The importance of ECM binding has however been investigated in the case of TGF β 1, where a mouse has been generated with a cysteine 33 to serine substitution⁷ in the TGF β pro-peptide. These mice only secreted LAP-TGF β that was not covalently bound to LTBP, but their cells still released TGF β at similar levels to WT cells [177]. These animals demonstrated a very similar phenotype to the TGF β 1 null mice, with multi-organ inflammation and lack of skin langerhans cells, but they do survive longer than their TGF β 1 null counterparts. This demonstrates that covalent association of TGF β with other factors (most likely LTBPs) plays a significant role in promoting TGF β 1 signalling *in vivo*.

Clues in to physiologically relevant mechanisms of TGF β activation come from another TGF β 1 knock in mouse, which has had its RGD motif replaced with an RGE motif to disrupt integrin binding [149]. This mouse phenocopies the TGF β 1 null mouse, suggesting that integrin binding is essential for all TGF β 1 activation, (at least where TGF β 1 has roles that are not redundant with other TGF β isoforms). Activation of

⁷ The residue responsible for covalent attachment to LTBP as previously discussed.

TGF β 1 and 3 by both integrins α v β 6 and α v β 8 has been demonstrated, and mice in which these integrins subunits are knocked out also demonstrate some similar phenotypes to the TGF β 1 and TGF β 3 null mice [150]. Since covalent attachment of TGF β 1 to LTBP1, and ECM localisation, have both been demonstrated as essential for integrin mediated TGF β 1 activation in cell culture [9, 135], it may be a reasonable assumption that LTBP1 localising TGF β to the ECM may also be important for TGF β activation by integrins *in vivo*, potentially by acting as anchors for traction mediated release of TGF β by integrins [135, 148, 151]. However this has not yet been tested, but together these *in vitro* and *in vivo* studies strongly suggest that LTBPs are important for both TGF β secretion and activation.

While the phenotype of the TGF β 1-RGE mouse implies integrins are essential for all non-redundant TGF β 1 activation, not all of this activation may be traction mediated or reliant on ECM anchoring of TGF β by LTBP. Incomplete reliance on traction is suggested by the less severe phenotype in the TGF β 1 C33S mice compared to TGF β 1 nulls and the RGE animals. Also some evidence for ECM independent LTBP mediated activation of TGF β 1/2/3 comes from a transgenic mouse engineered to express a cbEGF12-TB2-cbEGF13 fragment of LTBP1, under control of a keratin promoter [178]. This transgenic LTBP lacks all regions necessary for ECM localisation, and just retains the region able to bind latent TGF β . These mice have specific expression of transgenic protein in the basal epidermis and outer root sheet of the hair follicle. The mice display shorter hair, especially around P15, and histology of these mice showed a shorter period of hair shaft formation (anagen) and early induction of hair shaft involution (catagen). Some growth phase specific increases in nuclear pSMAD were seen, and TGF β has been implicated as a promoter of catagen in other

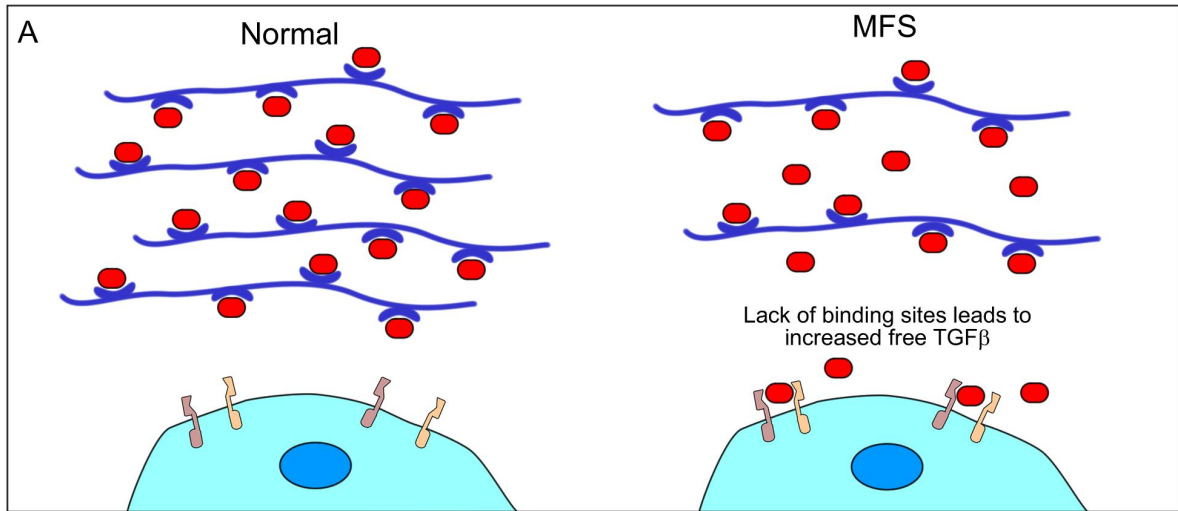
studies [179], suggesting expression of this truncated LTBP1 is increasing TGF β signalling in this case, perhaps by aiding its presentation for traction independent activation mechanisms, or by generally increasing TGF β 1 secretion from cells.

Together these studies of mouse models and human genetic disease highlight the important biological roles of the different LTBP isoforms in regulating TGF β secretion and activation. They also identify TGF β -independent activities, such as the putative role for LTBP4 in aiding elastic fibre assembly. However many questions remain over how the individual LTBPs perform these roles, and how their roles might contribute to pathologies such as Marfan syndrome and other fibrillinopathies, where TGF β is seen to be dysregulated.

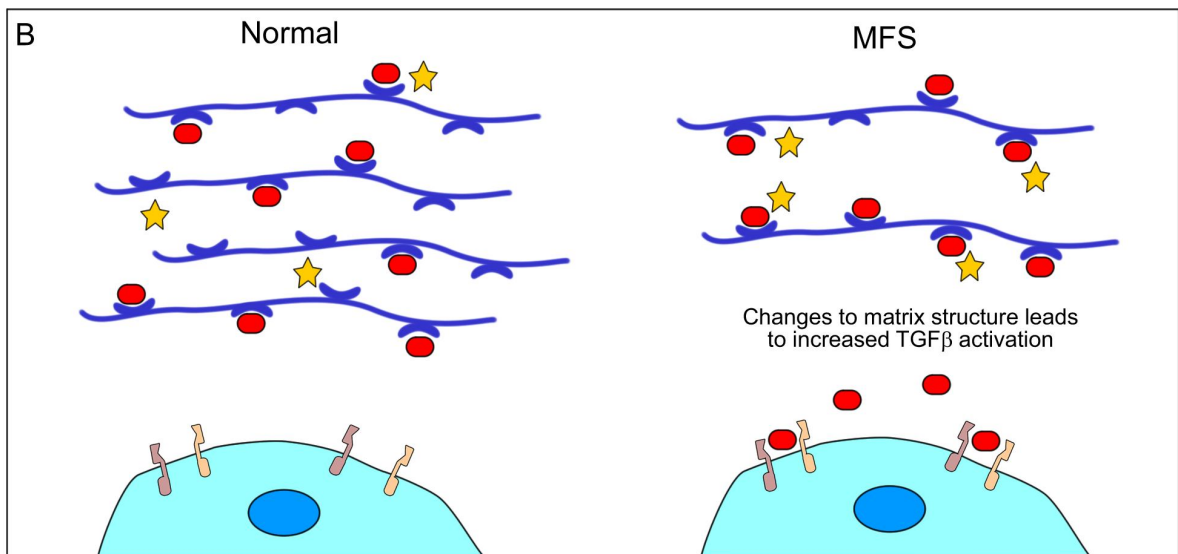
1.6 – Role of the fibrillin-LTBP interaction in TGF β signalling

As Marfan patients and MFS mice show increased TGF β signalling in numerous tissues [60,209], the initial observation of an interaction between LTBPs and the fibrillin matrix suggested a model where sequestration of TGF β to fibrillin microfibrils may be important for maintaining TGF β latency (Figure 1.15 A). In MFS, it was then proposed that the increase in TGF β signalling observed was due to the loss of a critical number of LTBP binding sites in the fibrillin microfibril network, which then led to insufficient latency of TGF β , and to excess active TGF β observed. Some support for this model comes from competition studies, where other proteins, such as MAGP1 and other fibrillin1 fragments [14, 147], that compete for the LTBP binding site on fibrillin, were shown to increase pSMAD2/3 levels in cell culture assays.

Latency perturbation model



Activation perturbation model



Localisation perturbation model

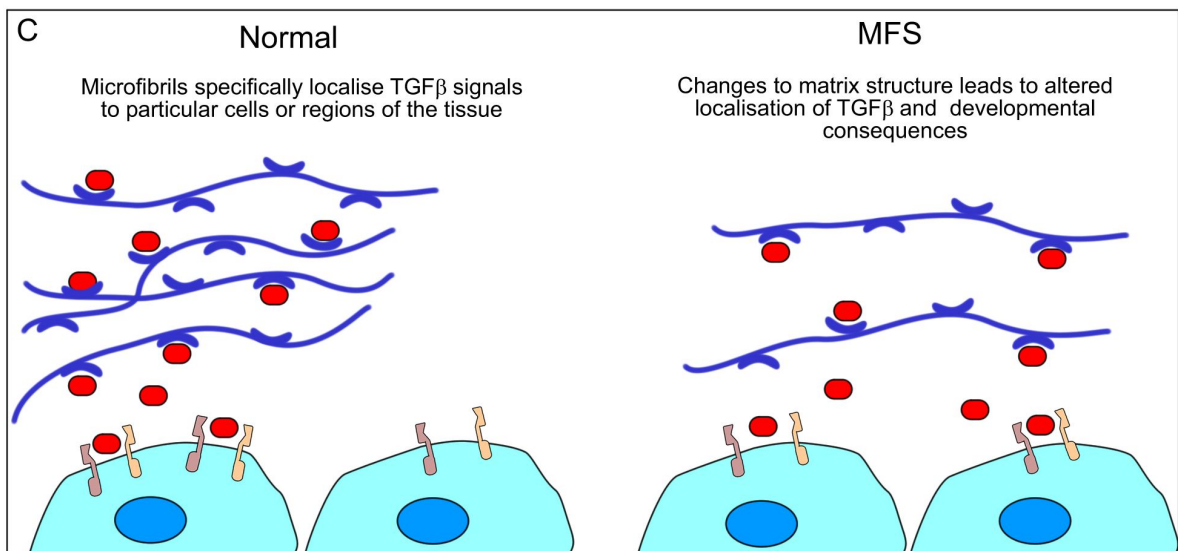


Figure 1.15 - Models for the dysregulation of TGFβ in Marfan syndrome

TGFβ is represented as red ellipsoids, fibrillin as blue lines (LTBPs excluded for simplicity). Cells with TGFβ receptors are also shown in blue. **A)** Latency perturbation model, where in MFS causes a shortage of binding sites for latent TGFβ. **B)** Activation perturbation model, where the reduction in fibrillin fibres leads to an increase in specific TGFβ activation (represented by yellow stars) perhaps via altered concentration of TGFβ activators, or altered matrix biomechanics. **C)** Localisation perturbation model, where changes in matrix structure affect the targeting of TGFβ signals.

However this latency model does not address why TGF β immobilisation to ECM fibres is necessary for inhibition of signalling, when the soluble complex of TGF β bound LAP is equally unable to bind its receptor. It is possible that ECM localization may protect against TGF β activation by specific factors, but this has not yet been demonstrated. Generally the diverse genetic diseases affecting fibrillin1 and TGF β levels, such as MFS and AD/GD, which both cause increases in TGF β activation but result in contrasting phenotypes, suggests that the localisation model for fibrillin regulation of TGF β is over simplistic.

Fibrillin may in fact have a more complex role than just maintaining the general latency of TGF β . For example one hypothesis could be that fibrillin plays a direct role in modulating TGF β activation (Figure 1.15 B), and disruption of the fibrillin network in MFS may directly alter fibrillin-dependant activation mechanisms. Alternatively microfibril structures may localise TGF β signals to specific cells and regions of tissue, which may, in turn, be responsible for establishing important patterning signals in tissue development (Figure 1.15C). Significant further research is needed to understand the mechanisms connecting fibrillin and TGF β , and a structural understanding of the protein interactions involved has the potential to be highly valuable, especially when attempting to build a detailed mechanistic picture of the function of the fibrillin-LTBP-TGF β complex.

1.7 – Thesis aims

Fibrillins and LTBP have important roles in the structure of the ECM, and also in cell signalling through regulation of specific growth factors. However many questions still remain over the detailed molecular mechanisms by which they perform both structural

and regulatory roles. This thesis examines the molecular basis of the fibrillin1 – LTBP1 interaction, initially using bioinformatics techniques to investigate the emergence of the fibrillins and LTBP1s and look at specific sequence conservation. Limited fragments of LTBP1 and fibrillin1 have then been expressed and purified, and various biochemical and biophysical techniques employed to investigate their structure and dynamics. SPR studies and NMR titrations have also been used to determine the precise binding sites involved in the LTBP1 – fibrillin1 interaction, and to generate structural models for the LTBP1 – fibrillin1 complex. The validity of this interaction model has then been tested by site-directed mutagenesis of putative salt bridges, highlighting a single consistent structural model for the LTBP1 – fibrillin1 interaction.

References

1. Robertson, I., S. Jensen, and P. Handford, *TB domain proteins: evolutionary insights into the multifaceted roles of fibrillins and LTBP1s*. Biochem J, 2011. **433**(2): p. 263-76.
2. Hyytiäinen, M., C. Penttinen, and J. Keski-Oja, *Latent TGF-beta binding proteins: extracellular matrix association and roles in TGF-beta activation*. Crit Rev Clin Lab Sci, 2004. **41**(3): p. 233-64.
3. Saharinen, J., et al., *Latent transforming growth factor-beta binding proteins (LTBP1s)--structural extracellular matrix proteins for targeting TGF-beta action*. Cytokine Growth Factor Rev, 1999. **10**(2): p. 99-117.
4. Keene, D.R., et al., *Extraction of extendable beaded structures and their identification as fibrillin-containing extracellular matrix microfibrils*. J Histochem Cytochem, 1991. **39**(4): p. 441-9.
5. Sakai, L.Y., et al., *Purification and partial characterization of fibrillin, a cysteine-rich structural component of connective tissue microfibrils*. J Biol Chem, 1991. **266**(22): p. 14763-70.
6. Ramirez, F. and D.B. Rifkin, *Extracellular microfibrils: contextual platforms for TGFbeta and BMP signaling*. Curr Opin Cell Biol, 2009. **21**(5): p. 616-22.
7. Miyazono, K., et al., *Latent high molecular weight complex of transforming growth factor beta 1. Purification from human platelets and structural characterization*. J Biol Chem, 1988. **263**(13): p. 6407-15.
8. Miyazono, K., et al., *A role of the latent TGF-beta 1-binding protein in the assembly and secretion of TGF-beta 1*. EMBO J, 1991. **10**(5): p. 1091-101.
9. Annes, J.P., et al., *Integrin alphaVbeta6-mediated activation of latent TGF-beta requires the latent TGF-beta binding protein-1*. J Cell Biol, 2004. **165**(5): p. 723-34.
10. Saharinen, J. and J. Keski-Oja, *Specific sequence motif of 8-Cys repeats of TGF-beta binding proteins, LTBP1s, creates a hydrophobic interaction surface for binding of small latent TGF-beta*. Mol Biol Cell, 2000. **11**(8): p. 2691-704.
11. Chen, Y., et al., *Amino acid requirements for formation of the TGF-beta-latent TGF-beta binding protein complexes*. J Mol Biol, 2005. **345**(1): p. 175-86.

12. Dabovic, B., et al., *Dual functions for LTBP in lung development: LTBP-4 independently modulates elastogenesis and TGF-beta activity*. J Cell Physiol, 2009. **219**(1): p. 14-22.
13. Isogai, Z., et al., *Latent transforming growth factor beta-binding protein 1 interacts with fibrillin and is a microfibril-associated protein*. J Biol Chem, 2003. **278**(4): p. 2750-7.
14. Massam-Wu, T., et al., *Assembly of fibrillin microfibrils governs extracellular deposition of latent TGF beta*. J Cell Sci. **123**(Pt 17): p. 3006-18.
15. Ono, R.N., et al., *Latent transforming growth factor beta-binding proteins and fibulins compete for fibrillin-1 and exhibit exquisite specificities in binding sites*. J Biol Chem, 2009. **284**(25): p. 16872-81.
16. Jensen, S.A., et al., *Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins*. Structure, 2009. **17**(5): p. 759-68.
17. Lee, S.S., et al., *Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization*. Structure, 2004. **12**(4): p. 717-29.
18. Yuan, X., et al., *Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils*. EMBO J, 1997. **16**(22): p. 6659-66.
19. Lack, J., et al., *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, 2003. **334**(2): p. 281-91.
20. Smallridge, R.S., et al., *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, 2003. **278**(14): p. 12199-206.
21. Downing, A.K., et al., *Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders*. Cell, 1996. **85**(4): p. 597-605.
22. Jensen, S.A., et al., *Ca²⁺-dependent interface formation in fibrillin-1*. J Biol Chem, 2005. **280**(14): p. 14076-84.
23. Jovanovic, J., et al., *Fibrillin-integrin interactions in health and disease*. Biochem Soc Trans, 2008. **36**(Pt 2): p. 257-62.
24. Bax, D.V., et al., *Cell adhesion to fibrillin-1: identification of an Arg-Gly-Asp-dependent synergy region and a heparin-binding site that regulates focal adhesion formation*. J Cell Sci, 2007. **120**(Pt 8): p. 1383-92.
25. Jovanovic, J., et al., *alphaVbeta6 is a novel receptor for human fibrillin-1. Comparative studies of molecular determinants underlying integrin-rgd affinity and specificity*. J Biol Chem, 2007. **282**(9): p. 6743-51.
26. Pfaff, M., et al., *Cell adhesion and integrin binding to recombinant human fibrillin-1*. FEBS Lett, 1996. **384**(3): p. 247-50.
27. Sakamoto, H., et al., *Cell-type specific recognition of RGD- and non-RGD-containing cell binding domains in fibrillin-1*. J Biol Chem, 1996. **271**(9): p. 4916-22.
28. Loeys, B.L., et al., *Mutations in fibrillin-1 cause congenital scleroderma: stiff skin syndrome*. Sci Transl Med, 2010. **2**(23): p. 23ra20.
29. Gleizes, P.E., et al., *Identification and characterization of an eight-cysteine repeat of the latent transforming growth factor-beta binding protein-1 that mediates bonding to the latent transforming growth factor-beta1*. J Biol Chem, 1996. **271**(47): p. 29891-6.
30. Ogiso, H., et al., *Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains*. Cell, 2002. **110**(6): p. 775-87.
31. Lu, Y., et al., *Tissue specific differences in fibrillin microfibrils analysed using single particle image analysis*. J Struct Biol, 2006. **155**(2): p. 285-93.
32. Handford, P.A., *Fibrillin-1, a calcium binding protein of extracellular matrix*. Biochim Biophys Acta, 2000. **1498**(2-3): p. 84-90.
33. Hartner, A., et al., *Characterization of the renal phenotype in a mouse model of Marfan syndrome*. Virchows Arch, 2004. **445**(4): p. 382-8.
34. Sawada, T., *Ultrastructural localization of fibrillin-1 and fibrillin-2 in oxytalan fibers in periodontal ligament of Japanese Macaca fuscata monkey*. J Mol Histol. **41**(4-5): p. 225-31.
35. Saito, M. and T. Tsuji, *[Molecular mechanisms for the improvement of wound healing ability of periodontal ligament in Marfan's syndrome]*. Clin Calcium. **22**(1): p. 35-42.
36. Ritty, T.M., et al., *Processing of the fibrillin-1 carboxyl-terminal domain*. J Biol Chem, 1999. **274**(13): p. 8933-40.

37. Raghunath, M., et al., *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, 1999. **112 (Pt 7)**: p. 1093-100.
38. Ashworth, J.L., et al., *Regulation of fibrillin carboxy-terminal furin processing by N-glycosylation, and association of amino- and carboxy-terminal sequences*. J Cell Sci, 1999. **112 (Pt 22)**: p. 4163-71.
39. Wallis, D.D., et al., *Profibrillin-1 maturation by human dermal fibroblasts: proteolytic processing and molecular chaperones*. J Cell Biochem, 2003. **90**(3): p. 641-52.
40. Handford, P.A., et al., *Fibrillin: from domain structure to supramolecular assembly*. Matrix Biol, 2000. **19**(6): p. 457-70.
41. Baldock, C., et al., *Nanostructure of fibrillin-1 reveals compact conformation of EGF arrays and mechanism for extensibility*. Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11922-7.
42. Kielty, C.M., et al., *Fibrillin: from microfibril assembly to biomechanical function*. Philos Trans R Soc Lond B Biol Sci, 2002. **357**(1418): p. 207-17.
43. Jensen, S.A., I.B. Robertson, and P.A. Handford, *Dissecting the fibrillin microfibril: structural insights into organization and function*. Structure, 2012. **20**(2): p. 215-25.
44. Lin, G., et al., *Homo- and heterotypic fibrillin-1 and -2 interactions constitute the basis for the assembly of microfibrils*. J Biol Chem, 2002. **277**(52): p. 50795-804.
45. Marson, A., et al., *Homotypic fibrillin-1 interactions in microfibril assembly*. J Biol Chem, 2005. **280**(6): p. 5013-21.
46. Hubmacher, D., et al., *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, 2008. **105**(18): p. 6548-53.
47. Ashworth, J.L., et al., *Fibrillin assembly: dimer formation mediated by amino-terminal sequences*. J Cell Sci, 1999. **112 (Pt 20)**: p. 3549-58.
48. Trask, T.M., et al., *N-terminal domains of fibrillin 1 and fibrillin 2 direct the formation of homodimers: a possible first step in microfibril assembly*. Biochem J, 1999. **340 (Pt 3)**: p. 693-701.
49. Reinhardt, D.P., et al., *Initial steps in assembly of microfibrils. Formation of disulfide-cross-linked multimers containing fibrillin-1*. J Biol Chem, 2000. **275**(3): p. 2205-10.
50. Qian, R.Q. and R.W. Glanville, *Alignment of fibrillin molecules in elastic microfibrils is defined by transglutaminase-derived cross-links*. Biochemistry, 1997. **36**(50): p. 15841-7.
51. Sabatier, L., et al., *Fibrillin assembly requires fibronectin*. Mol Biol Cell, 2009. **20**(3): p. 846-58.
52. Tsuruga, E., et al., *Integrin alphavbeta3 regulates microfibril assembly in human periodontal ligament cells*. Tissue Cell, 2009. **41**(2): p. 85-9.
53. Dirk Hubmacher, D.P.R., *Microfibrils and Fibrillin*, in *The Extracellular Matrix: an Overview*. 2011, Springer-Verlag Berlin Heidelberg.
54. Cain, S.A., et al., *Fibrillin-1 interactions with heparin. Implications for microfibril and elastic fiber assembly*. J Biol Chem, 2005. **280**(34): p. 30526-37.
55. Tsutsui, K., et al., *ADAMTSL-6 is a novel extracellular matrix protein that binds to fibrillin-1 and promotes fibrillin-1 fibril formation*. J Biol Chem, 2011. **285**(7): p. 4870-82.
56. Saito, M., et al., *ADAMTSL6beta protein rescues fibrillin-1 microfibril disorder in a Marfan syndrome mouse model through the promotion of fibrillin-1 assembly*. J Biol Chem, 2011. **286**(44): p. 38602-13.
57. Gabriel, L.A., et al., *ADAMTSL4, a secreted glycoprotein widely distributed in the eye, binds fibrillin-1 microfibrils and accelerates microfibril biogenesis*. Invest Ophthalmol Vis Sci, 2012. **53**(1): p. 461-9.
58. Ramirez, F. and L.Y. Sakai, *Biogenesis and function of fibrillin assemblies*. Cell Tissue Res. **339**(1): p. 71-82.
59. Ramirez, F., et al., *Extracellular microfibrils in development and disease*. Cell Mol Life Sci, 2007. **64**(18): p. 2437-46.
60. Matt, P., et al., *Circulating transforming growth factor-beta in Marfan syndrome*. Circulation, 2009. **120**(6): p. 526-32.
61. Loeys, B.L., et al., *A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in TGFBR1 or TGFBR2*. Nat Genet, 2005. **37**(3): p. 275-81.
62. Robinson, P.N., et al., *The molecular genetics of Marfan syndrome and related disorders*. J Med Genet, 2006. **43**(10): p. 769-87.

63. Robinson, P.N. and M. Godfrey, *The molecular genetics of Marfan syndrome and related microfibrilopathies*. J Med Genet, 2000. **37**(1): p. 9-25.
64. Dietz, H.C., et al., *Recent progress towards a molecular understanding of Marfan syndrome*. Am J Med Genet C Semin Med Genet, 2005. **139C**(1): p. 4-9.
65. Stheneur, C., et al., *Prognosis factors in probands with an FBN1 mutation diagnosed before the age of 1 year*. Pediatr Res. **69**(3): p. 265-70.
66. Faivre, L., et al., *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, 2007. **81**(3): p. 454-66.
67. Tiecke, F., et al., *Classic, atypically severe and neonatal Marfan syndrome: twelve mutations and genotype-phenotype correlations in FBN1 exons 24-40*. Eur J Hum Genet, 2001. **9**(1): p. 13-21.
68. Hatchwell, E., *Shprintzen-Goldberg syndrome results from mutations in fibrillin-1, not monosomy 22q11*. J Pediatr, 1997. **131**(1 Pt 1): p. 164-5.
69. Ades, L.C., et al., *Segregation of a novel FBN1 gene mutation, G1796E, with kyphoscoliosis and radiographic evidence of vertebral dysplasia in three generations*. Am J Med Genet, 2002. **109**(4): p. 261-70.
70. Ahram, D., et al., *A homozygous mutation in ADAMTSL4 causes autosomal-recessive isolated ectopia lentis*. Am J Hum Genet, 2009. **84**(2): p. 274-8.
71. Comeglio, P., et al., *Identification of FBN1 gene mutations in patients with ectopia lentis and marfanoid habitus*. Br J Ophthalmol, 2002. **86**(12): p. 1359-62.
72. Hasham, S.N., D.C. Guo, and D.M. Milewicz, *Genetic basis of thoracic aortic aneurysms and dissections*. Curr Opin Cardiol, 2002. **17**(6): p. 677-83.
73. Dietz, H.C., et al., *Four novel FBN1 mutations: significance for mutant transcript level and EGF-like domain calcium binding in the pathogenesis of Marfan syndrome*. Genomics, 1993. **17**(2): p. 468-75.
74. Morales, J., et al., *Homozygous mutations in ADAMTS10 and ADAMTS17 cause lenticular myopia, ectopia lentis, glaucoma, spherophakia, and short stature*. Am J Hum Genet, 2009. **85**(5): p. 558-68.
75. Haji-Seyed-Javadi, R., et al., *LTBP2 mutations cause Weill-Marchesani and Weill-Marchesani-like syndrome and affect disruptions in the extracellular matrix*. Hum Mutat. **33**(12): p. 1985-94.
76. Sengle, G., et al., *Microenvironmental regulation by fibrillin-1*. PLoS Genet, 2012. **8**(1): p. e1002425.
77. Faivre, L., et al., *In frame fibrillin-1 gene deletion in autosomal dominant Weill-Marchesani syndrome*. J Med Genet, 2003. **40**(1): p. 34-6.
78. Le Goff, C., et al., *Mutations in the TGFbeta binding-protein-like domain 5 of FBN1 are responsible for acromicric and geleophysic dysplasias*. Am J Hum Genet, 2011. **89**(1): p. 7-14.
79. Le Goff, C., et al., *ADAMTSL2 mutations in geleophysic dysplasia demonstrate a role for ADAMTS-like proteins in TGF-beta bioavailability regulation*. Nat Genet, 2008. **40**(9): p. 1119-23.
80. Goldblatt, J., et al., *Further evidence for a marfanoid syndrome with neonatal progeroid features and severe generalized lipodystrophy due to frameshift mutations near the 3' end of the FBN1 gene*. Am J Med Genet A. **155A**(4): p. 717-20.
81. Graul-Neumann, L.M., et al., *Marfan syndrome with neonatal progeroid syndrome-like lipodystrophy associated with a novel frameshift mutation at the 3' terminus of the FBN1-gene*. Am J Med Genet A. **152A**(11): p. 2749-55.
82. Horn, D. and P.N. Robinson, *Progeroid facial features and lipodystrophy associated with a novel splice site mutation in the final intron of the FBN1 gene*. Am J Med Genet A. **155A**(4): p. 721-4.
83. Jordan, C.D., et al., *Fibrillins in adult human ovary and polycystic ovary syndrome: is fibrillin-3 affected in PCOS?* J Histochem Cytochem. **58**(10): p. 903-15.
84. Segars, J.H. and A.H. DeCherney, *Is there a genetic basis for polycystic ovary syndrome?* J Clin Endocrinol Metab. **95**(5): p. 2058-60.
85. Raja-Khan, N., et al., *A variant in the fibrillin-3 gene is associated with TGF-beta and inhibin B levels in women with polycystic ovary syndrome*. Fertil Steril. **92**(5): p. 1455-61.
86. Jordan, C.D., et al., *Fibrillins in Human Adult Ovary and Polycystic Ovary Syndrome: Is Fibrillin-3 Affected in PCOS?* J Histochem Cytochem. **58**(10): p. 903-15.
87. Kumar, A., et al., *A homozygous mutation in LTBP2 causes isolated microspherophakia*. Hum Genet. **128**(4): p. 365-71.

88. Ali, M., et al., *Null mutations in LTBP2 cause primary congenital glaucoma*. Am J Hum Genet, 2009. **84**(5): p. 664-71.
89. Narooie-Nejad, M., et al., *Loss of function mutations in the gene encoding latent transforming growth factor beta binding protein 2, LTBP2, cause primary congenital glaucoma*. Hum Mol Genet, 2009. **18**(20): p. 3969-77.
90. Desir, J., et al., *LTBP2 null mutations in an autosomal recessive ocular syndrome with megalocornea, spherophakia, and secondary glaucoma*. Eur J Hum Genet, 2010.
91. Azmanov, D.N., et al., *LTBP2 and CYP1B1 mutations and associated ocular phenotypes in the Roma/Gypsy founder population*. Eur J Hum Genet, 2010. **19**(3): p. 326-33.
92. Khan, A.O., M.A. Aldahmesh, and F.S. Alkuraya, *Congenital megalocornea with zonular weakness and childhood lens-related secondary glaucoma - a distinct phenotype caused by recessive LTBP2 mutations*. Mol Vis. **17**: p. 2570-9.
93. Noor, A., et al., *Oligodontia is caused by mutation in LTBP3, the gene encoding latent TGF-beta binding protein 3*. Am J Hum Genet, 2009. **84**(4): p. 519-23.
94. Urban, Z., et al., *Mutations in LTBP4 Cause a Syndrome of Impaired Pulmonary, Gastrointestinal, Genitourinary, Musculoskeletal, and Dermal Development*. Am J Hum Genet, 2009.
95. Walton, K.L., et al., *Two distinct regions of latency-associated peptide coordinate stability of the latent transforming growth factor-beta1 complex*. J Biol Chem, 2010. **285**(22): p. 17029-37.
96. Shi, M., et al., *Latent TGF-beta structure and activation*. Nature, 2011. **474**(7351): p. 343-9.
97. Janssens, K., et al., *Camurati-Engelmann disease: review of the clinical, radiological, and molecular data of 24 families and implications for diagnosis and treatment*. J Med Genet, 2006. **43**(1): p. 1-11.
98. Janssens, K., et al., *Mutations in the gene encoding the latency-associated peptide of TGF-beta 1 cause Camurati-Engelmann disease*. Nat Genet, 2000. **26**(3): p. 273-5.
99. Kinoshita, A., et al., *Domain-specific mutations in TGFB1 result in Camurati-Engelmann disease*. Nat Genet, 2000. **26**(1): p. 19-20.
100. Beffagna, G., et al., *Regulatory mutations in transforming growth factor-beta3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1*. Cardiovasc Res, 2005. **65**(2): p. 366-73.
101. Cardoso, S., S.P. Robertson, and P.B. Daniel, *TGFB1 mutations associated with Loeys-Dietz syndrome are inactivating*. J Recept Signal Transduct Res, 2012.
102. Attias, D., et al., *Comparison of clinical presentations and outcomes between patients with TGFB2 and FBN1 mutations in Marfan syndrome and related disorders*. Circulation, 2009. **120**(25): p. 2541-9.
103. Goudie, D.R., et al., *Multiple self-healing squamous epithelioma is caused by a disease-specific spectrum of mutations in TGFB1*. Nat Genet, 2011. **43**(4): p. 365-9.
104. Coucke, P.J., et al., *Mutations in the facilitative glucose transporter GLUT10 alter angiogenesis and cause arterial tortuosity syndrome*. Nat Genet, 2006. **38**(4): p. 452-7.
105. Lee, B., et al., *Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes*. Nature, 1991. **352**(6333): p. 330-4.
106. Dietz, H.C. and R.E. Pyeritz, *Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders*. Hum Mol Genet, 1995. **4 Spec No**: p. 1799-809.
107. Hilhorst-Hofstee, Y., et al., *The clinical spectrum of complete FBN1 allele deletions*. Eur J Hum Genet, 2010. **19**(3): p. 247-52.
108. Judge, D.P., et al., *Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome*. J Clin Invest, 2004. **114**(2): p. 172-81.
109. Schrijver, I., et al., *Premature termination mutations in FBN1: distinct effects on differential allelic expression and on protein and clinical phenotypes*. Am J Hum Genet, 2002. **71**(2): p. 223-37.
110. Nijbroek, G., et al., *Fifteen novel FBN1 mutations causing Marfan syndrome detected by heteroduplex analysis of genomic amplicons*. Am J Hum Genet, 1995. **57**(1): p. 8-21.
111. Hutchinson, S., et al., *Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype?* Hum Mol Genet, 2003. **12**(18): p. 2269-76.
112. Whiteman, P., et al., *Cellular and molecular studies of Marfan syndrome mutations identify co-operative protein folding in the cbEGF12-13 region of fibrillin-1*. Hum Mol Genet, 2007. **16**(8): p. 907-18.

113. Whiteman, P. and P.A. Handford, *Defective secretion of recombinant fragments of fibrillin-1: implications of protein misfolding for the pathogenesis of Marfan syndrome and related disorders*. Hum Mol Genet, 2003. **12**(7): p. 727-37.
114. Godfrey, M., et al., *Cosegregation of elastin-associated microfibrillar abnormalities with the Marfan phenotype in families*. Am J Hum Genet, 1990. **46**(4): p. 652-60.
115. Godfrey, M., et al., *Unilateral microfibrillar abnormalities in a case of asymmetric Marfan syndrome*. Am J Hum Genet, 1990. **46**(4): p. 661-71.
116. Hollister, D.W., et al., *Immunohistologic abnormalities of the microfibrillar-fiber system in the Marfan syndrome*. N Engl J Med, 1990. **323**(3): p. 152-9.
117. Pereira, L., et al., *Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1*. Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3819-23.
118. Pereira, L., et al., *Targetting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome*. Nat Genet, 1997. **17**(2): p. 218-22.
119. Carta, L., et al., *Fibrillins 1 and 2 perform partially overlapping functions during aortic development*. J Biol Chem, 2006. **281**(12): p. 8016-23.
120. Ramirez, F., et al., *Fibrillin microfibrils: multipurpose extracellular networks in organismal physiology*. Physiol Genomics, 2004. **19**(2): p. 151-4.
121. Arteaga-Solis, E., et al., *Regulation of limb patterning by extracellular microfibrils*. J Cell Biol, 2001. **154**(2): p. 275-81.
122. Ng, C.M., et al., *TGF-beta-dependent pathogenesis of mitral valve prolapse in a mouse model of Marfan syndrome*. J Clin Invest, 2004. **114**(11): p. 1586-92.
123. Cohn, R.D., et al., *Angiotensin II type 1 receptor blockade attenuates TGF-beta-induced failure of muscle regeneration in multiple myopathic states*. Nat Med, 2007. **13**(2): p. 204-10.
124. Habashi, J.P., et al., *Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome*. Science, 2006. **312**(5770): p. 117-21.
125. Habashi, J.P., et al., *Angiotensin II type 2 receptor signaling attenuates aortic aneurysm in mice through ERK antagonism*. Science. 2011. **332**(6027): p. 361-5.
126. Siracusa, L.D., et al., *A tandem duplication within the fibrillin 1 gene is associated with the mouse tight skin mutation*. Genome Res, 1996. **6**(4): p. 300-13.
127. De Backer, J., et al., *Utility of molecular analyses in the exploration of extreme intrafamilial variability in the Marfan syndrome*. Clin Genet, 2007. **72**(3): p. 188-98.
128. Annes, J.P., J.S. Munger, and D.B. Rifkin, *Making sense of latent TGFbeta activation*. J Cell Sci, 2003. **116**(Pt 2): p. 217-24.
129. Herpin, A., C. Lelong, and P. Favrel, *Transforming growth factor-beta-related proteins: an ancestral and widespread superfamily of cytokines in metazoans*. Dev Comp Immunol, 2004. **28**(5): p. 461-85.
130. Kahlem, P. and S.J. Newfeld, *Informatics approaches to understanding TGFbeta pathway regulation*. Development, 2009. **136**(22): p. 3729-40.
131. Vitt, U.A., S.Y. Hsu, and A.J. Hsueh, *Evolution and classification of cystine knot-containing hormones and related extracellular signaling molecules*. Mol Endocrinol, 2001. **15**(5): p. 681-94.
132. Penttinen, C., et al., *Secretion of human latent TGF-beta-binding protein-3 (LTBP-3) is dependent on co-expression of TGF-beta*. J Cell Sci, 2002. **115**(Pt 17): p. 3457-68.
133. Kantola, A.K., et al., *Independent regulation of short and long forms of latent TGF-beta binding protein (LTBP)-4 in cultured fibroblasts and human tissues*. J Cell Physiol, 2010.
134. Unsold, C., et al., *Latent TGF-beta binding protein LTBP-1 contains three potential extracellular matrix interacting domains*. J Cell Sci, 2001. **114**(Pt 1): p. 187-197.
135. Fontana, L., et al., *Fibronectin is required for integrin alphavbeta6-mediated activation of latent TGF-beta complexes containing LTBP-1*. FASEB J, 2005. **19**(13): p. 1798-808.
136. Kantola, A.K., J. Keski-Oja, and K. Koli, *Fibronectin and heparin binding domains of latent TGF-beta binding protein (LTBP)-4 mediate matrix targeting and cell adhesion*. Exp Cell Res, 2008. **314**(13): p. 2488-500.
137. Dallas, S.L., et al., *Fibronectin regulates latent transforming growth factor-beta (TGF beta) by controlling matrix assembly of latent TGF beta-binding protein-1*. J Biol Chem, 2005. **280**(19): p. 18871-80.
138. Koli, K., et al., *Sequential deposition of latent TGF-beta binding proteins (LTBPs) during formation of the extracellular matrix in human lung fibroblasts*. Exp Cell Res, 2005. **310**(2): p. 370-82.

139. Chen, Q., et al., *Potential role for heparan sulfate proteoglycans in regulation of transforming growth factor-beta (TGF-beta) by modulating assembly of latent TGF-beta-binding protein-1*. J Biol Chem, 2007. **282**(36): p. 26418-30.
140. Oklu, R., et al., *Loss of a consensus heparin binding site by alternative splicing of latent transforming growth factor-beta binding protein-1*. FEBS Lett, 1998. **425**(2): p. 281-5.
141. Hyttiainen, M. and J. Keski-Oja, *Latent TGF-beta binding protein LTBP-2 decreases fibroblast adhesion to fibronectin*. J Cell Biol, 2003. **163**(6): p. 1363-74.
142. Nunes, I., et al., *Latent transforming growth factor-beta binding protein domains involved in activation and transglutaminase-dependent cross-linking of latent transforming growth factor-beta*. J Cell Biol, 1997. **136**(5): p. 1151-63.
143. Olofsson, A., et al., *Efficient association of an amino-terminally extended form of human latent transforming growth factor-beta binding protein with the extracellular matrix*. J Biol Chem, 1995. **270**(52): p. 31294-7.
144. Hirani, R., E. Hanssen, and M.A. Gibson, *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, 2007. **26**(4): p. 213-23.
145. Zilberberg, L., et al., *Specificity of latent TGF-ss binding protein (LTBP) incorporation into matrix: role of fibrillins and fibronectin*. J Cell Physiol, 2012.
146. Vehvilainen, P., M. Hyttiainen, and J. Keski-Oja, *Matrix association of latent TGF-beta binding protein-2 (LTBP-2) is dependent on fibrillin-1*. J Cell Physiol, 2009. **221**(3): p. 586-93.
147. Chaudhry, S.S., et al., *Fibrillin-1 regulates the bioavailability of TGFbeta1*. J Cell Biol, 2007. **176**(3): p. 355-67.
148. Munger, J.S., et al., *The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis*. Cell, 1999. **96**(3): p. 319-28.
149. Yang, Z., et al., *Absence of integrin-mediated TGFbeta1 activation in vivo recapitulates the phenotype of TGFbeta1-null mice*. J Cell Biol, 2007. **176**(6): p. 787-93.
150. Aluwihare, P., et al., *Mice that lack activity of alphavbeta6- and alphavbeta8-integrins reproduce the abnormalities of Tgfb1- and Tgfb3-null mice*. J Cell Sci, 2009. **122**(Pt 2): p. 227-32.
151. Wipff, P.J., et al., *Myofibroblast contraction activates latent TGF-beta1 from the extracellular matrix*. J Cell Biol, 2007. **179**(6): p. 1311-23.
152. Ahamed, J., et al., *In vitro and in vivo evidence for shear-induced activation of latent transforming growth factor-beta1*. Blood, 2008. **112**(9): p. 3650-60.
153. Buscemi, L., et al., *The single-molecule mechanics of the latent TGF-beta1 complex*. Curr Biol. **21**(24): p. 2046-54.
154. Lepparanta, O., et al., *Regulation of TGF-beta storage and activation in the human idiopathic pulmonary fibrosis lung*. Cell Tissue Res, 2012.
155. Chandramouli, A., et al., *Choreographing metastasis to the tune of LTBP*. J Mammary Gland Biol Neoplasia, 2011. **16**(2): p. 67-80.
156. Haji-Seyed-Javadi, R., et al., *LTBP2 mutations cause Weill-Marchesani and Weill-Marchesani-like syndrome and affect disruptions in the extracellular matrix*. Hum Mutat, 2012.
157. Giltay, R., G. Kostka, and R. Timpl, *Sequence and expression of a novel member (LTBP-4) of the family of latent transforming growth factor-beta binding proteins*. FEBS Lett, 1997. **411**(2-3): p. 164-8.
158. Saharinen, J., et al., *Identification and characterization of a new latent transforming growth factor-beta-binding protein, LTBP-4*. J Biol Chem, 1998. **273**(29): p. 18459-69.
159. Drews, F., et al., *Disruption of the latent transforming growth factor-beta binding protein-1 gene causes alteration in facial structure and influences TGF-beta bioavailability*. Biochim Biophys Acta, 2008. **1783**(1): p. 34-48.
160. Todorovic, V., et al., *Long form of latent TGF-beta binding protein 1 (Ltbp1L) is essential for cardiac outflow tract septation and remodeling*. Development, 2007. **134**(20): p. 3723-32.
161. Shipley, J.M., et al., *Developmental expression of latent transforming growth factor beta binding protein 2 and its requirement early in mouse development*. Mol Cell Biol, 2000. **20**(13): p. 4879-87.
162. Todorovic, V. and D.B. Rifkin, *LTBPs, more than just an escort service*. J Cell Biochem, 2012. **113**(2): p. 410-8.

163. Dabovic, B., et al., *Bone abnormalities in latent TGF-[beta] binding protein (Ltbp)-3-null mice indicate a role for Ltbp-3 in modulating TGF-[beta] bioavailability.* J Cell Biol, 2002. **156**(2): p. 227-32.
164. Dabovic, B., et al., *Bone defects in latent TGF-beta binding protein (Ltbp)-3 null mice; a role for Ltbp in TGF-beta presentation.* J Endocrinol, 2002. **175**(1): p. 129-41.
165. Koli, K., et al., *Disruption of LTBP-4 function reduces TGF-beta activation and enhances BMP-4 signaling in the lung.* J Cell Biol, 2004. **167**(1): p. 123-33.
166. Sterner-Kock, A., et al., *Disruption of the gene encoding the latent transforming growth factor-beta binding protein 4 (LTBP-4) causes abnormal lung development, cardiomyopathy, and colorectal cancer.* Genes Dev, 2002. **16**(17): p. 2264-73.
167. Todorovic, V., et al., *Long form of latent TGF-beta binding protein 1 (Ltbp1L) regulates cardiac valve development.* Dev Dyn, 2011. **240**(1): p. 176-87.
168. Chen, Y., et al., *Growth retardation as well as spleen and thymus involution in latent TGF-beta binding protein (Ltbp)-3 null mice.* J Cell Physiol, 2003. **196**(2): p. 319-25.
169. Colarossi, C., et al., *Lung alveolar septation defects in Ltbp-3-null mice.* Am J Pathol, 2005. **167**(2): p. 419-28.
170. Dabovic, B., et al., *Osteopetrosis-like phenotype in latent TGF-beta binding protein 3 deficient mice.* Bone, 2005. **37**(1): p. 25-31.
171. Zhou, Y., et al., *Latent TGF-beta binding protein 3 identifies a second heart field in zebrafish.* Nature, 2011. **474**(7353): p. 645-8.
172. Dabovic, B., et al., *Control of lung development by latent TGF-beta binding proteins.* J Cell Physiol, 2011. **226**(6): p. 1499-509.
173. Wang, R., et al., *GARP regulates the bioavailability and activation of TGFbeta.* Mol Biol Cell, 2012. **23**(6): p. 1129-39.
174. Kulkarni, A.B., et al., *Transforming growth factor beta 1 null mutation in mice causes excessive inflammatory response and early death.* Proc Natl Acad Sci U S A, 1993. **90**(2): p. 770-4.
175. Sanford, L.P., et al., *TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes.* Development, 1997. **124**(13): p. 2659-70.
176. Kaartinen, V., et al., *Abnormal lung development and cleft palate in mice lacking TGF-beta 3 indicates defects of epithelial-mesenchymal interaction.* Nat Genet, 1995. **11**(4): p. 415-21.
177. Yoshinaga, K., et al., *Perturbation of transforming growth factor (TGF)-beta1 association with latent TGF-beta binding protein yields inflammation and tumors.* Proc Natl Acad Sci U S A, 2008. **105**(48): p. 18758-63.
178. Mazzieri, R., et al., *Expression of truncated latent TGF-beta-binding protein modulates TGF-beta signaling.* J Cell Sci, 2005. **118**(Pt 10): p. 2177-87.
179. Tsuji, Y., et al., *A potential suppressor of TGF-beta delays catagen progression in hair follicles.* J Invest Dermatol Symp Proc, 2003. **8**(1): p. 65-8.
180. Gregory, K.E., et al., *The prodomain of BMP-7 targets the BMP-7 complex to the extracellular matrix.* J Biol Chem, 2005. **280**(30): p. 27970-80.
181. Sengle, G., et al., *Targeting of bone morphogenetic protein growth factor complexes to fibrillin.* J Biol Chem, 2008. **283**(20): p. 13874-88.
182. Sengle, G., et al., *Prodomains of transforming growth factor beta (TGFbeta) superfamily members specify different functions: extracellular matrix interactions and growth factor bioavailability.* J Biol Chem, 2011. **286**(7): p. 5087-99.
183. Hanssen, E., et al., *MAGP-2 has multiple binding regions on fibrillins and has covalent periodic association with fibrillin-containing microfibrils.* J Biol Chem, 2004. **279**(28): p. 29185-94.
184. Jensen, S.A., et al., *Protein interaction studies of MAGP-1 with tropoelastin and fibrillin-1.* J Biol Chem, 2001. **276**(43): p. 39661-6.
185. El-Hallous, E., et al., *Fibrillin-1 interactions with fibulins depend on the first hybrid domain and provide an adaptor function to tropoelastin.* J Biol Chem, 2007. **282**(12): p. 8935-46.
186. Choudhury, R., et al., *Differential regulation of elastic fiber formation by fibulin-4 and -5.* J Biol Chem, 2009. **284**(36): p. 24553-67.
187. Hirai, M., et al., *Fibulin-5/DANCE has an elastogenic organizer activity that is abrogated by proteolytic cleavage in vivo.* J Cell Biol, 2007. **176**(7): p. 1061-71.
188. Trask, B.C., et al., *The microfibrillar proteins MAGP-1 and fibrillin-1 form a ternary complex with the chondroitin sulfate proteoglycan decorin.* Mol Biol Cell, 2000. **11**(5): p. 1499-507.

189. Tiedemann, K., et al., *Microfibrils at basement membrane zones interact with perlecan via fibrillin-1*. J Biol Chem, 2005. **280**(12): p. 11404-12.
190. Isogai, Z., et al., *Versican interacts with fibrillin-1 and links extracellular microfibrils to other connective tissue networks*. J Biol Chem, 2002. **277**(6): p. 4565-72.
191. Kutz, W.E., et al., *ADAMTS10 protein interacts with fibrillin-1 and promotes its deposition in extracellular matrix of cultured fibroblasts*. J Biol Chem. **286**(19): p. 17156-67.
192. Cain, S.A., et al., *Defining elastic fiber interactions by molecular fishing: an affinity purification and mass spectrometry approach*. Mol Cell Proteomics, 2009. **8**(12): p. 2715-32.
193. Kassner, A., et al., *Molecular structure and interaction of recombinant human type XVI collagen*. J Mol Biol, 2004. **339**(4): p. 835-53.
194. Kassner, A., et al., *Discrete integration of collagen XVI into tissue-specific collagen fibrils or beaded microfibrils*. Matrix Biol, 2003. **22**(2): p. 131-43.
195. Hubmacher, D. and S.S. Apte, *Genetic and functional linkage between ADAMTS superfamily proteins and fibrillin-1: a novel mechanism influencing microfibril assembly and function*. Cell Mol Life Sci, 2011. **68**(19): p. 3137-48.
196. Rock, M.J., et al., *Molecular basis of elastic fiber formation. Critical interactions and a tropoelastin-fibrillin-1 cross-link*. J Biol Chem, 2004. **279**(22): p. 23748-58.
197. Groppe, J., et al., *Cooperative assembly of TGF-beta superfamily signaling complexes is mediated by two disparate mechanisms and distinct modes of receptor binding*. Mol Cell, 2008. **29**(2): p. 157-68.
198. Lopez-Casillas, F., J.L. Wrana, and J. Massague, *Betaglycan presents ligand to the TGF beta signaling receptor*. Cell, 1993. **73**(7): p. 1435-44.
199. Denner, S., et al., *Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene*. EMBO J, 1998. **17**(11): p. 3091-100.
200. Holm, T.M., et al., *Noncanonical TGFbeta signaling contributes to aortic aneurysm progression in Marfan syndrome mice*. Science, 2011. **332**(6027): p. 358-61.
201. Gui, Y. and L.J. Murphy, *Interaction of insulin-like growth factor binding protein-3 with latent transforming growth factor-beta binding protein-1*. Mol Cell Biochem, 2003. **250**(1-2): p. 189-95.
202. Hirai, M., et al., *Latent TGF-beta-binding protein 2 binds to DANCE/fibulin-5 and regulates elastic fiber assembly*. EMBO J, 2007. **26**(14): p. 3283-95.
203. Brooke, J.S., J.H. Cha, and L. Eidels, *Latent transforming growth factor beta-binding protein-3 and fibulin-1C interact with the extracellular domain of the heparin-binding EGF-like growth factor precursor*. BMC Cell Biol, 2002. **3**: p. 2.
204. Anderson, S.B., A.L. Goldberg, and M. Whitman, *Identification of a novel pool of extracellular pro-myostatin in skeletal muscle*. J Biol Chem, 2008. **283**(11): p. 7027-35.
205. Rifkin, D.B., *Latent transforming growth factor-beta (TGF-beta) binding proteins: orchestrators of TGF-beta availability*. J Biol Chem, 2005. **280**(9): p. 7409-12.
206. Todorovic, V., et al., *Latent TGF-beta binding proteins*. Int J Biochem Cell Biol, 2005. **37**(1): p. 38-41.
207. Schrijver, I., et al., *Cysteine substitutions in epidermal growth factor-like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes*. Am J Hum Genet, 1999. **65**(4): p. 1007-20.
208. Liu, W.O., et al., *Denaturing HPLC-identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders*. Genet Test, 1997. **1**(4): p. 237-42.
209. Neptune, E.R., et al., *Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome*. Nat Genet, 2003. **33**(3): p. 407-11.
210. Lacro, R.V., et al., *Rationale and design of a randomized clinical trial of beta-blocker therapy (atenolol) versus angiotensin II receptor blocker therapy (losartan) in individuals with Marfan syndrome*. Am Heart J, 2007. **154**(4): p. 624-31.

Chapter 2- Evolution of fibrillins and LTBP

2.1 - Introduction

While there is a large amount of information available on the biological significance of fibrillins and LTBP, the evolutionary story of how these TB domain containing proteins emerged, and acquired their important physiological functions, is less well understood. In order to investigate the evolutionary history of the fibrillins and LTBP, protein sequence information from a variety of sources was collated and examined. In this chapter the broad evolutionary history of these proteins is investigated, and the conservation of specific regions relevant to the fibrillin - LTBP interaction is examined in detail. This evolutionary data was also used to write a more general review on the evolution of fibrillins and LTBP and their roles in matrix biology [2]¹.

2.2- Evolutionary history of the TB domain protein family

The large and constantly growing number of available genome sequences has allowed a picture to be built showing how the TB domain protein family emerged and diversified during the evolution of the animal kingdom. Much uncertainty still remains, as there are still many key genomes to be sequenced, and very few genomes that have been comprehensively mapped and annotated. However the picture that can be assembled is still highly informative when considering the functions of these proteins in the ECM.

For these analyses, searches for TB domain proteins were carried out using basic local alignment search tool (BLAST) algorithms against a variety of protein (blastp) and nucleotide (tblastn) libraries, with sequences of selected TB domains as the

¹ Included at the back of this thesis.

search query (see materials and methods Section 8.1). Searches were also carried out with longer sections of the fibrillin and LTBP proteins, but results from these are often uninformative, as such searches often detect numerous proteins that contain multiple EGF/cbEGF repeats. While these EGF repeat proteins score highly as homologues of fibrillin and LTBP, they may be clearly distinct when their overall domain structure is examined, and crucially often lack TB or hybrid-like domains, which define membership of this family.

Figure 2.1 summarises the TB domain proteins found in a wide variety of species, and their domain structures. Species were chosen to give a broad representation of the animal kingdom, and key points of divergence in the evolution of *Homo sapiens*. Figure 2.1 also displays a tree representing evolutionary relationships of the species shown, (according to the Tree of life web project (<http://tolweb.org/tree/>)). In all available genomes (as of March 2012), the organisms most distantly related to humans that contain TB domain proteins are the cnidarians². It is possible that TB domain proteins are restricted to the eumetazoans³, as they could not be found among the available sequences from simpler metazoans, such as placazoa⁴, porifera⁵, or fungi, plants or prokaryotes. However the draft nature of the *A. queenslandica* sponge and *Trichoplax* genomes, mean that TB domain proteins may still be present, but their genes could have gone un-sequenced or un-detected. It is clear though that TB

² Cnidaria are organisms like jellyfish possessing only radial symmetry, giving them just a “top and bottom”

³ Clade of organisms in which the tissues are arranged into distinct germ layers, and where embryos go through a gastrulation stage. Traditionally comprised of Cnidarians, Bilateria and Ctenophora, (although recent studies of the *Mnemiopsis leidyi* genome suggest that Cnidarians may be more closely related to placazoa than Ctenophora [3])

⁴ Simple organisms with two cell layers and no basement membrane, eg *Trichoplax adhaerens*.

⁵ Sponges- eg *Amphimendon queenslandica* (genome sequenced)

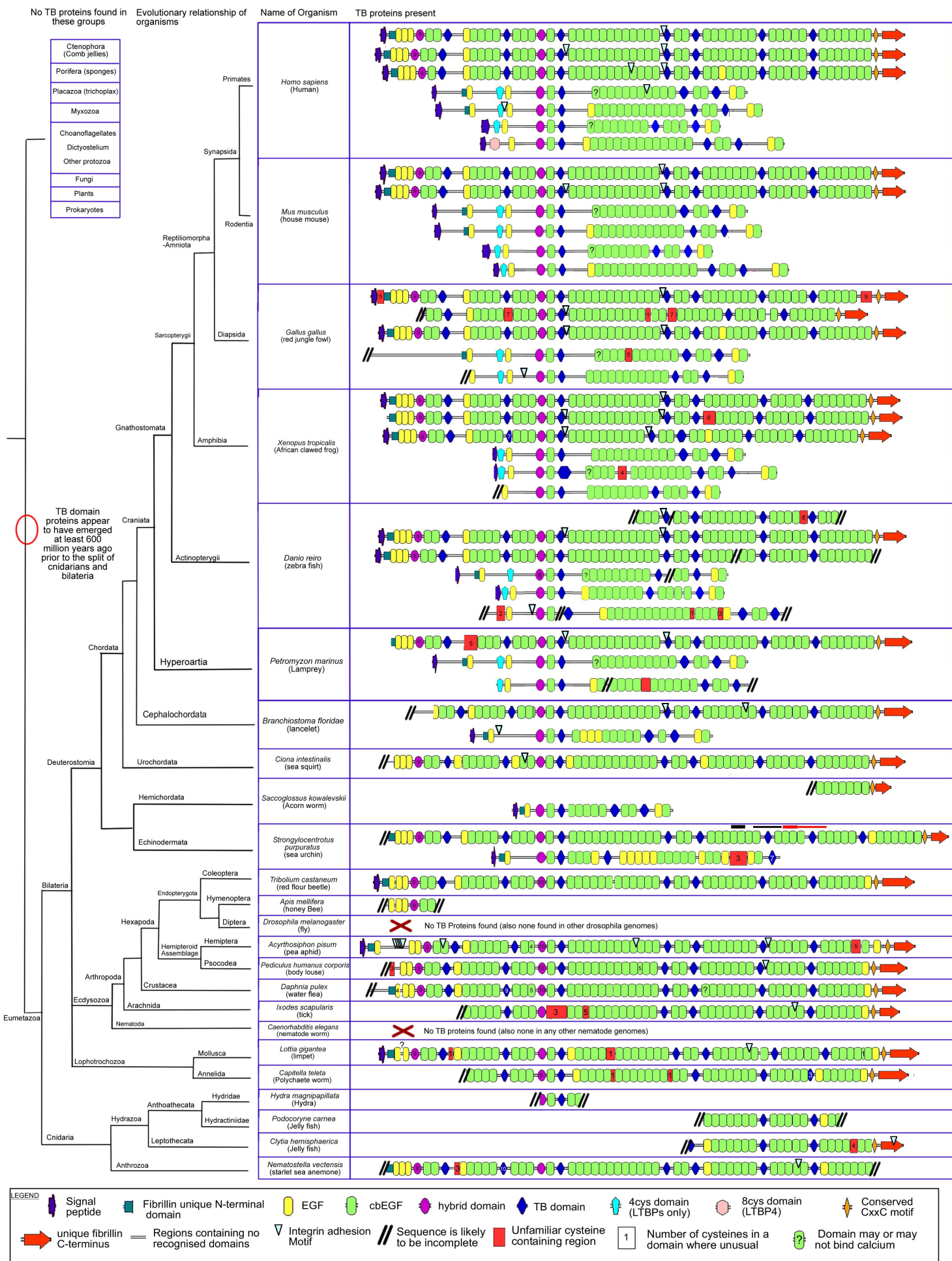


Figure 2.1 (previous page)- Domain organisation of TB proteins in a variety of organisms, and the evolutionary relationships between them.

Proteins were detected as described in materials and methods Section 8.1. Domains were annotated initially using the SMART web based domain prediction tool [1], followed by manual inspection of regions of interest, and location of cysteines not associated with domains. It should be noted that many of these protein sequences are predicted from shotgun genome sequences of varying quality, and are not backed up by cDNAs, so many may be missing exons or include other errors. These errors likely give rise to the various unrecognised cysteine-containing domains shown in orange. Where possible re-prediction of proteins was carried out from genomic DNA, using the WISE2 web based tool (<http://www.ebi.ac.uk/Tools/Wise>) with human fibrillin-1 as an example sequence. This gave significant improvements to the protein sequence predictions in some cases, such as in the fibrillin of *N. vectensis* and *I. scapularis*. Sequence references are given in Appendix 4. The phylogenetic tree shown is based on tree of life web project (<http://tolweb.org/tree/>).

domain proteins emerged before the divergence of bilateria⁶ and cnidaria, which occurred over 600 million years ago [4], as these proteins are present in both groups.

The genome sequence for the starlet sea anemone *Nematostella vectensis* contains a near complete sequence of a fibrillin-like protein which is remarkably similar to our own in terms of domain organisation. Several fragmented fibrillin sequences are also available from other cnidarians, and are also consistent with the human fibrillin1 domain organisation. In fact the extent of conservation of domain architecture across all fibrillin-like proteins is quite remarkable. For example, the only differences in domain structure between sea anemone fibrillin and human fibrillin1, are two cbEGF domains being replaced by EGF domains in the anemone, and in the anemone the “proline rich” region found in human fibrillin1 is absent. Across all organisms looked at the only significant divergence from the signature ~47 EGF/cbEGF, 7 TB and 2 hybrid domain organisation of the fibrillins is seen in the sea urchin *Strongylocentrotus purpuratus*, with the inclusion of an extra TB domain and four cbEGF domains. However alignment of these additional domains with the rest of the protein demonstrates they are 100% identical in amino acid sequence to other regions of the same molecule (duplicate domains highlighted with red and black bars above

⁶ Bilateria are organisms such as ourselves with bilateral symmetry, possessing a “front and a back” as well as a “top and a bottom”

the relevant protein in Figure 2.1). Although there is reportedly 3.9% variation in the nucleotide sequence between these duplicated regions [5], the close similarity either suggests a very recent evolutionary event, or an error in the sequencing process. cDNA sequencing would be informative to confirm these extra domains are actually expressed in *S. purpuratus*.

The conservation of the fibrillin domain structure may relate to its importance in microfibril structure and assembly (Section 1.3.1). Microfibrils have been demonstrated to be present in cnidaria (specifically in the jellyfish *Podocoryne carnea* [6]), where rotary shadowing electron microscopy shows that they form beaded string structures, similar to those seen in humans. Other sequence features that have been implicated in microfibril assembly are also conserved throughout the fibrillin sequences available, including the presence of an additional single cysteine in the first hybrid domain [2].

The lack of comprehensive genome information for the early metazoans limits investigations into the initial appearance of the TB domain proteins. However the large number of TB domains seen in the sea anemone and other early fibrillins, would have required a significant number of exon duplications and rearrangements before reaching the well-conserved fibrillin domain organisation we see in genomes today. This suggests that the first TB domain might have emerged quite some time before the split of cnidaria and bilateria.

Although BLAST searches could not detect any TB domain proteins outside of the eumetazoa, the structure of human follistatin [7] has revealed that its N-terminal domain is structurally similar to fibrillin hybrid and TB domains (Figure 2.2 A and C),

although the lack of sequence homology means follistatins are not detected with TB domain blast queries. BLAST searches with the follistatin domain sequence detect follistatin-like proteins in cnidarians, and also in the simpler metazoan *Trichoplax adhaerens* (fig 2.2 B and D). If the follistatin N-terminal domain and fibrillin TB domains both evolved from a single ancestor protein, instead of emerging independently, then this ancestor protein must have evolved prior to the divergence of placazoa and eumetazoa.

A surprising feature of the evolution of TB proteins as a whole is their absence from both of the model organisms *Drosophila melanogaster* and *Caenorhabditis elegans*, which have been extensively studied and their genomes fully sequenced. Further analysis also reveals TB proteins to be absent from predicted protein and EST libraries of various other related *Drosophila* and nematode species. With little information on the precise role of fibrillin microfibrils outside of vertebrates, it is difficult to hypothesise why these organisms have lost fibrillin when it is so well conserved elsewhere. However the fact that both these organisms have been demonstrated to have rapidly evolving genomes [8, 9], suggests they have undergone many changes to their biochemistry in adapting to their respective niches.

While it appears then that TB domain proteins are quite ancient, and fibrillin-like proteins are present in many organisms, they do not appear to diversify until the emergence of the deuterostomes⁷, where the LTBP domain organisation is first seen, and in vertebrates multiple LTBP and fibrillin isoforms also appear.

⁷ A group of organisms defined by the fact that the anus forms before the mouth during gastrulation, thought to have emerged over 542million years ago (prior to the Cambrian explosion)

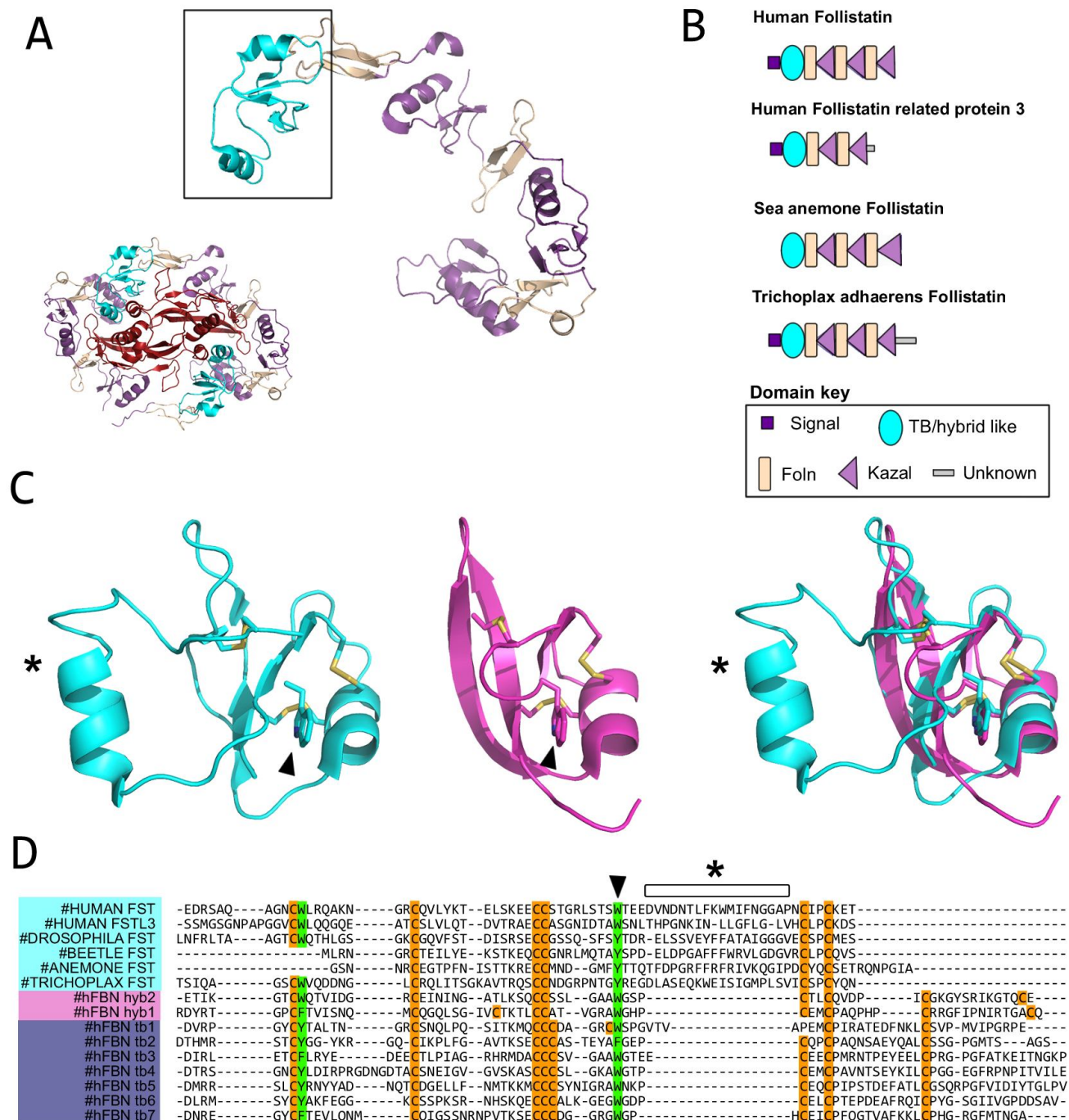


Figure 2.2- Comparison of follistatin N-terminal domain with hybrid and TB domains.

A) Structure of the follistatin- activin complex (PDB 2BOU), (bottom left) the activin dimer is coloured dark red while the follistatin protein is coloured by domain, and shown in detail in the top right. The follistatin N-terminal domain is coloured light blue and boxed, the EGF-like follistatin (“folln”) domains are coloured cream, and the “kazal” type domains are coloured mauve. **B)** The domain organisations of follistatin sequences from selected organisms, demonstrating highly similar domain arrangements. **C)** Direct comparison of the structures of the follistatin N-terminal domain (light blue) and fibrillin hyb2 domain (magenta) (PDB 2W86), with cysteine and core tryptophan residues shown as sticks. An arrow head highlights the core tryptophan while an asterisk highlights the extra alpha helix unique to the follistatin domain. **D)** Alignments of follistatin N-terminal domains from various species with human fibrillin hybrid and TB domains. Key cysteine residues are highlighted in orange, while structurally significant aromatic residues are highlighted in green. The arrowhead and asterisk highlight the same features as in C.

The most evolutionarily remote LTBP-like proteins from human LTBPs were detected in the sea urchin, *S. purpuratus*, and in the acorn worm, *Saccoglossus kowalevskii*. Generally the LTBP domain organisation appears less well conserved than that of fibrillin, although there are many common domain features. The domain organisation of the *S. purpuratus* LTBP is particularly unusual, with a large stretch of extra EGFs and cbEGFs predicted between its second and third TB domain, where other LTBPs only have 1-4 cbEGFs (Figure 2.1).

In the early LTBPs the number of EGF/cbEGF domains in the tandem array between the first and second TB domain varies significantly, and this is also the case between the different LTBP isoforms that emerged later. Additionally the “4-cysteine” domain that is unique to LTBPs (blue pentagon in Figure 2.1) can only be found in the sequences from members of the craniata clade⁸, starting with lamprey, perhaps suggesting this domain evolved later, and is involved in a more recently acquired function of the LTBPs.

Despite their differences in domain organisation, all these early LTBP-like proteins contain a unique two residue sequence insertion between the 6th and 7th cysteines of the second TB domain, which has been shown to be crucial for covalent binding of the TGF β propeptide by LTBP (as discussed in Chapter 1 Section 1.4.2). The conservation of this insert is shown in Figure 2.3, and further supports the functional importance of TGF β binding by these LTBP proteins.

⁸ Clade of chordates including lampreys, distinguished by having a hard skull made of either bone or cartilage.

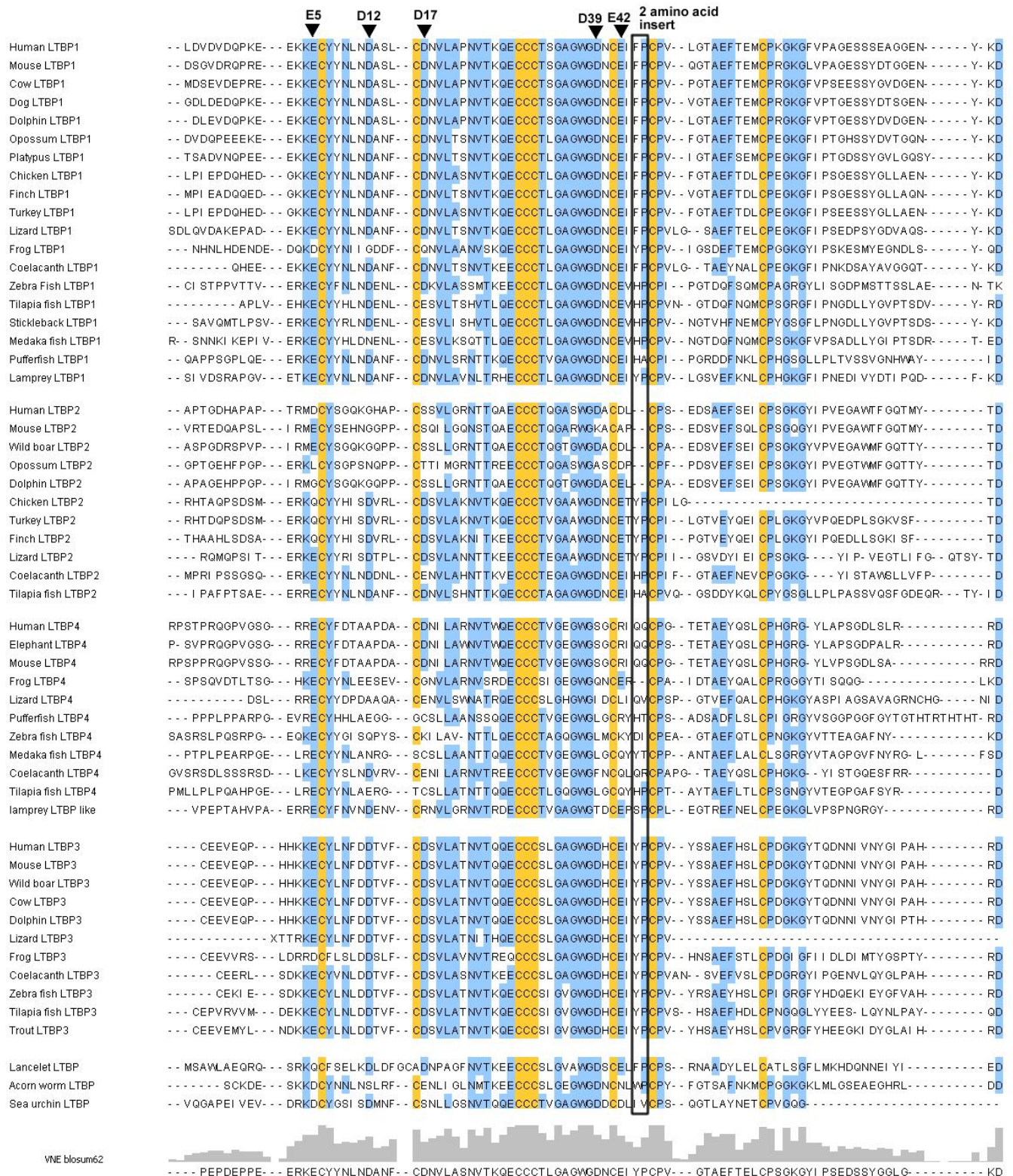


Figure 2.3 - Alignment of the second TB domain from various LTBP

Sequences from a variety of organisms representing all LTBP isoforms were aligned with the MEGA 5 (see section 8.3). The background behind residues present in >60% of all sequences shown was coloured blue, or yellow when cysteine. Alignment divided into blocks by isoform. Arrowheads highlight acidic residues implicated in binding the TGFβ propeptide (numbered according to Chen et al [20]), and the two amino acid insert shown to be important for TGFβ binding is surrounded by a black box. The level of conservation is quantified at the bottom of the alignment in a chart based on the BLOSUM 62 matrix.

Searches of TGF β like sequences from various organisms suggests that true TGF β growth factors⁹ also first emerged in the deuterostomes, and are also first seen in sea urchins [2]. It is possible then that LTBP β s and TGF β have co-evolved in adopting the key roles they now play in mammalian development and homeostasis.

2.3- Phylogenetic analysis of fibrillins and LTBP β s

While the overall fibrillin domain structure has remained relatively unchanged since the split of cnidaria and bilateria, the primary amino acid sequence has still changed considerably over time, for example sea anemone fibrillin only shares 42% sequence homology¹⁰ with human fibrillin1. To examine the relative homologies between the fibrillins, a phylogenetic tree was generated (Figure 2.4) using the maximum likelihood method, from alignments of the entire fibrillin molecule (materials and methods Section 8.4).

The groupings of fibrillin sequences in this tree closely reflect the phylogenetic groupings of the organisms from which they originate. For example all the chordate fibrillins are contained in one sub-tree, while arthropod fibrillins are grouped in another. As well as supporting the evolutionary relationships between these organisms (Figure 2.1), this also suggests that the variation in fibrillin sequences seen between organisms may be primarily due to random background mutation rather than any novel selection pressures.

⁹ “True TGF β growth factors” here refers to proteins that are more homologous to TGF β 1, 2 and 3, than to any other member of the other TGF β superfamily (such as the BMPs, activins, and myostatins). TGF β superfamily members are present throughout the metazoa.

¹⁰ As calculated by the lalign pair-wise alignment server (www.ch.embnet.org/software/LALIGN_form.html)

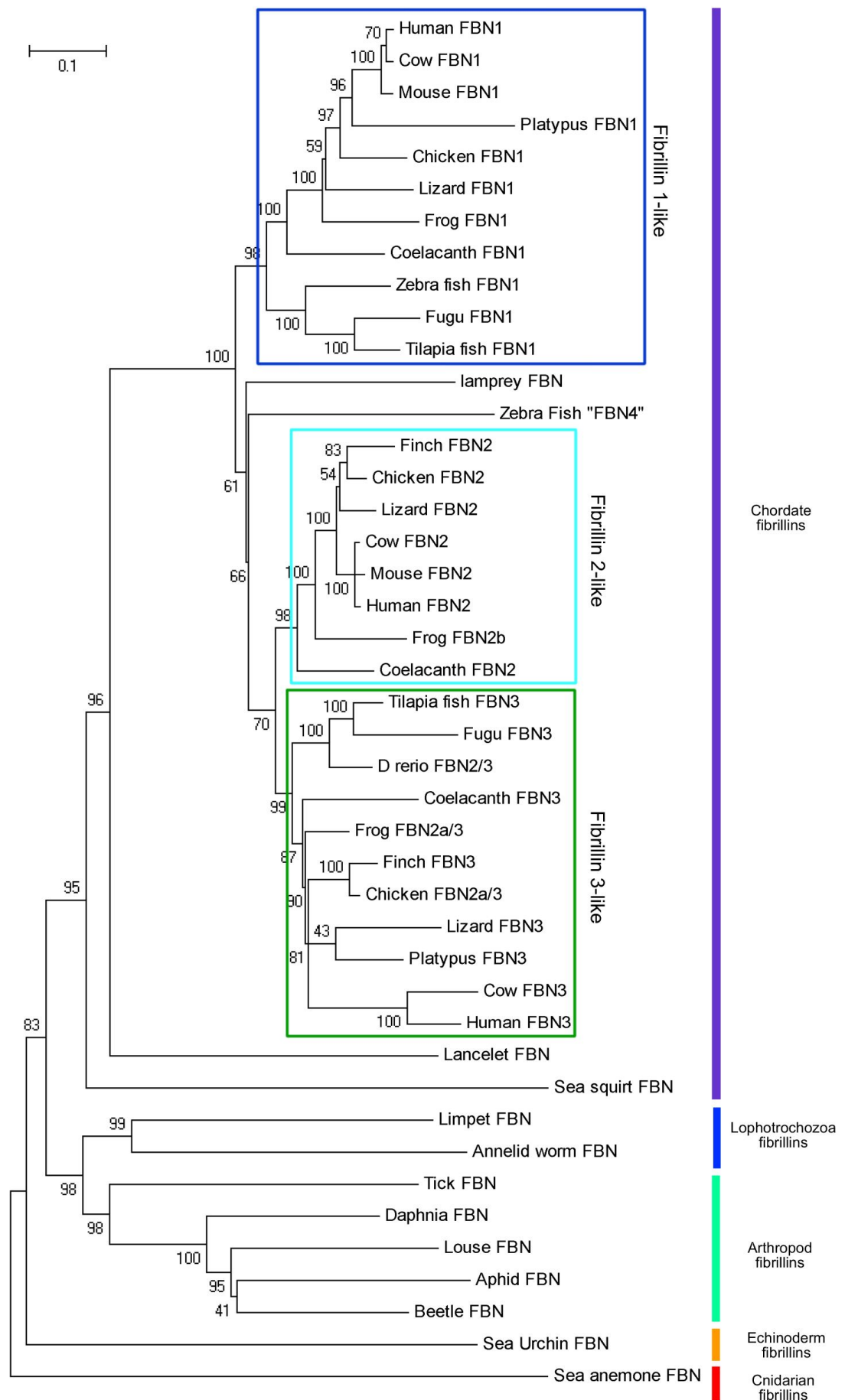


Figure 2.4 - Fibrillin phylogenetic tree

Coloured boxes highlight groupings of specific fibrillin isoforms, while coloured lines to the right highlight their source taxonomic groups. The bar at the top right gives the distance scale used, ie the difference between sequences (0.1 corresponds to a 10% divergence). Numbers at the root of each branch represent statistical confidence in that sub-tree based on bootstrapping replicates. Details of parameters used given in Section 8.4

Three different fibrillin isoforms appear with the bony fish (Figure 2.1), and in Figure 2.4 these form two main sub-trees, consisting of those fibrillins that resemble fibrillin 1, and those which resemble fibrillin 2 and 3. While only one fibrillin sequence could be found in the lamprey genome, several fish genomes contain three different fibrillin genes. From their homologies they appear to be the predecessors of human fibrillin1, 2 and 3. This would mean just two gene duplication events have led to the three human fibrillins we have today, which may have occurred after the ancestors of lamprey and bony fish diverged.

Some of the fibrillins within the sub-tree containing fibrillin2 and fibrillin3 appear more distinct from the rest, for example one of the zebra fish fibrillin genes is not closely homologous to human fibrillin2 or 3, which led to it being initially named “fibrillin4” [5], although the lack of a fibrillin2 in zebra fish leaves the possibility that this “fibrillin4” arose due to significant changes in the fibrillin2 gene. The mammalian fibrillin3 proteins also appear more distantly removed from other proteins in the fibrillin3 sub-tree, maybe as a result of different selection pressures. Separate analyses performed recently by Piha-Gossak et al. also support these observations [5], using concatenated TB domains in their alignments instead of full-length sequences, and a more mammalian focused set of sequences.

Phylogenetic trees were also constructed to investigate the sequence relationships between LTBP isoforms from different organisms (Figure 2.5). As LTBPs are restricted to the deuterostomes, LTBP-like sequences from a wider variety of these organisms were included than with the fibrillins, in order to give a more comprehensive picture of LTBP evolution. The tree shown in Figure 2.5 is based on alignments of the available LTBP sequences, starting from the beginning of the central EGF array, through to the

C-terminus. These restricted alignments were used as they exclude the highly variable N-terminal regions, which were difficult to align between all LTBP isoforms.

Specific sub-trees appear to represent the LTBP1, LTBP2, LTBP3, and LTBP4 isoforms, with the internal branches of these sub-trees generally reflecting the evolutionary relationships between the source organisms. Sea urchin, acorn worm, and lancelets each only seem to have a single LTBP-like protein, and these three LTBPs cluster to a separate sub-tree, and do not appear to have particularly close homology with any specific LTBP isoforms. Lampreys are more closely related to fish and the higher vertebrates, and have two LTBP isoforms, one with closest homology to LTBP1, and the other appears closest to an LTBP4-like protein in this tree.

Some fish genomes, such as tilapia and coelacanth, contain four LTBPs, while in zebrafish only three LTBP sequences could be found, with homology to LTBP1, 3 and 4. Also only LTBP1, 3 and 4-like proteins could be found in the amphibian *Xenopus tropicalis* genome. From construction of several different phylogenetic trees there was some ambiguity over the classification of LTBP2-like proteins, as the fish LTBP2 sequences do not appear closely homologous to the reptile, bird and mammalian LTBP2 like proteins, and would not always locate to the LTBP2 sub-tree when more restricted alignments were used. To investigate this further the genomic context of LTBP2 was examined. Similar genes were seen flanking LTBP2 in the Tilapia fish genome, and in the lizard and human genomes (eg FCF1 and SYNDIG1L), strongly suggesting the fish LTBP2 genes and mammalian LTBP2 genes arose from a single ancestral gene¹¹.

¹¹ Other LTBP genes all flanked by a different non-homologous genes, eg in humans LTBPs have the following upstream genes: LTBP1- RASGRP3, LTBP4- SHKBP1 , LTBP3- SSCA1.

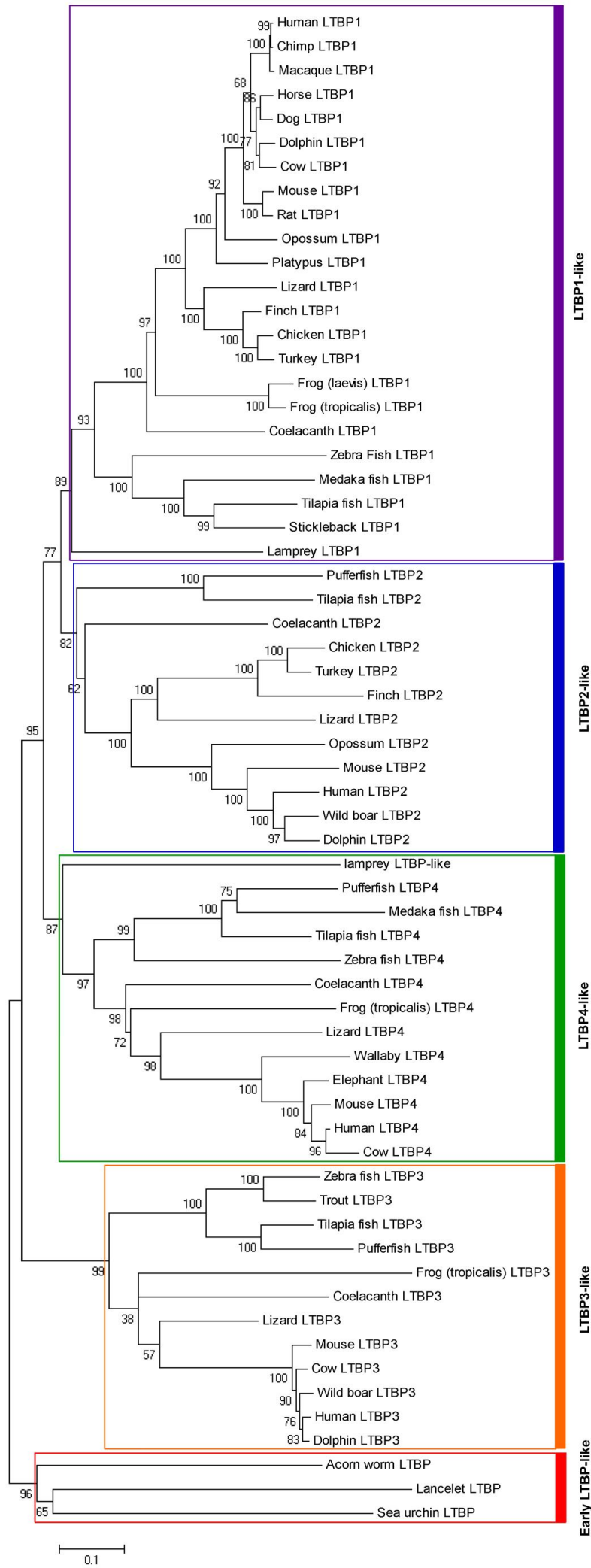


Figure 2.5 - LTBP Phylogenetic tree

Coloured boxes highlight the sub trees specific to particular LTBP isoforms. The bar at the bottom of the tree gives the distance scale used, ie the difference between sequences (0.1 corresponds to a 10% divergence). Numbers at the root of each branch represent statistical confidence in that sub tree based on bootstrapping replicates. Details of parameters used to generate this tree are given in Section 8.4

While all four LTBP3s can be found in various mammalian genomes, no LTBP3 or LTBP4-like proteins could be found in any available bird genomes (turkey, chicken and zebra finch). Interestingly the bird, reptile and fish LTBP2-like proteins all contain the two residue insert between TB2 C6-C7, important for TGF β binding, as well as at least four of the five acidic residues involved in this interaction (Figure 2.3). The C6-C7 insert is only lost in the mammalian LTBP2 proteins, and in *Xenopus* LTBP4. This presents the possibility that non-mammalian LTBP2-like proteins may still be able to bind TGF β , which may have significant implications for their biological function.

It is an interesting coincidence that LTBP3, which has been shown by mouse models to be important for bone development, is first detected in the bony fishes, and may also be lost in birds. Fish were the first organisms to evolve a skeleton composed of hard bone rather than cartilage, while birds have evolved light hollow bones to aid flight. However the multitude of anatomical differences between the organisms looked at mean it is not possible to correlate the emergence of specific fibrillin and LTBP isoforms with the evolution of particular physiological adaptations.

An understanding of the evolutionary history of fibrillins and LTBP3s will hopefully be informative for further biological and biochemical studies, and may also be useful for selecting suitable model organisms for studying their role *in vivo*. Figure 2.6 summarises the emergence and diversification of both the fibrillins and LTBP3s, based on the information from the phylogenetic trees in Figure 2.4 and 2.5.

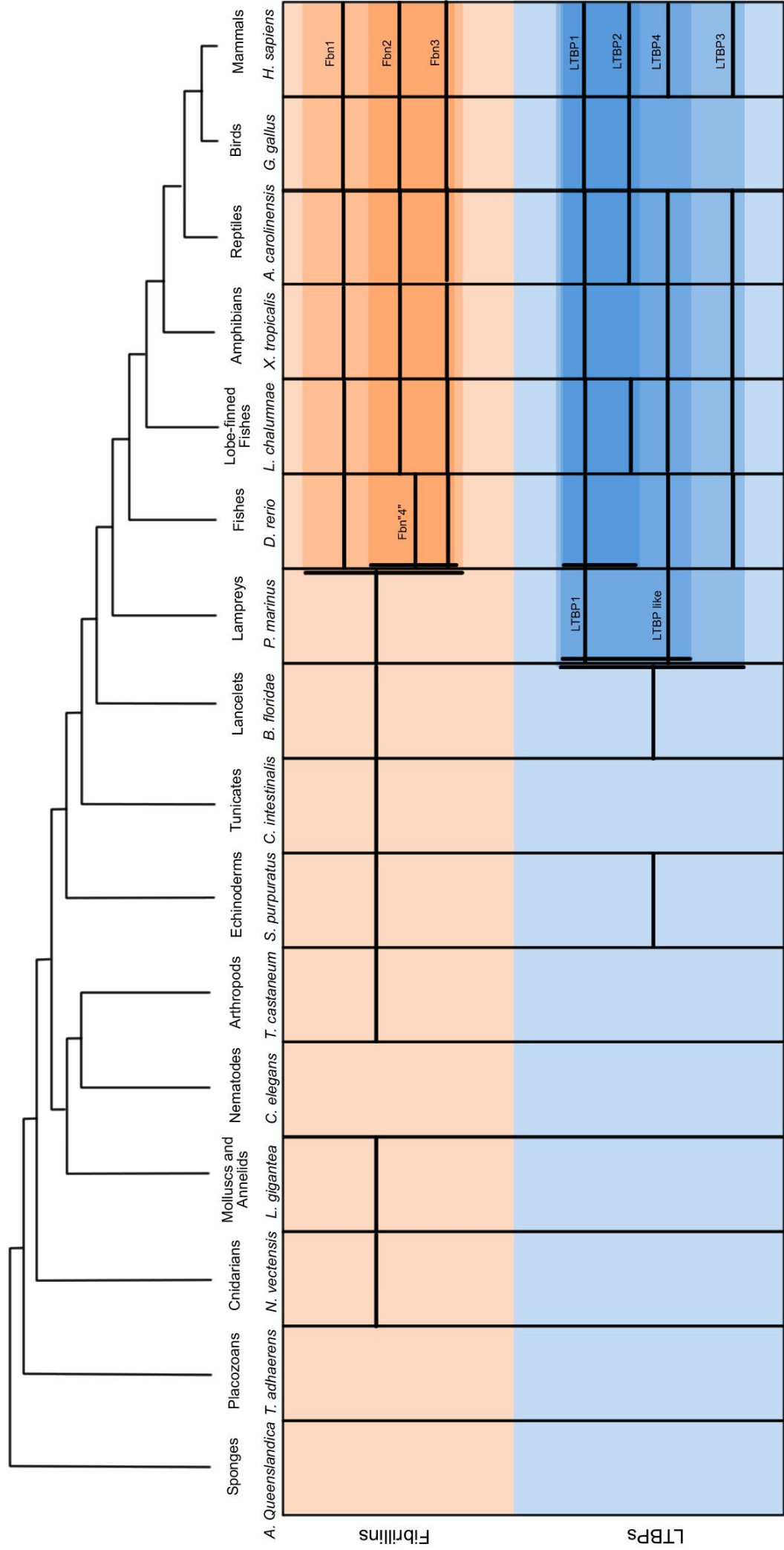


Figure 2.6- Summary of the emergence and diversification of fibrillins and LTBPs through evolution

This tabular figure consists of a column for each species, arranged by evolutionary distance, which is also summarised by the unscaled tree above the figure. Presence of the protein of interest is depicted by a black line. Lines are placed in series with their closest isoforms, and background shading depicts more closely homologous groups of isoforms, based on their placement within the phylogenetic trees (see figures 2.5 and 2.6 and Appendix 2 for full methodology).

2.4- Specific insights into the LTBP-Fibrillin interaction

As well as being helpful for establishing the evolutionary history of fibrillins and LTBPs, the genomic record can also inform on the molecular basis of fibrillin and LTBP activities by the presence or absence of sequence conservation in regions of interest.

Conservation in a number of interesting regions of the fibrillins and LTBPs is investigated in more detail in Robertson et al [2], but for the purposes of this thesis the rest of this chapter will focus on conservation in the fibrillin N-terminus and LTBP C-terminus, which are the regions essential for the fibrillin – LTBP interaction. In order to sequester latent TGF β to the ECM, LTBPs must interact with various insoluble matrix components (discussed in Chapter 1 Section 1.4.3). In humans the only interaction currently characterised in detail is the binding of the C-terminus of LTBP1, 2 and 4 to the N-terminus of fibrillin. No specific LTBP residues have yet been implicated as involved in the interaction, but in LTBP1 and 4 the binding site has been localised to the C-terminal region including TB3, EGF3 and cbEGF15 in LTBP1, and the equivalent domains in LTBP4 [10, 11].

As well as the conservation of cysteines and other key structural residues, there are a number of other conserved non-structural residues in the C-terminal domains of LTBP1, and other LTBP isoforms (Figure 2.7). It is possible that these residues play a role in a biologically significant protein-protein interaction, potentially with fibrillin, and are conserved for this reason. For example in the TB3 domain there is a conserved

Figure 2.7 (following 2 pages) - Alignment of LTBP C-terminus

Alignment of the C-terminal region of LTBPs implicated in fibrillin binding, including TB3, EGF3, cbEGF15 (numbered according to human LTBP1L). Due to the large size of this region the alignment is divided into two panels, A and B. Alignment was generated as described in Section 8.3. Residues that are present in >60% of all sequences used are highlighted blue, unless they are a cysteine residues, which are highlighted yellow.

A

TB3

□ 1515 □ 1525 □ 1535 □ 1545 □ 1555 □ 1565 □ 1575 □ 1585 □ 1595 □ 1605

Human_LTBP1

Mouse_LTBP1

Cow_LTBP1

Dog_LTBP1

Dolphin_LTBP1

Platyus_LTBP1

Chicken_LTBP1

Finch_LTBP1

Turkey_LTBP1

Lizard_LTBP1

Frog_(tropicalis)_LTBP1

Zebra_fish_LTBP1

Stickleback_LTBP1

Medaka_fish_LTBP1

Salmon_LTBP1

Lamprey_LTBP1

LTBP1

Human_LTBP2

Mouse_LTBP2

Dolphin_LTBP2

Opossum_LTBP2

Chicken_LTBP2

Turkey_LTBP2

LTBP2

Human_LTBP4

Elephant_LTBP4

Mouse_LTBP4

Lizard_LTBP4

Frog_(tropicalis)_LTBP4

Medaka_fish_LTBP4

LTBP4

Human_LTBP3

Mouse_LTBP3

Cow_LTBP3

Dolphin_LTBP3

Frog_(tropicalis)_LTBP3

Zebra_fish_LTBP3

Trout_LTBP3

Pufferfish_LTBP3

LTBP3

Lancelet_LTBP

Acom_worm_LTBP

Early

vME bloom62

B

EGF3

cbEGF15

	□ 1610	□ 1620	□ 1630	□ 1640	□ 1650	□ 1660	□ 1670	□ 1680	□ 1690	□ 1700	□ 1710	□ 1720
Human_L1TBP1	YFI QDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDTAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTDGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Mouse_L1TBP1	YFI QDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYI	PSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---
Cow_L1TBP1	YFYVDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Dog_L1TBP1	YFI QDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Dolphin_L1TBP1	YFI QDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Platyus_L1TBP1	YFYVDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Chicken_L1TBP1	YFYVDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Finch_L1TBP1	YFYVDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Turkey_L1TBP1	YFYVDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Lizard_L1TBP1	YFYVDRF LNSF EELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---	---
Frog_(tropicalis)_L1TBP1	YFI PERY	---	SYEELQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---
Zebra_fish_L1TBP1	YDEPRNPYDAF	EGLRAEEOGV	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---
Stickleback_L1TBP1	YPTADRLKA	FEGLRAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---
Medaka_fish_L1TBP1	TKTKPKYKAF	EGLRAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---
Salmon_L1TBP	YEDEENPYDAF	EGLRAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---
Lamprey_L1TBP1	ITFSSGLDSF	EGLQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---	---	---	---

L1TBP2

Human_L1TBP2	HSADRTPI	LESPLQPS	ELQPHYVA	SHPEPPAGF	EGLQAEEOGI	LNGCENGRCVRVQEG	YTCDCF EGFQLDAAHMA	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---
Mouse_L1TBP2	QPQDNTPV	LEPPLQPS	ELQPHYVA	SHPEPPAGF	EGLQAEEOGI	LNGCENGRCVRVQEG	YTCDCF EGFQLDAAHMA	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---
Dolphin_L1TBP2	RPEDRTPV	LEPPLQPS	ELQPHYVA	SHPEPPAGF	EGLQAEEOGI	LNGCENGRCVRVQEG	YTCDCF EGFQLDAAHMA	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---
Opossum_L1TBP2	HPVANL	PALLEPPLQPS	ELQPHYVA	SHPEPPAGF	EGLQAEEOGI	LNGCENGRCVRVQEG	YTCDCF EGFQLDAAHMA	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE	---
Chicken_L1TBP2	SSEF	ROSPTFH	---	GPORPPRYL	PSOPGLYEGF	EGLQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE
Turkey_L1TBP2	SSEF	ROSPTFH	---	GPORPPRYL	PSOPGLYEGF	EGLQAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE

L1TBP4

Human_L1TBP4	E	DDGGPYGESEAP	PPGPGTRWPYRS	ROTTRS	---	FPEPEEP	EGSYAGSLAEPY	EELEAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE
Elephant_L1TBP4	E	DDGGPYGESEAP	PPGPGTRWPYRS	ROTTRS	---	FPEPEEL	PEGGHNGALAPHE	LEAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE
Mouse_L1TBP4	E	DDGGPYGESEAP	PPGPGTRWPYRS	ROTTRS	---	FPEPEES	ERSGYTGALSEPY	ELEAEEOGI	LNGCENGRCVRVQEG	YTCDCF DGYHLDMAKMT	CDVAVNECDELNNRMS	LCKNAKCI	NTEGSYKCLCLPGYVPSDK	PNYCTPLNTALNLEKDSDE
Lizard_L1TBP4	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Frog_(tropicalis)_L1TBP4	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Medaka_fish_L1TBP4	---	---	---	---	---	---	---	---	---	---	---	---	---	---

L1TBP3

Human_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Mouse_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Cow_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Dolphin_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Frog_(tropicalis)_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Zebra_fish_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Trout_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Pufferfish_L1TBP3	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Early L1TBP

Lancelet_L1TBP	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Acom_worm_L1TBP	---	---	---	---	---	---	---	---	---	---	---	---	---	---



“TYT” motif (residues 1547-1549) just prior to the three core cysteines. The tyrosine in particular is a bulky hydrophobic amino acid at a position that is not normally conserved (see TB2 in Figure 2.3 for comparison). There are also a particularly large number of conserved residues in the EGF3 domain, and in the eight residues at the end of the interdomain linker, just before the first cysteine of EGF3. In human LTBP1 this region has the sequence “FEELQAEE” (residues 1618-1625) and includes a number of conserved acidic and hydrophobic residues. Several acidic and hydrophobic residues are also conserved within the EGF3 domain. For example, a conserved “EG” motif (residues 1642-1643) is present at the end of the main β -hairpin, and an “LD” motif (residues 1654-1655) is present in the C-terminal loop of the domain, which is also conserved in LTBP3 and the early LTBP3-like sequences. While many of these potential interaction residues are shared between LTBP1, LTBP2, LTBP4, and even in some cases in the early LTBP3-like sequences, they are often not present in the LTBP3-like sequences. The LTBP3 C-terminus has been shown not to interact with the fibrillin1 N-terminus (Section 1.4.4), and so the lack of conservation in LTBP3 would still be consistent with a role for these residues in the LTBP-fibrillin interaction.

The cbEGF15 domain at the C-terminus also contains several conserved residues, although not as many as in EGF3. The residues required for calcium binding in this domain are conserved in all sequences, suggesting this domain is consistently present as a calcium binding EGF. Conversely there is a consistent absence of calcium binding motif residues in EGF3.

The interdomain linker between TB3 and EGF3 is quite poorly conserved and varies significantly in length between the different LTBP isoforms. The longest linkers are found in the LTBP4-like proteins, followed by those in LTBP2-like proteins, and then

the LTBP1-like proteins, with LTBP3-like proteins having the shortest linkers. One interesting feature of these regions is that they are quite rich in aromatic residues. When certain gaps are introduced into the alignment (as in Figure 2.7) there appears to be some moderate conservation of “YG” and “YXP” motifs in the linker region, which may be functionally significant. However the general shortage of conserved residues in the linker region, along with its significant variability in length would suggest it does not play a structural role. Despite this, all of the LTBPs examined do contain a linker region of some description, which even in LTBP3 is significantly longer than linker regions between fibrillin TB-cbEGF pairs¹². This might mean that there is some functional requirement for a linker region in between these two domains, leading to its overall conservation, even if its specific sequence features are not well preserved through evolution.

The presence of this linker between the TB3 and EGF3 domains might suggest a lack of precise interdomain packing between them. Although interdomain linker length has previously been shown to be unrelated to TB-cbEGF interactions, as measured by calcium affinity [12], those experiments were conducted with smaller linkers of between 15 and 23 amino acids. Additionally, the LTBP TB3 domain lacks a GY/F packing motif near its C-terminus, which is consistently absent across isoforms and species, and often replaced with charged or other hydrophobic residues. As this motif is absent, if an TB3 - EGF3 interaction did occur it would likely be via a novel mechanism. In contrast the GY/F packing motif at the end of EGF3 is highly conserved, and so is suggestive of a conserved packing interaction between EGF3 and cbEGF15.

¹² Between the last cysteine of TB3 and first cysteine of EGF3 there are 33 residues in human LTBP3 compared to around 15-23 residues between the equivalent cysteines in fibrillin TB-cbEGF pairs.

All these conserved features of the LTBP C-terminus may have potential implications for the molecular mechanism of the LTBP - fibrillin interaction. However their significance must be tested biochemically before any conclusions can be drawn.

The fibrillin N-terminus has been shown to interact with LTBPs in the region around its first hybrid domain (hyb1). Generally the molecular details of this interaction have not been explored, but it has been shown that one residue in particular, asparagine 164 in EGF3, has a significant affect on the affinity of the interaction with LTBP 4 if substituted with serine [11]. This substitution has also been associated with dominant ectopia lentis and Marfan syndrome [13]. Alignments show that this residue is conserved in species that also possess LTBPs (Figure 2.8), but is replaced by a structurally similar aspartate residue in arthropod and cnidarian fibrillins, and by a glycine residue in the sea squirt *Ciona intestinalis*. However the conservation of asparagine 164 in species with LTBPs would be consistent with a conserved role for this residue in the fibrillin-LTBP interaction.

Generally the entire region of fibrillin implicated in the LTBP interaction is highly conserved, and the EGF-EGF-hyb-cbEGF domain pattern is consistently present. While there are many conserved residues in the fibrillin N-terminus that are not likely to be of structural importance, the large number of protein interactions that have been linked to this region of fibrillin means this conservation does not necessarily suggest they are involved in the LTBP interaction. It is interesting to note though that there are a number of conserved basic arginine and lysine residues in EGF3 and hyb1, in contrast to the LTBP C-terminus, where there are a number of conserved acidic residues. This is also reflected in the pI of protein fragments from this region (Chapter 8 Table 8.5), which tend to be around pH 5 for LTBP C-terminal fragments, and pH 8

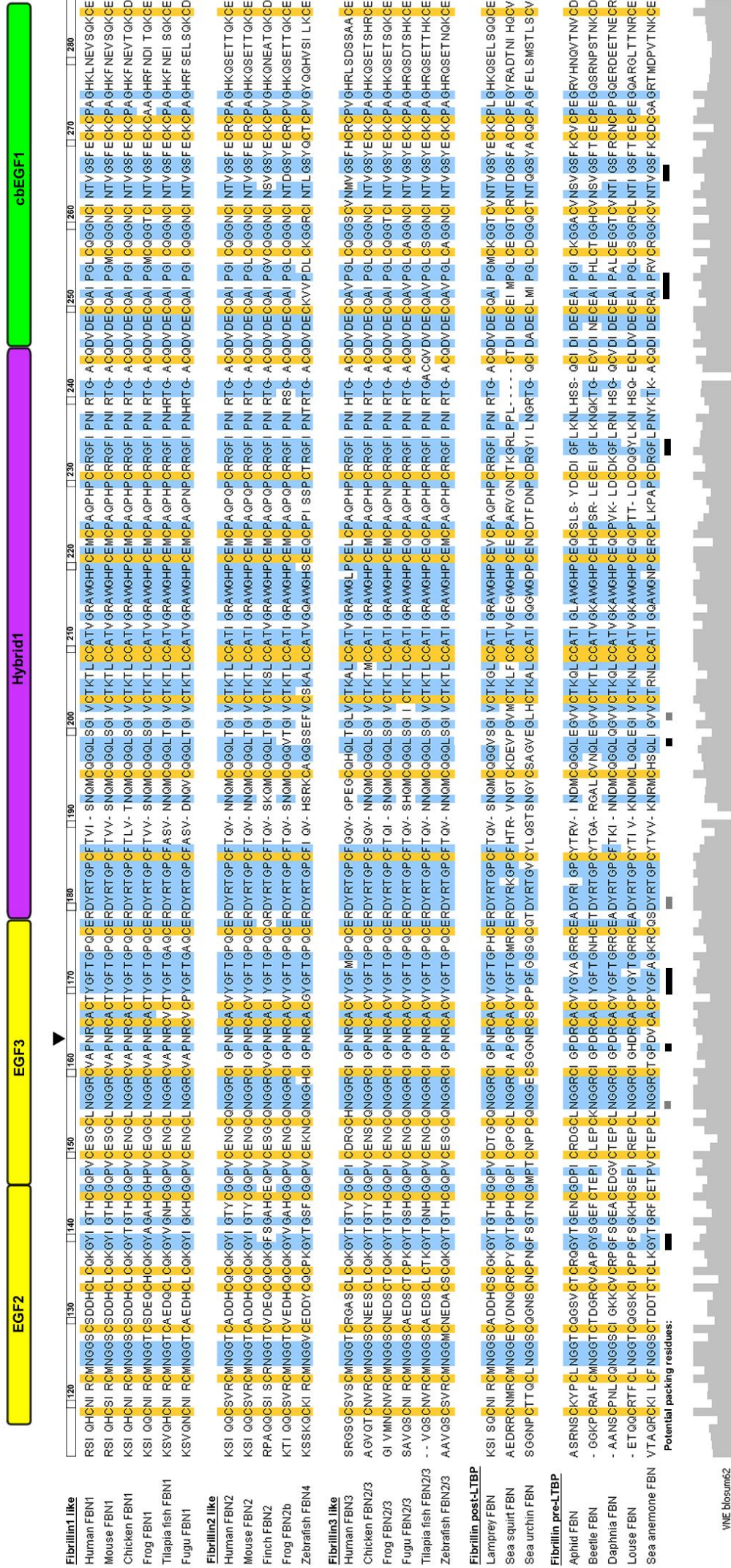


Figure 2.8 - Alignment of fibrillin hyb1 and adjacent domains

Alignment of the N-terminal region of fibrillin implicated in binding LTBP, including EGF2, EGF3, hyb1 and cbEGF1 (numbered according to human fibrillin1). Alignments were generated as described in Section 8.3. Residues present in >80% of all sequences are highlighted with a blue background, or yellow in the case of cysteines. Asparagine 164 implicated in LTBP binding is highlighted with an arrow head above the alignment, and conserved regions potentially involved in interdomain packing are highlighted by lines at the bottom of the alignment (grey boxes are implicated based on novel data in Chapter 3).

for fibrillin fragments. From this it could be hypothesised that an LTBP-fibrillin interaction may involve ionic interactions and the formation of salt bridges.

The structure of this region of fibrillin is also unknown, and the nature of the interdomain interfaces is of particular interest, as they may affect potential protein interaction sites. Residue positions that may be involved in interdomain interfaces, (based on interfaces in known structures [14-19]), are highlighted under the alignment in Figure 2.8. The packing consensus GY/F residues at the end of EGF2 are conserved in all fibrillins, and a proline is conserved at the end of the main β -hairpin of EGF3, which may contribute to the packing site with EGF2. At the C-terminus of the hyb1 domain the GY/F packing motif is also conserved, which may pack with the conserved VG motif on the main β -hairpin of cbEGF1. In the structure of fibrillin cbEGF9-hyb2-cbEGF10 [17], other extensive hydrophobic contacts were also seen between the hyb2 domain and cbEGF10, including contacts between isoleucine 867 in the second β -hairpin (C2-C3) of hyb2, and phenylalanine 917 in the initial α -helical region of cbEGF10. In the hyb1 and cbEGF1 sequences there are also consistently large hydrophobic residues in the equivalent positions (L199 and I253).

The crystal structure of cbEGF9-hyb2-cbEGF10 is less informative about the nature of the interdomain interaction between the upstream cbEGF9 domain and hyb2. While the structure does display limited contacts between these two domains, the side chains involved do not clearly suggest any hydrophobic or salt bridge contacts that would stabilise the interaction in solution, and could potentially be crystallisation artefacts. The C-terminus of cbEGF9 is also quite different in sequence to the C-terminus of EGF3, as cbEGF9 does not have a GY/F packing motif, while EGF3 does have this motif, which is well conserved, and is also preceded by an additional

conserved tyrosine. The N-terminus of hyb1 also exhibits a conserved “DYRTGP” motif, which not present in hyb2, and while there is little precedent to suggest this motif might be involved in interdomain packing it could be positioned in a suitable location for packing at the start of the hyb1 domain.

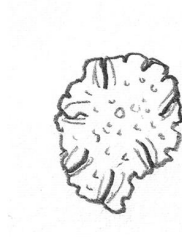
2.6- Conclusions

Analysis of fibrillin and LTBP-like sequences from the genomic record has provided a great deal of useful information, illuminating the general picture of TB domain protein evolution, and giving potential insights into their structure and interaction mechanisms. This information, together with information gleaned from the evolution of fibrillin interaction partners [2] is summarised in Figure 2.9, in order to provide a general picture of how the different roles of fibrillins and LTBPs may have evolved in the ECM.

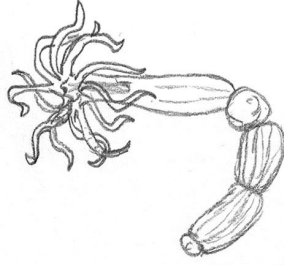
Fibrillin appears first and may initially play a structural role in forming microfibrils in the ECM. Later divergence of the LTBPs and TGF β then lead to the emergence of the fibrillin-LTBP-TGF β apparatus, creating a complex extracellular sensor, which is now critical to the regulation of a wide variety of human development and disease processes. Subsequent to the divergence of LTBPs, the fibrillins also acquired a vital role in directing elastic fiber assembly, and now perform this parallel structural function in a variety of tissues. In vertebrates, fibrillins and LTBPs have also diversified into several isoforms, potentially introducing redundancy, and allowing them to specialise in their functional niches. The release of more genome sequences, especially those of basal metazoans, should cast more light on the initial emergence of TB domain proteins and their early roles in the ECM.

Our common ancestor with:

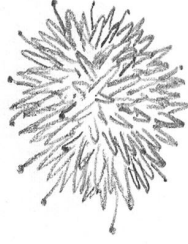
Unknown



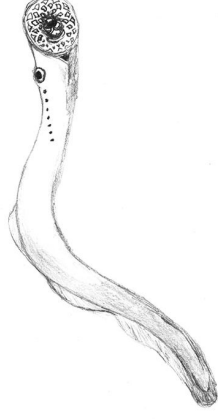
Anemone



Sea urchin



Lamprey



Fish

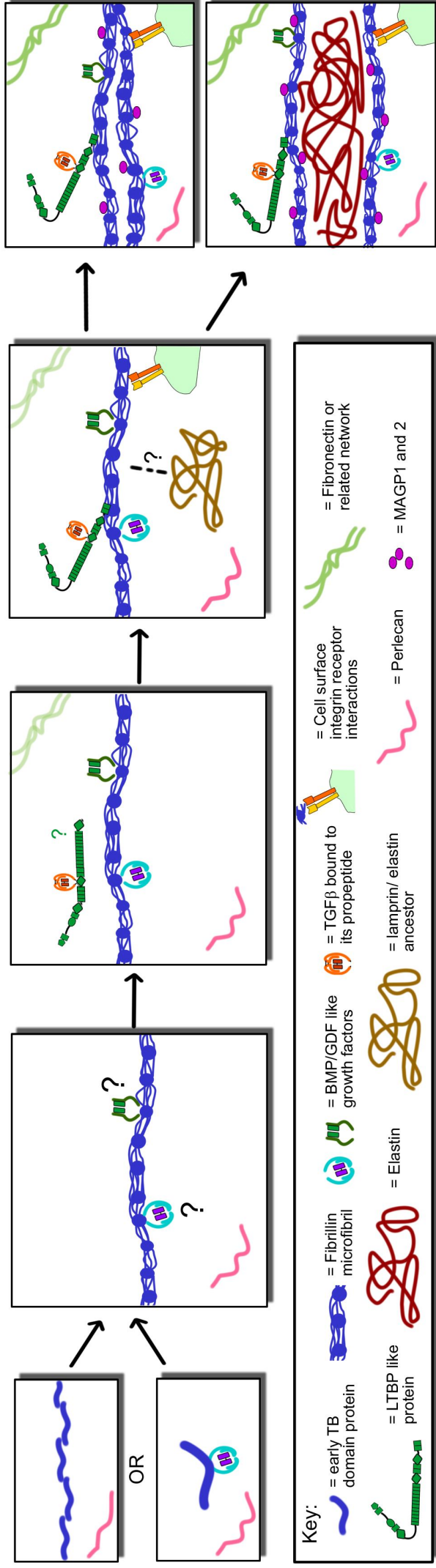


Figure 2.9- Summary of Fibrillin and LTBP evolution.

Adapted from Robertson et al [2]. Boxes contain possible functions acquired by fibrillin microfibrils at each stage of their evolution. The unknown organism represents the first organism to utilise TB domain proteins, where it is unclear if they might have first acquired a structural role, or initially played a role in growth factor binding.

The conservation of specific residues in the interacting regions of fibrillin and LTBP also give interesting hints into conserved elements of their structure and the residues that may be important for the LTBP - fibrillin interaction. Both of these aspects will be explored further in the following chapters.

References

1. Letunic, I., T. Doerks, and P. Bork, *SMART 6: recent updates and new developments*. Nucleic Acids Res, 2009. **37**(Database issue): p. D229-32.
2. Robertson, I., S. Jensen, and P. Handford, *TB domain proteins: evolutionary insights into the multifaceted roles of fibrillins and LTBPs*. Biochem J, 2011. **433**(2): p. 263-76.
3. Ryan, J.F., et al., *The homeodomain complement of the ctenophore Mnemiopsis leidyi suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa*. Evodevo, 2010. **1**(1): p. 9.
4. Peterson, K.J. and N.J. Butterfield, *Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record*. Proc Natl Acad Sci U S A, 2005. **102**(27): p. 9547-52.
5. Piha-Gossack, A., W. Sossin, and D.P. Reinhardt, *The evolution of extracellular fibrillins and their functional domains*. PLoS One, 2012. **7**(3): p. e33560.
6. Reber-Muller, S., et al., *An extracellular matrix protein of jellyfish homologous to mammalian fibrillins forms different fibrils depending on the life stage of the animal*. Dev Biol, 1995. **169**(2): p. 662-72.
7. Thompson, T.B., et al., *The structure of the follistatin:activin complex reveals antagonism of both type I and type II receptor binding*. Dev Cell, 2005. **9**(4): p. 535-43.
8. Coghlan, A. and K.H. Wolfe, *Fourfold faster rate of genome rearrangement in nematodes than in Drosophila*. Genome Res, 2002. **12**(6): p. 857-67.
9. Ranz, J.M., F. Casals, and A. Ruiz, *How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus Drosophila*. Genome Res, 2001. **11**(2): p. 230-9.
10. Isogai, Z., et al., *Latent transforming growth factor beta-binding protein 1 interacts with fibrillin and is a microfibril-associated protein*. J Biol Chem, 2003. **278**(4): p. 2750-7.
11. Ono, R.N., et al., *Latent transforming growth factor beta-binding proteins and fibulins compete for fibrillin-1 and exhibit exquisite specificities in binding sites*. J Biol Chem, 2009. **284**(25): p. 16872-81.
12. Jensen, S.A., et al., *Ca²⁺-dependent interface formation in fibrillin-1*. J Biol Chem, 2005. **280**(14): p. 14076-84.
13. Comeglio, P., et al., *Identification of FBN1 gene mutations in patients with ectopia lentis and marfanoid habitus*. Br J Ophthalmol, 2002. **86**(12): p. 1359-62.
14. Lee, S.S., et al., *Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization*. Structure, 2004. **12**(4): p. 717-29.
15. Yuan, X., et al., *Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils*. EMBO J, 1997. **16**(22): p. 6659-66.
16. Lack, J., et al., *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, 2003. **334**(2): p. 281-91.
17. Jensen, S.A., et al., *Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins*. Structure, 2009. **17**(5): p. 759-68.
18. Smallridge, R.S., et al., *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, 2003. **278**(14): p. 12199-206.
19. Downing, A.K., et al., *Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders*. Cell, 1996. **85**(4): p. 597-605.
20. Chen, Y., et al., *Amino acid requirements for formation of the TGF-beta-latent TGF-beta binding protein complexes*. J Mol Biol, 2005. **345**(1): p. 175-86.

Chapter 3 - Structure and dynamics of the fibrillin N-terminus

3.1 Introduction

The N-terminus of fibrillin, especially the first hybrid domain, is a region of particular interest in microfibril biology, as it has been shown to be important for binding both LTBP1 and 4, as well as fibulins and other factors [1-3] (Figure 1.7). It has also been shown to contain an additional cysteine (cysteine 204), which has been predicted to be solvent exposed [4], and may be involved in microfibril assembly [5]. This region is also highly conserved, as discussed in Chapter 2, underlining its potential biological importance. A structural understanding of this region is crucial for developing a detailed understanding of the molecular basis for the interactions it mediates.

3.2 Production of recombinant fibrillin

In order to examine this region, the DNA coding for the EGF2-EGF3-hyb1-cbEGF1 (FBN^{e2cb1}) domains of fibrillin 1 was cloned into the prokaryotic expression construct pQE30 (see materials and methods Section 8.7). To remove the free sulphydryl introduced by cysteine 204, which could interfere with *in vitro* refolding, the codon of this residue was altered by site-directed mutagenesis to a codon of a serine residue (Section 8.8). The context of this fragment in the whole fibrillin molecule is shown in Figure 3.1, along with the overlapping FBN^{NE3} construct, which is also referred to in this chapter. Successful cloning and mutagenesis was confirmed by DNA sequencing (Section 8.7), and the protein was then expressed in *E. coli* BL21 cells as a His-tagged fusion protein (Section 8.9).

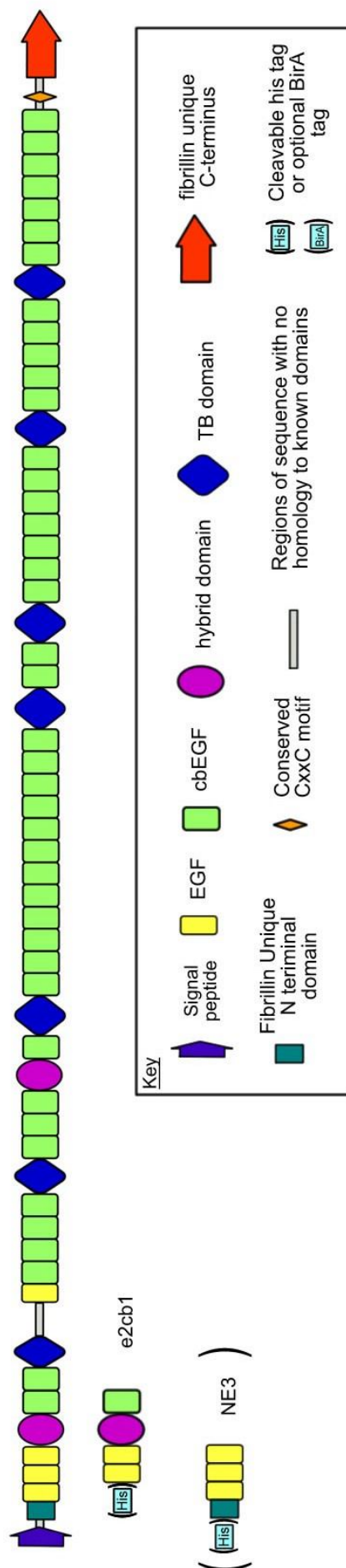


Figure 3.1 - Constructs used in structural studies of the fibrillin-1 N-terminus.

This chapter focuses on structural studies of the FBN^{e2cb1} fragment (fibrillin-1 residues 113-287), but data from the FBN^{NE3} fragment is also referred to at several points.

Inclusion body extraction (Section 8.12) was found to give the highest yields for this construct when protein was expressed in both rich and minimal media, and so this extraction technique was used to produce the majority of protein for this study. Guanidine-HCl lysis followed by nickel column purification was also found to be a viable extraction method (Section 8.11).

Protein was refolded by an oxido-shuffling system (Section 8.15), and purified by reverse phase HPLC. The protein's His-tag was then removed by factor Xa cleavage (Section 8.16), and the resulting protein was purified by ion exchange FPLC and HPLC (Section 8.17). Example chromatograms are shown in Figure 3.2.

3.3 Initial characterisation and assessment of fold

The mass and purity of the final FBN^{e2cb1} sample was confirmed by reducing and non-reducing SDS-PAGE (Section 8.19) and ESI mass spectrometry (Section 8.20), (Figure 3.3 A and B). While SDS-PAGE and mass spectrometry demonstrated the protein was of the correct size, and appeared homogenous, this alone was not necessarily confirmation that all domains of FBN^{e2cb1} were correctly folded. As the protein has no predicted enzymatic activity this could not be assayed to confirm fold, but several other techniques were employed as preliminary tests of folding, these are described below :

1) Calcium chelation

As the protein is predicted to contain a calcium-binding EGF domain it was possible to use a calcium chelation assay to determine whether the refolded protein binds calcium (Section 8.22). The calcium affinity of the hyb1-cbEGF1 domain pair had previously been measured, and was observed to have a high calcium affinity (K_d 8.7 nM) [4], this was also reflected in chelation studies of FBN^{e2cb1} with Quin2 chelator (Figure 3.3 C),

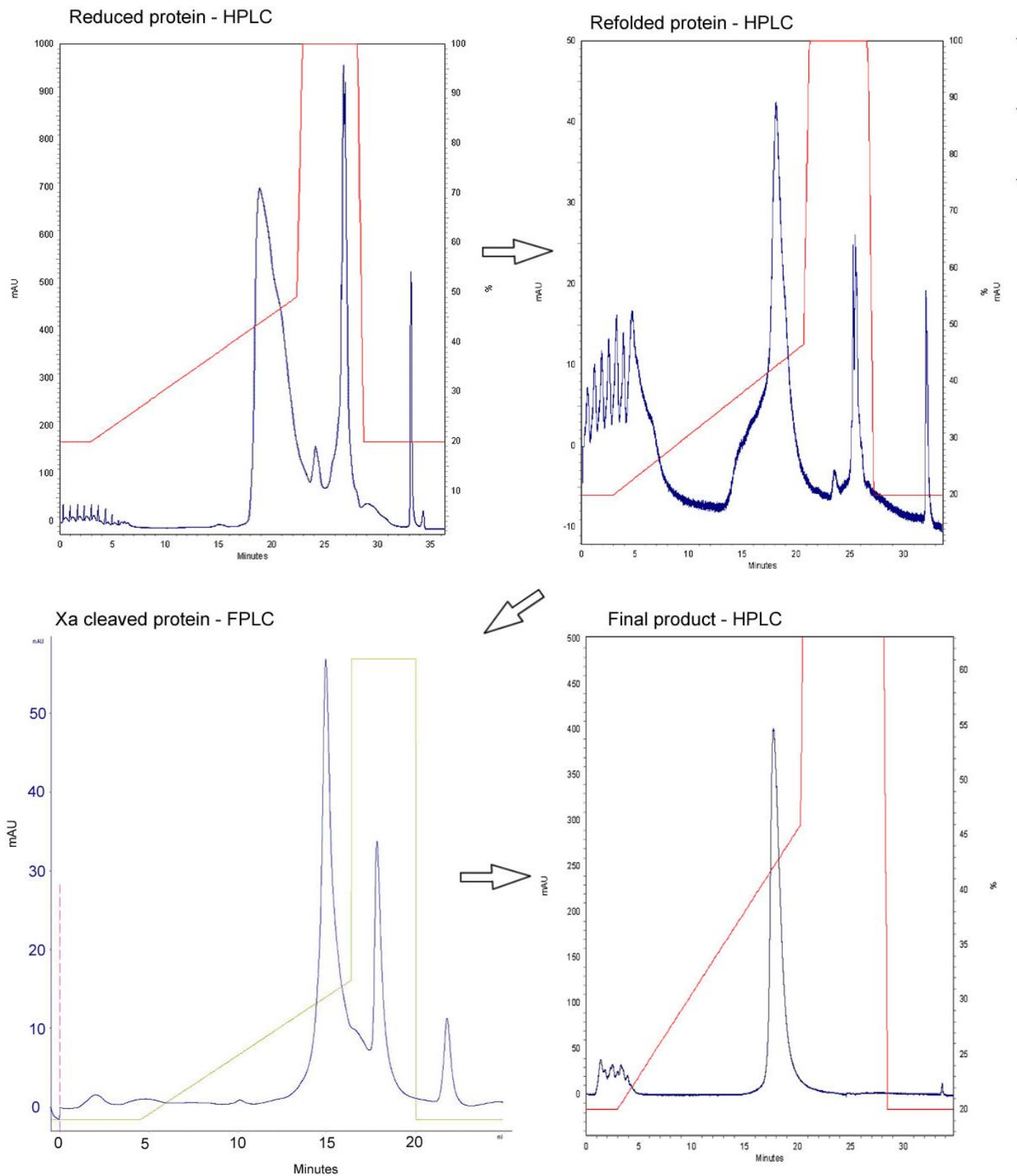


Figure 3.2 - Chromatograms obtained during the purification of FBN^{e2cb1}

Example chromatograms (blue) are shown here from each stage of the purification process detailed in Section 8.9 - 8.17. Red and green lines represent A to B buffer gradients. HPLC was carried out with a C8 reverse phase column, from which proteins were eluted by an acetonitrile gradient, (based on hydrophobicity). FPLC was carried out with a ion exchange "MonoS" (GE healthcare) column with cationic surface chemistry, proteins were eluted from this column with an NaCl gradient.

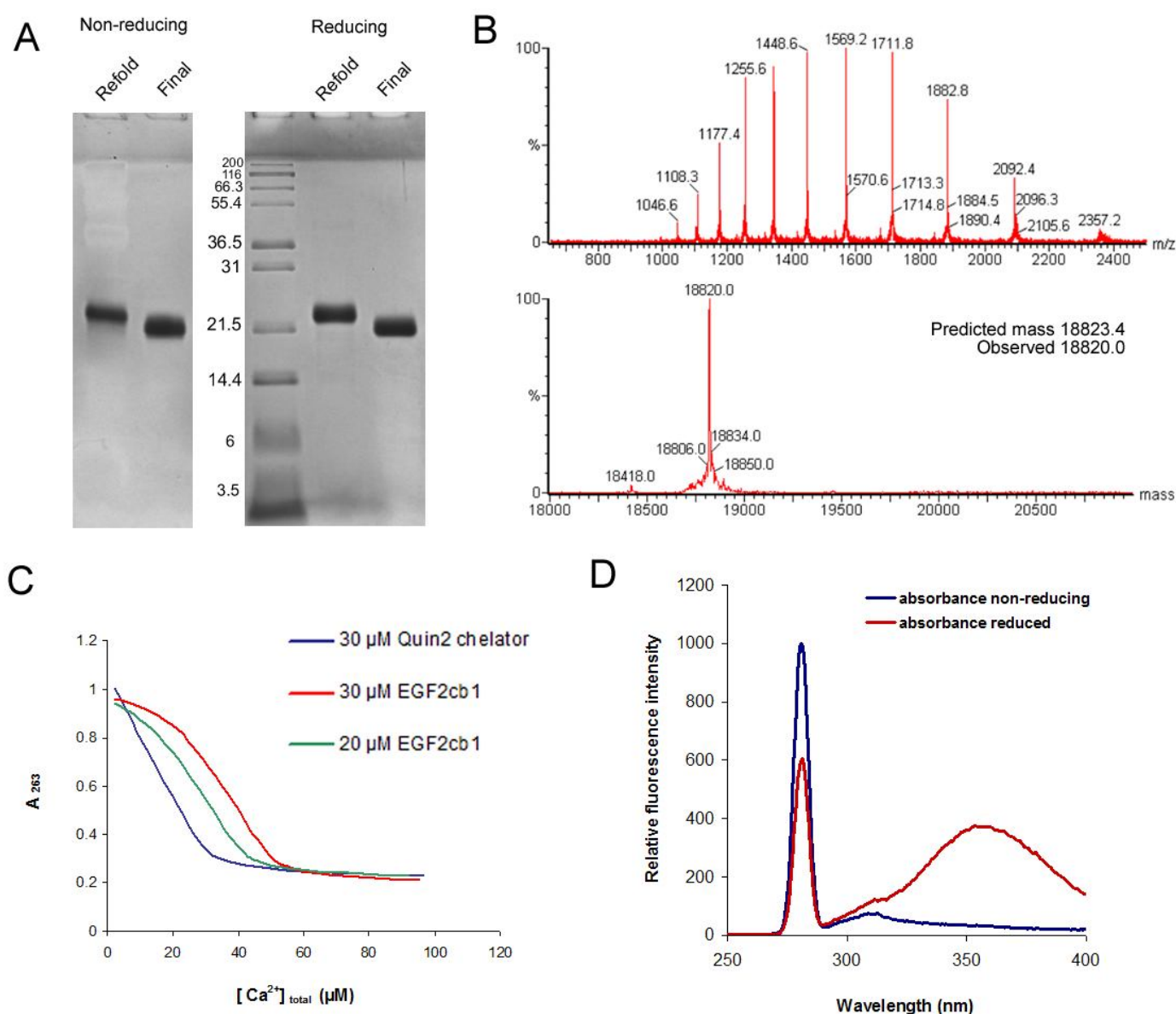


Figure 3.3 - Initial characterisation of FBN^{e2cb1} mass and fold

A) SDS-PAGE analysis of FBN^{e2cb1} purified directly from refold by HPLC ("Refold") and after Xa cleavage of His tag and further purification ("Final"). Samples run under both reducing and non-reducing conditions, (samples run on the same gel). **B)** ESI- mass spectrometry results of FBN^{e2cb1} with the His-tag removed. Mass/charge series is also shown to demonstrate homogeneous material **C)** Calcium chelation experiments with FBN^{e2cb1} demonstrating competition for calcium binding with the high affinity Quin2 chelator. **D)** Fluorimetry experiments demonstrating quenching of tryptophan fluorescence in FBN^{e2cb1} under non-reducing conditions, although fluorescence is seen when the disulphides in FBN^{e2cb1} are broken by reducing agents. Data for panels B, C, and D in this figure were collected by part II student Isabelle Osuch.

which is a high affinity calcium chelator ($K_d = 0.35 \mu\text{M}$). FBN^{e2cb1} was able to compete with Quin2 for calcium, as demonstrated by the slower loss of chelator absorbance seen when FBN^{e2cb1} was present during titration with calcium (Figure 3.3C).

2) Fluorimetry

FBN^{e2cb1} also contains a tryptophan residue, which is predicted to form the hydrophobic core of the hybrid 1 domain. The intrinsic fluorescence of tryptophan can be exploited to examine fold, as fluorescence quenching can be observed when a fluorophore is buried within a protein instead of being exposed to solvent. Figure 3.3 D demonstrates a clear absence of tryptophan fluorescence in a FBN^{e2cb1} sample, but significant fluorescence at 350 nm is visible when the same protein sample is reduced (Section 8.23). This suggests that the tryptophan residue is buried within the protein as predicted, but becomes exposed when the FBN^{e2cb1} disulphide bonds are broken by reducing agents.

3) 2D NMR

The highest resolution evidence that the FBN^{e2cb1} fragment was correctly folded came from two dimensional ¹H NMR experiments. Figure 3.4 shows selected regions of 2D NOESY spectra of FBN^{e2cb1} (Section 8.28), collected in both the presence and absence of calcium. The dispersed nature of the peaks in the spectra is suggestive of a structured protein, and several cross-peaks in the 4.8-5.9 ppm region correspond to downfield shifted H^α protons (Figure 3.4 A), which are indicative of anti-parallel β-sheet structure. Additionally, various peaks in the spectra appear to be calcium dependant, which provides further evidence of a correctly folded cbEGF that can specifically bind calcium. Later assignment showed that these peaks are specific to cbEGF1 residues, or to hyb1 residues that pack with the cbEGF1 domain.

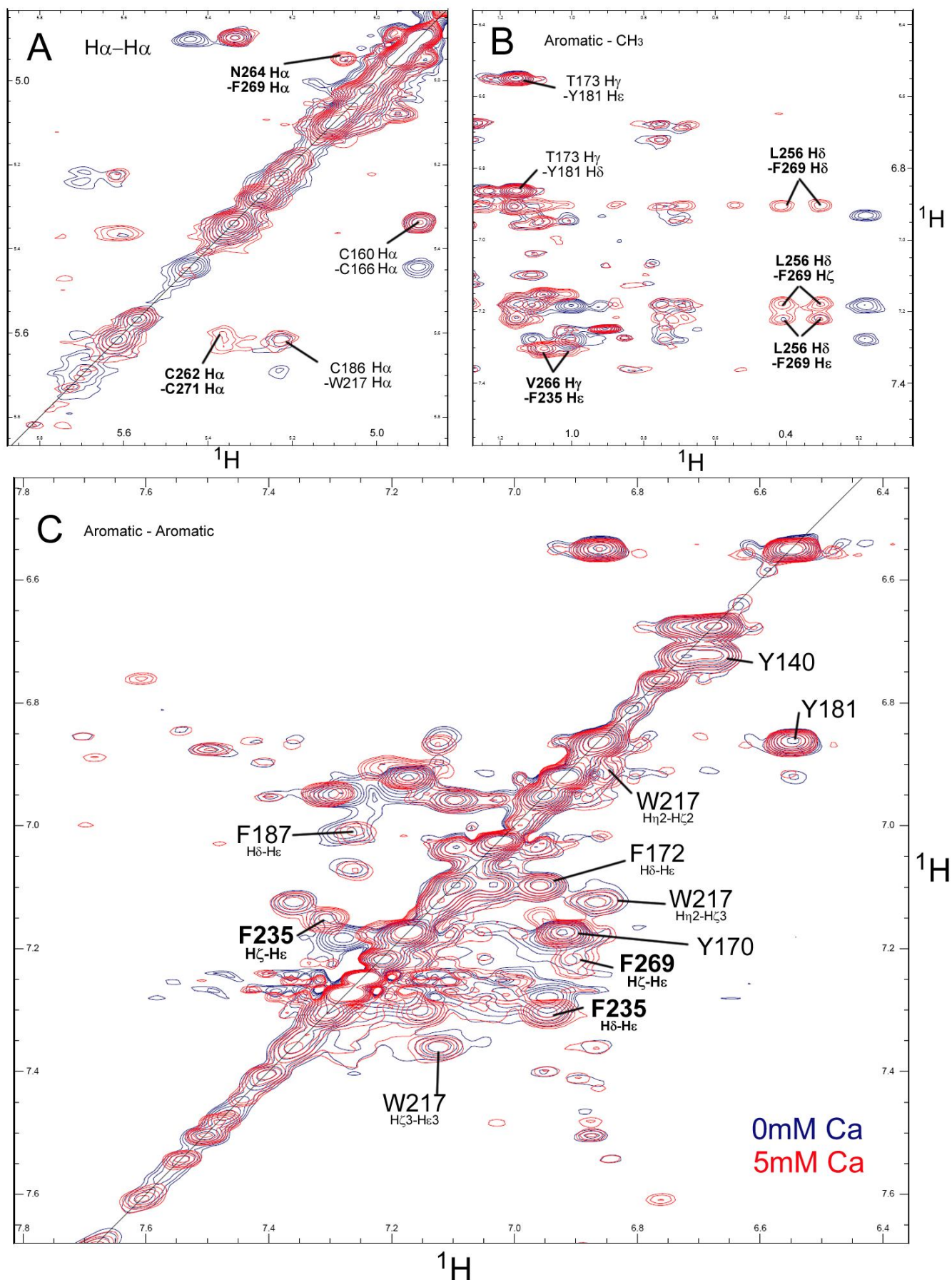


Figure 3.4 - 750 MHz 2D NOESY spectra demonstrating folding and calcium dependent peaks in FBN^{e2cb1}

Selected regions of NOESY spectra collected in D₂O at pH 5.4 with or without 5 mM calcium. Some assignments obtained later are also highlighted. Calcium sensitive assignments are shown in bold **A**) A number of NOE's between downfield shifted $H\alpha$ protons, which are generally indicative of anti-parallel β -sheet structure. **B**) NOEs between aromatic protons and up-field shifted methyl resonances. **C**) Aromatic - aromatic NOEs display a dispersed distribution and range of peak intensity.

Also when 2D NOESY spectra were compared with those of an overlapping fragment of fibrillin, FBN^{NE3} (Figure 3.1), many peaks were seen to overlay, suggesting similar domain folds for EGF2 and EGF3 in both FBN^{NE3} and FBN^{e2cb1} constructs (data not shown).

3.4 Backbone and side chain resonance assignment

In order to investigate the structure, dynamics and specific interactions of the FBN^{e2cb1} protein fragment, backbone and side chain resonance assignment was carried out. Due to the relatively large size of the protein (~19 kDa), isotopic labelling was required for assignment, which was achieved by protein expression in minimal medium (Section 8.10). Initially a sample labelled with ¹⁵N alone was produced, but the quality of the spectra were only sufficient to assign a limited number of resonances, and complete double labelling with ¹³C and ¹⁵N was necessary for the majority of assignments.

Figure 3.5 shows an annotated ¹⁵N-¹H - HSQC summarising the amide assignments for FBN^{e2cb1}. 3D ¹⁵N-edited TOCSY and NOESY spectra, and ¹³C HNCA, CBCANH, and CBCACONH spectra were all collected for sequential assignment of amides, C^α and C^β carbons, which were further confirmed with HNCOC and HNCACOC experiments (see Section 8.29 for experiment details). Some example strips from ¹⁵N-edited NOESY and TOCSY spectra used for sequential assignment are shown in Figure 3.6, and strips from ¹³C-edited HNCA, CBCACONH, HNCOC and HNCACOC experiments for the same region¹ are shown in Figure 3.7.

As the structure determination of FBN^{e2cb1} was of interest, side chain assignment was also carried out using the same ¹⁵N-edited TOCSY and NOESY experiments, a 3D

¹ This region was presented as the peaks had good intensity and were generally well separated from other peaks. This was not necessarily reflected in other regions of the protein.

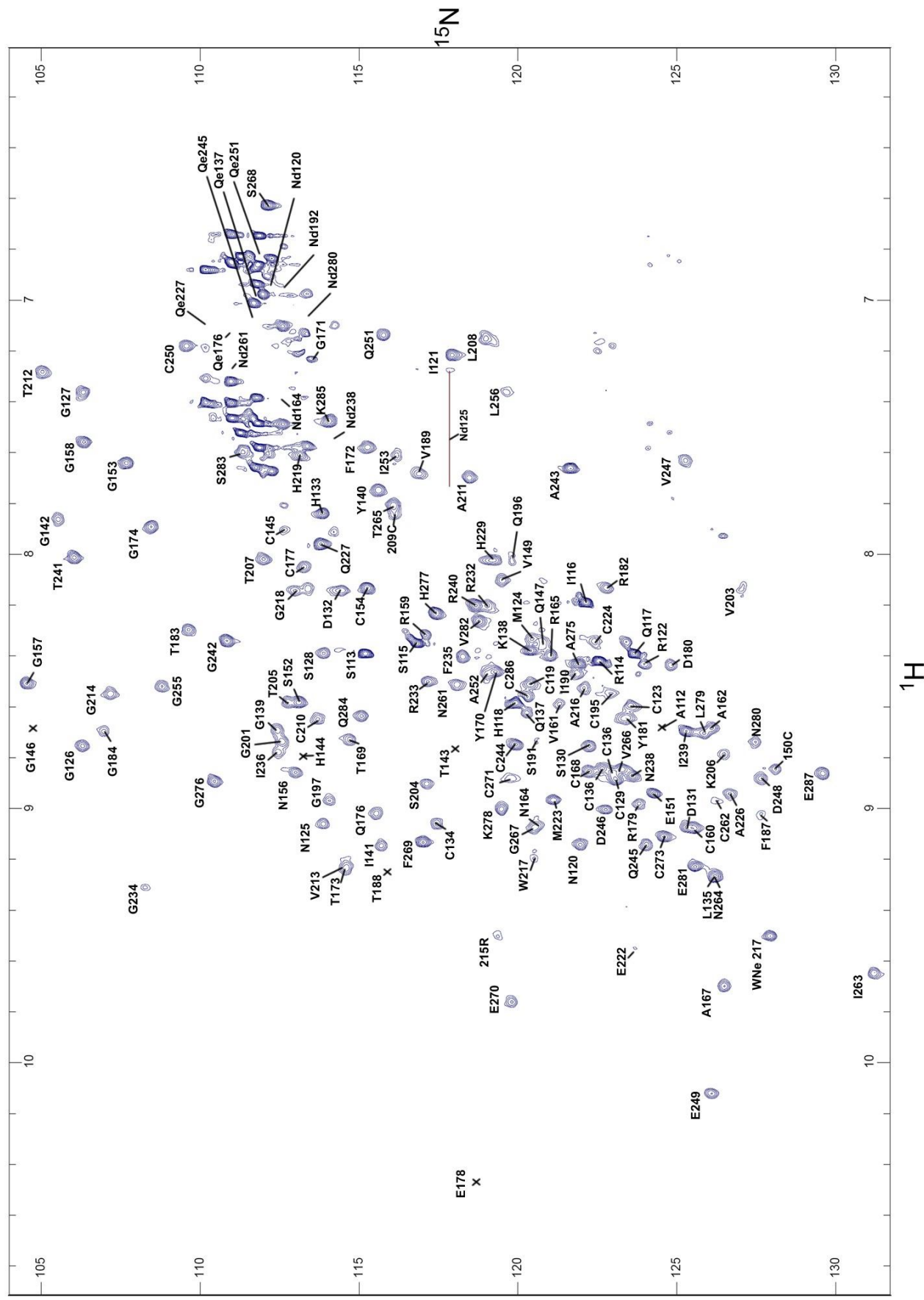


Figure 3.5 - 950 MHz HSQC spectrum showing peak assignments for FBN^{e2cb1}
HSQC collected in 95% H₂O: 5% D₂O at pH 5.4 with 5 mM CaCl₂. Peak assignments are shown above their respective peaks, except where otherwise indicated by lines. Small x's indicate peaks that were too weak to visualise in this spectrum.

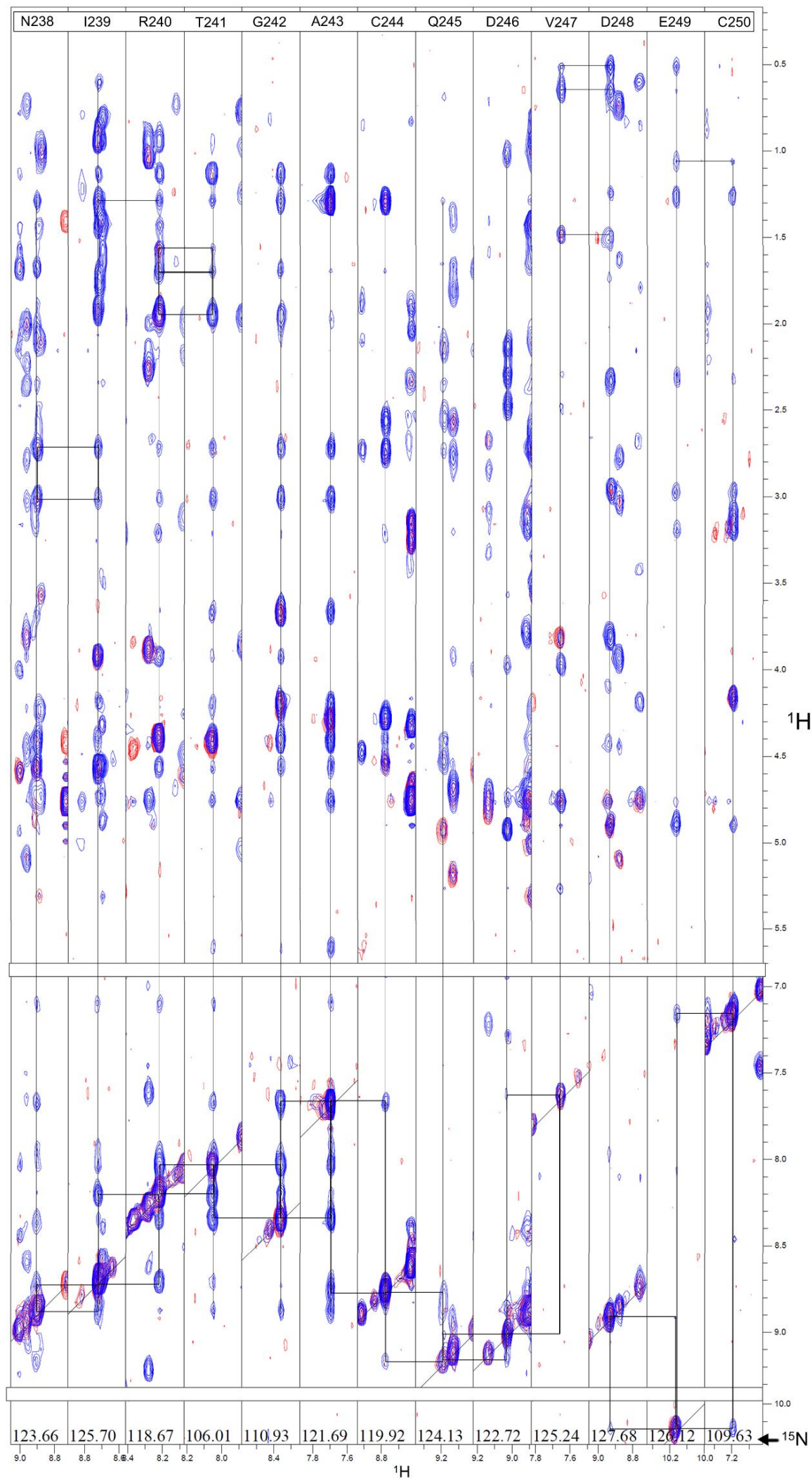


Figure 3.6 - Selected strips from 3D ^{15}N -edited NOESY and TOCSY spectra

Sequential strips assigned to residues in the C-terminal β -hairpin of hyb1, NOESY peaks are shown in blue while TOCSY peaks are shown in red. Assigned residues are given at the top of each strip, and the position in the ^{15}N dimension at the bottom. Black lines highlight sequential NH, H α , H β , and other connections. The good signal to noise seen with peaks shown here was not representative of peaks obtained for many other regions of the protein. (Spectra collected from the same sample and with the same instrument as the HSQC in Figure 3.5)

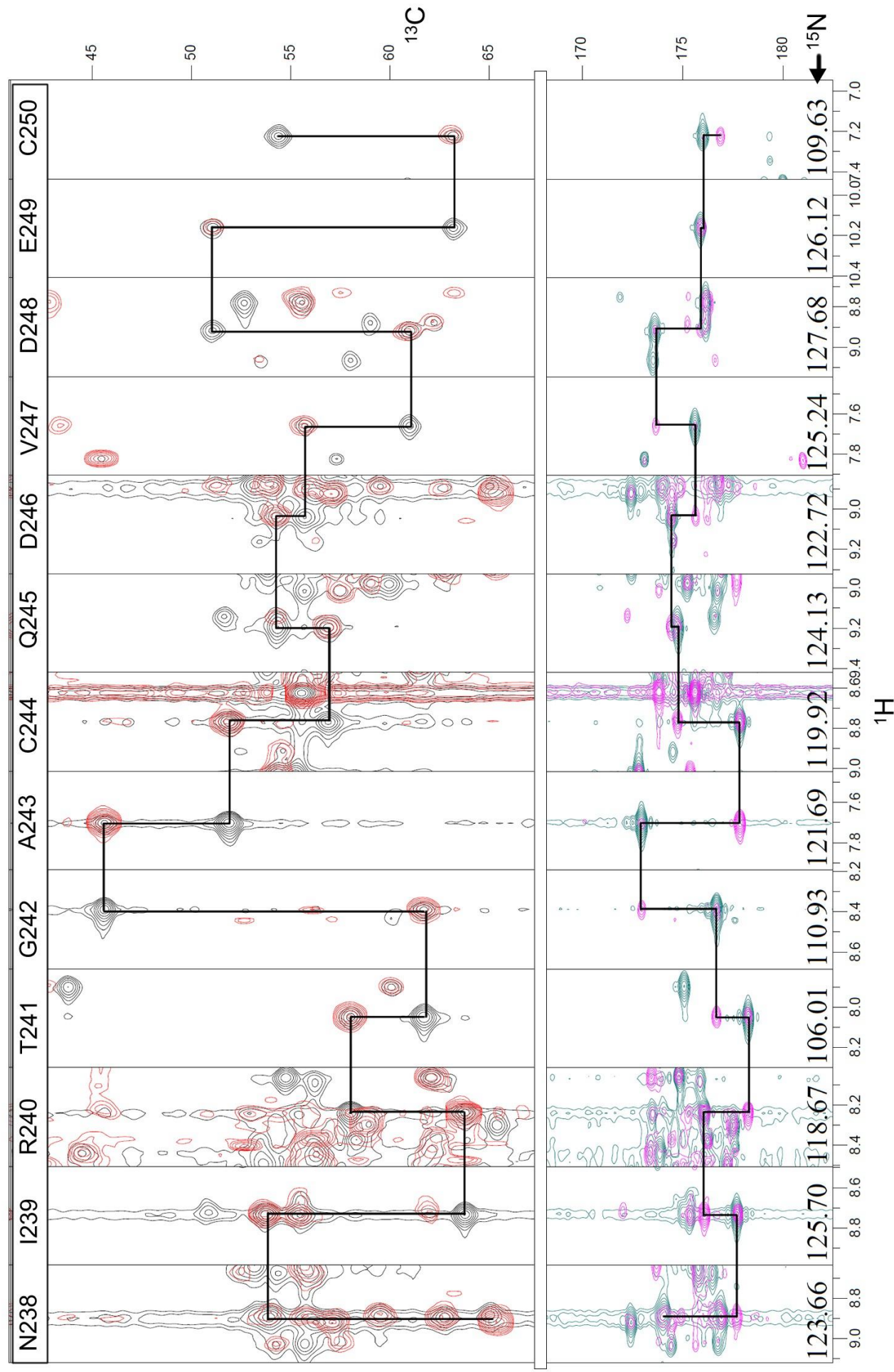


Figure 3.7 - Selected strips from 3D HNCA, CBCACONH, HNCACO, and HNCACO spectra of FBN^{e2cb1}

Strips of the same region of FBN^{e2cb1} examined in Figure 3.6, this time showing HNCA peaks (red), HNCACO peaks (black), CBCACONH peaks (teal), and HNCACO peaks (pink). Black lines highlight sequential C α and carbonyl connections. Spectra collected under the same conditions as in Figure 3.5 and 3.6, but using a Bruker 500 MHz spectrometer.

HBHACONH experiment for specific assignment of H^α and H^β protons, a 3D ^{13}C -edited HCCH-TOCSY experiment, and also a 3D ^{13}C -edited NOESY experiment for assignment of other side-chain resonances. All spectra for assignment were collected at pH 5.4, with the temperature set to 25 °C, and with 5mM calcium present. All assigned chemical shifts are summarised in Table A1.1 in Appendix 1.

Some residues gave peaks with extremely weak intensity (Appendix 1 Figure A1.1), and so, even with double labelling, it was not possible to get complete backbone and side chain assignments. In several cases it appeared peaks were absent, or masked by other peaks of greater intensity. However, despite the challenging nature of the protein, with perseverance it was possible to assign 90% of backbone amides and 87% of side chain protons (Appendix 1 Table A1.6).

Comparison of the H^α , H^β , and amide assignments of $\text{FBN}^{\text{e2cb1}}$ with those obtained for the overlapping FBN^{NE3} fragment at a similar pH (data kindly provided by D. Yadin) demonstrated similar chemical shifts for the majority of EGF2-3 domain resonances (Figure 3.8). This supports both domains adopting similar folds in FBN^{NE3} and $\text{FBN}^{\text{e2cb1}}$ constructs. However significant changes in peak position were seen in some areas, especially with peaks corresponding to the C-terminal loop of EGF3, such as in the “GF” motif, with the H^α protons of glycine 171 exhibiting a significant up-field shift in $\text{FBN}^{\text{e2cb1}}$, which could be due to ring current effects from packing with an aromatic side chain, induced by the presence of hyb1. The equivalent motif forms packing interactions with downstream domains in the structures of cbEGF12-13, cbEGF32-34, and between cbEGF22-TB4 [6-8] (Section 1.2.2 and 1.2.4).

The H^β protons of asparagine 156 also demonstrate a significant up-field shift in $\text{FBN}^{\text{e2cb1}}$, which is also suggestive of aromatic packing. While not at the C-terminus of the EGF3 sequence, this asparagine is predicted to be on a loop that would be close

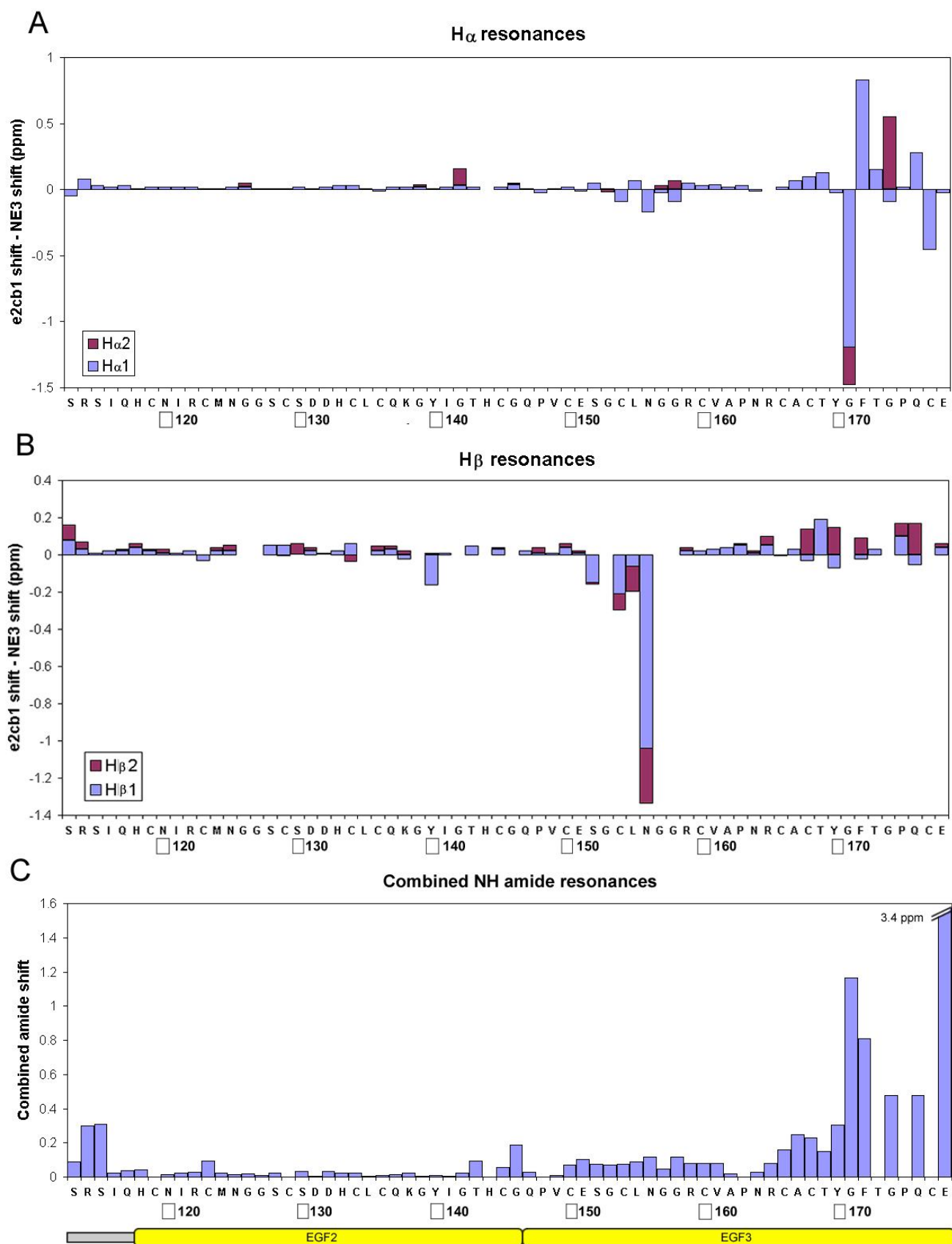


Figure 3.8 - Chemical shift differences between FBN^{e2cb1} and FBN^{NE3}

A) Difference in H α peak position (for glycines H α 1 corresponds to the more down-field resonance and H α 2 to the more up-field resonance). **B)** Difference in H β peak position (H β 1 corresponds to the more down-field resonance and H β 2 to the more up-field resonance). **C)** Combined amide shift generated by adding the difference of the peak position in the ^1H dimension, to the difference in position in ^{15}N dimension divided by 6.

to the C-terminus of the EGF3 domain when folded. Generally the differences in EGF3 peak positions seen between FBN^{NE3} and FBN^{e2cb1} strongly suggests domain interface formation between EGF3 and hyb1 in FBN^{e2cb1}.

3.5 Fast timescale Dynamics

To examine the fast time scale dynamics of FBN^{e2cb1}, a heteronuclear NOE experiment [15] was carried out (Figure 3.9), which gives information on regions that exhibit dynamics in the sub-nanosecond timescale. Only two regions of FBN^{e2cb1} displayed a particularly low heteronuclear NOE ratio, indicative of flexibility; the first of which corresponds to the seven residues at the N-terminus of the fragment, prior to the first cysteine of EGF2. This region has also been shown to form a flexible linker between EGF1 and EGF2 in the FBN^{NE3} construct (D. Yadin, personal communication), and so this result is consistent with this region being unstructured. The second flexible region corresponds to a loop at the C-terminus of the hybrid 1 domain, but it precedes the last cysteine of this domain, and so does not have the potential to act as a flexible interdomain linker. Generally the heteronuclear NOE values for FBN^{e2cb1} do not suggest any flexible linkers between any of the four domains present, and support the molecule forming a single rigid unit.

3.6 Calcium binding

Specific calcium-dependent changes were seen in HSQC spectra of FBN^{e2cb1} (Figure 3.10 A), as seen previously with 2D NOESY spectra (Figure 3.4). Residues with calcium-sensitive peaks are highlighted in the sequence and preliminary structure of FBN^{e2cb1} in Figure 3.10 B, and in the domain schematic in Figure 3.11 A. As well as peaks from the cbEGF, a number of peaks in the hybrid domain are also lost or displaced in the absence of calcium, such as serine 204, threonine 205, threonine 212, glycine 234, phenylalanine 235, and various residues in the C-terminal β -hairpin

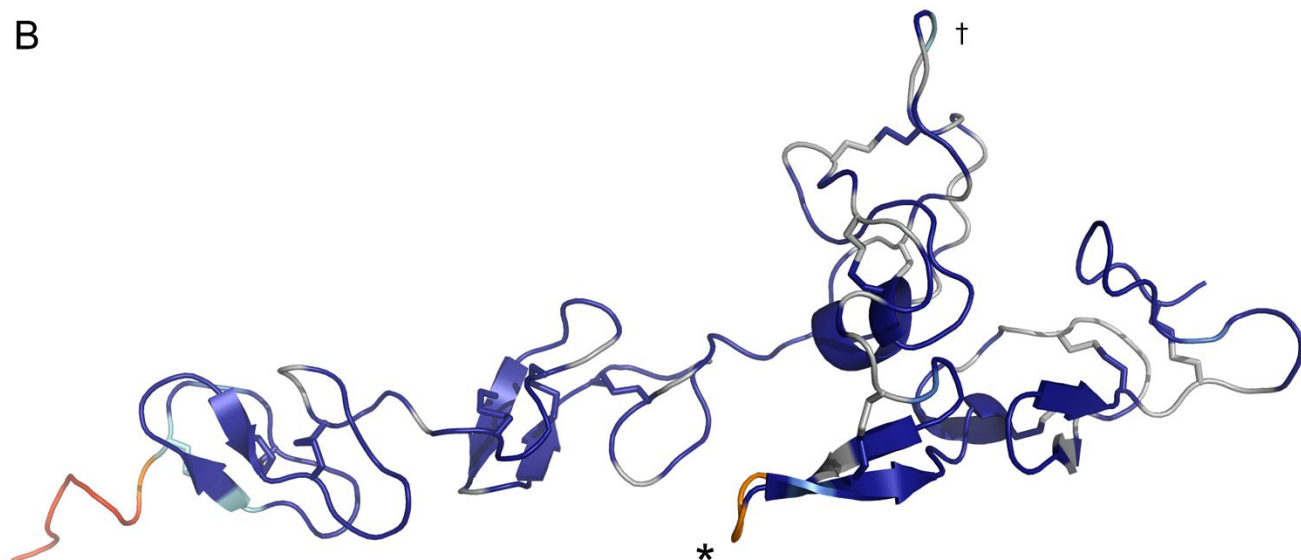
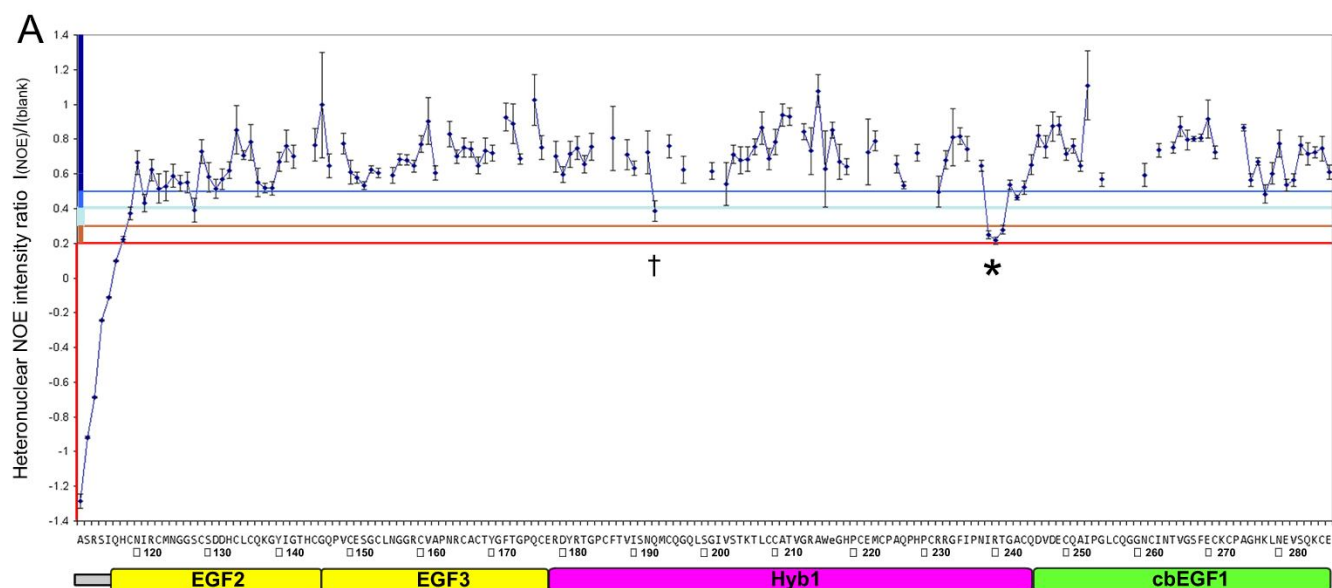


Figure 3.9 - Heteronuclear NOE experiment revealing fast time scale dynamics of FBN^{e2cb1} residues
A) Heteronuclear NOE values plotted as a function of sequence, where a low heteronuclear NOE intensity ratio demonstrates flexibility. Error bars represent the standard error introduced by background noise (Section 8.33). Coloured lines represent the arbitrary cut offs used to generate 3.9 B. **B)** Dynamics of FBN^{e2cb1} residues mapped onto a putative structure of this fragment (see section 3.12 (structure 40-3-3)). Flexible residues with ratios <0.2 are coloured red, those with ratios between 0.2 - 0.3 are coloured orange, those between 0.3 and 0.4 are coloured light blue, between 0.4 and 0.5 mid-blue, and those greater than 0.5 coloured dark blue. Residues that could not be assigned, or with peaks indistinguishable from noise in these experiments, are coloured light grey. Disulphide side chains are also included as stick representations. * and † signs highlight small flexible loops in the hybrid domain

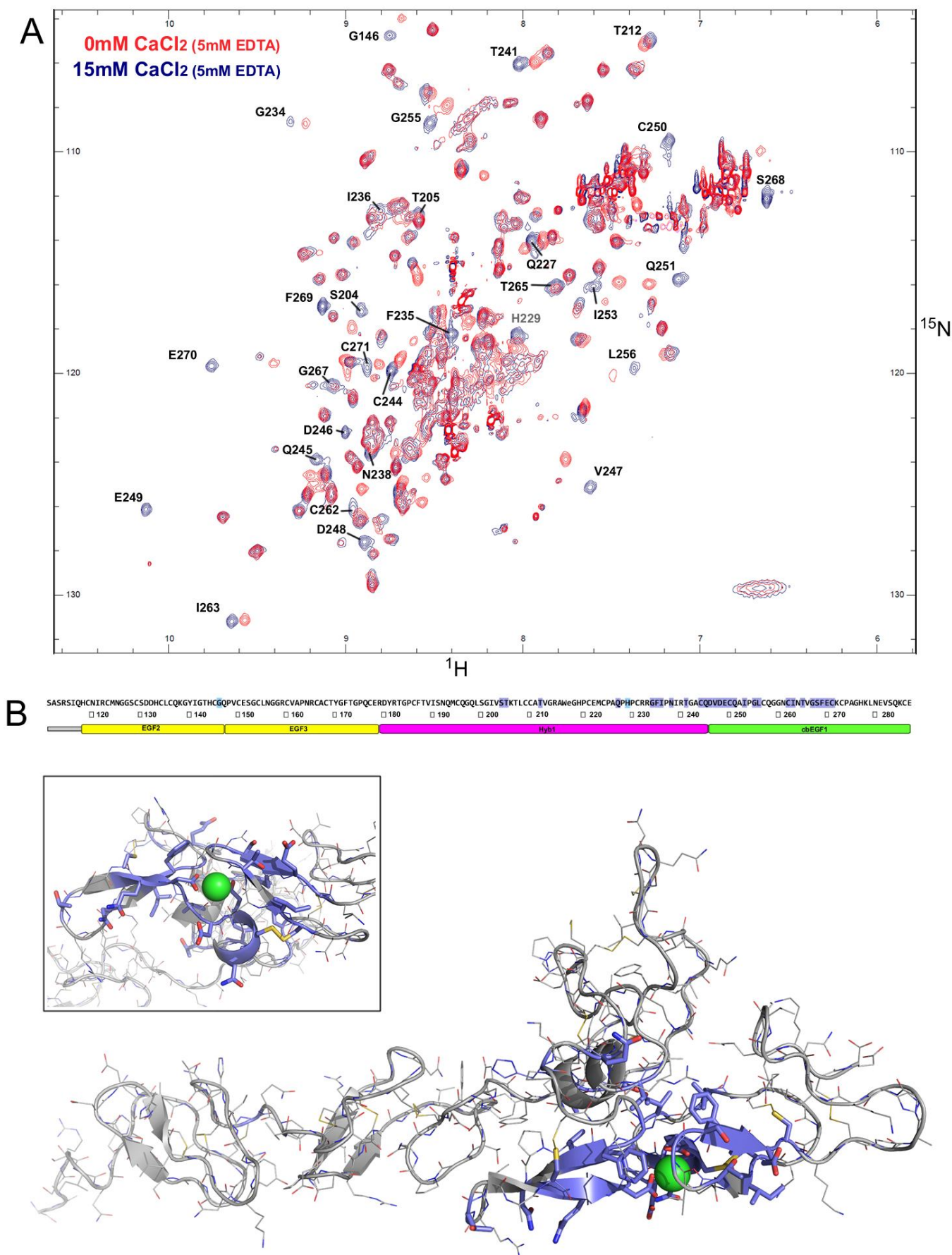
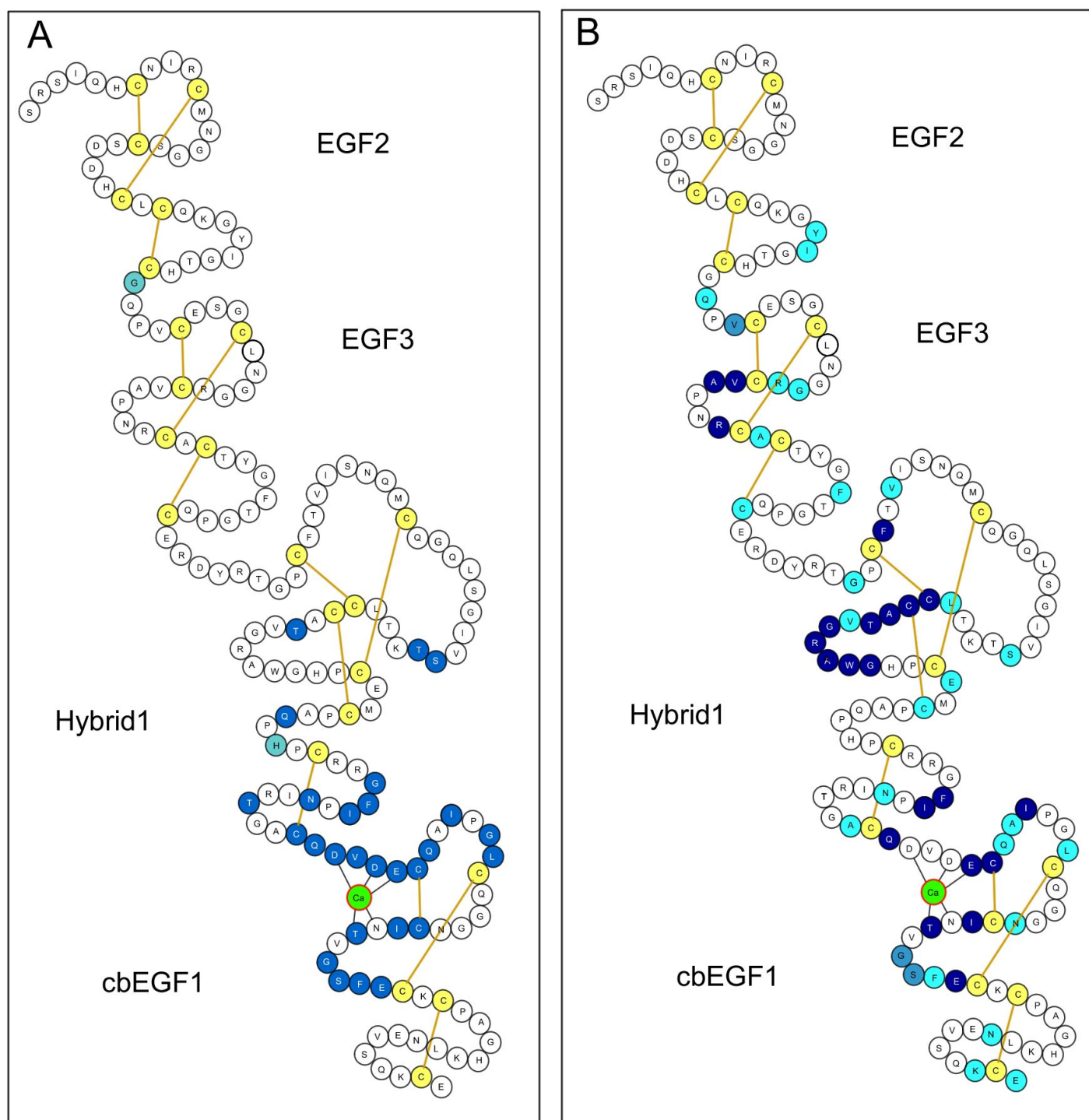


Figure 3.10 - Calcium sensitive HSQC peaks in FBN^{e2cb1}

A) 500 MHz HSQC spectra collected at pH 5.4 in the presence of 5 mM EDTA and no calcium, or 5 mM EDTA and 15 mM calcium. Assignments of calcium-sensitive peaks are highlighted (H229 is marked in grey as this residue is particularly pH sensitive) **B)** Assignments highlighted in A mapped onto a putative structure of this fragment (see section 3.12 (structure 40-3-3)), with calcium-sensitive residues coloured blue and their side chains highlighted as stick representations. Calcium is shown bound to the protein as a green sphere. The boxed image shows detail of the FBN^{e2cb1} calcium-binding site from a different angle.



of hyb1. These calcium-sensitive peaks suggest an interface between the hyb1 and cbEGF1 domains, which appears consistent with that predicted by the preliminary structure (Section 3.12). The peak of glycine 146 in EGF2 also appears to be affected by calcium in these spectra, which is unexpected as it is not predicted to be close to the calcium binding site, however the position of this peak also changes in spectra collected with different levels of sodium chloride, demonstrating that this peak is sensitive to general changes in ionic strength and not specifically calcium.

3.7 Backbone solvent accessibility

To look at solvent accessibility and hydrogen bonding of FBN^{e2cb1} amides, a D₂O exchange experiment was carried out, in which a freeze dried sample of ¹⁵N-labelled FBN^{e2cb1} was dissolved in D₂O, and successive HSQCs collected to examine the loss of peak intensity, as backbone amides exchange with the deuterium solvent (Section 8.32). Figure 3.12 A shows a number of peaks persisting for many hours and even days after the protein was dissolved in D₂O, while others disappeared more quickly. When the residues with slowly exchanging peaks are highlighted on the proposed structure of FBN^{e2cb1} (Figure 3.12 B) and in the FBN^{e2cb1} construct schematic (Figure 3.11 B), a large number of highly protected amides can be seen in the core of the hyb1 domain, and also in cbEGF1, while the EGF2 domain only has a couple of slowly exchanging amides near the interface with EGF3, and EGF3 has some slow exchanging amides initially, but most of them exchange on a more rapid timescale than those in the hyb1 and cbEGF1.

This suggests the presence of hydrogen bonds throughout the FBN^{e2cb1} structure, which are generally consistent with the preliminary structure shown. The particularly slow exchange of amides in regions of hyb1 and cbEGF1 support these domains

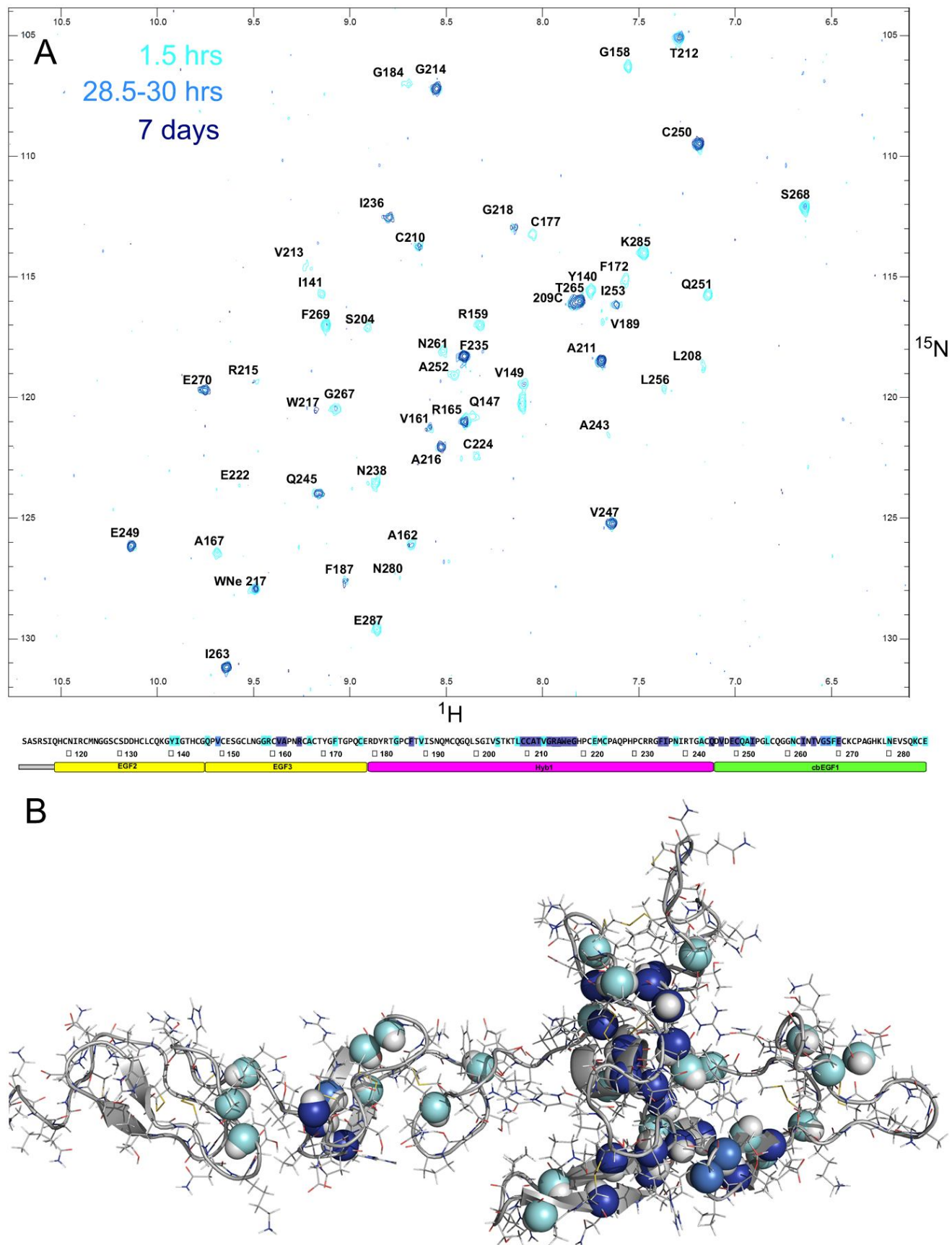


Figure 3.12 - Slowly exchanging amides in FBN^{e2cb1}

A) Successive 600 MHz HSQCs collected at pH 5.4 with 5 mM calcium after freeze dried protein was dissolved in D₂O (Section 8.32). Assignments for peaks that persist are highlighted. **B)** Slowly exchanging peaks mapped onto a preliminary structure of this fragment (see section 3.12 (structure 40-3-3)), amides represented as spheres and coloured according to the length of time their peaks persisted (see A for legend).

forming a solvent protected core, potentially helped by hyb1 - cbEGF1 domain packing.

3.8 Structure determination of FBN^{e2cb1} – General strategy

The pattern of slow exchanging amides, along with the calcium-sensitive peaks, and the similarity in chemical shift of FBN^{e2cb1} EGF2-3 peaks to those of FBN^{NE3} EGF2-3, all support the domains in FBN^{e2cb1} folding in a manner consistent with homologous domains. However the accuracy of homology modelling FBN^{e2cb1} is limited, especially with regard to interdomain interfaces. The nature of the EGF3-hyb1 interface is particularly difficult to predict, as the closest homologous structure is the cbEGF9-hyb2 domain pair [4], but this displays a very limited packing interaction and is missing many of the conserved putative packing residues found in EGF3 – hyb1 (Section 2.4). This makes further structure determination of the FBN^{e2cb1} fragment of particular interest, as it may reveal novel interdomain interaction mechanisms, as well as providing more specific information on the position of key residues such as cysteine/serine 204. A structure of FBN^{e2cb1} may also provide information on relative domain orientations, which may influence the availability of specific binding sites for other proteins, such as LTBP1.

The challenging spectra and relatively large size (~19 kDa) of FBN^{e2cb1} did not make it an ideal candidate for NMR structure determination. However significant efforts to crystallise this protein for structure determination by X-ray methods, were unsuccessful. This left structure determination by NMR as the most viable option for this fragment.

The existing structure of hyb2-cbEGF10 [4] is homologous to hyb1-cbEGF1, and the structure of the EGF2-EGF3 pair has also been determined (D. Yadin, personal communication). For initial structure calculations, and to aid with NOE assignment, a

homology model based on these two structures was used. XPLOR based refinement calculations were setup using this homology model as the starting structure (Section 8.31). These calculations had the advantage that they could be run significantly faster than full simulated annealing of a random structure, and so allowed quicker iterative improvement of the restraints used in the calculations. Once a good set of restraints were established, full simulated-annealing calculations were carried out using XPLOR (Section 8.31), with further refinement of the resulting low energy structures using residual dipolar coupling (RDC) data. Figure 3.13 summarises the overall structure determination process employed.

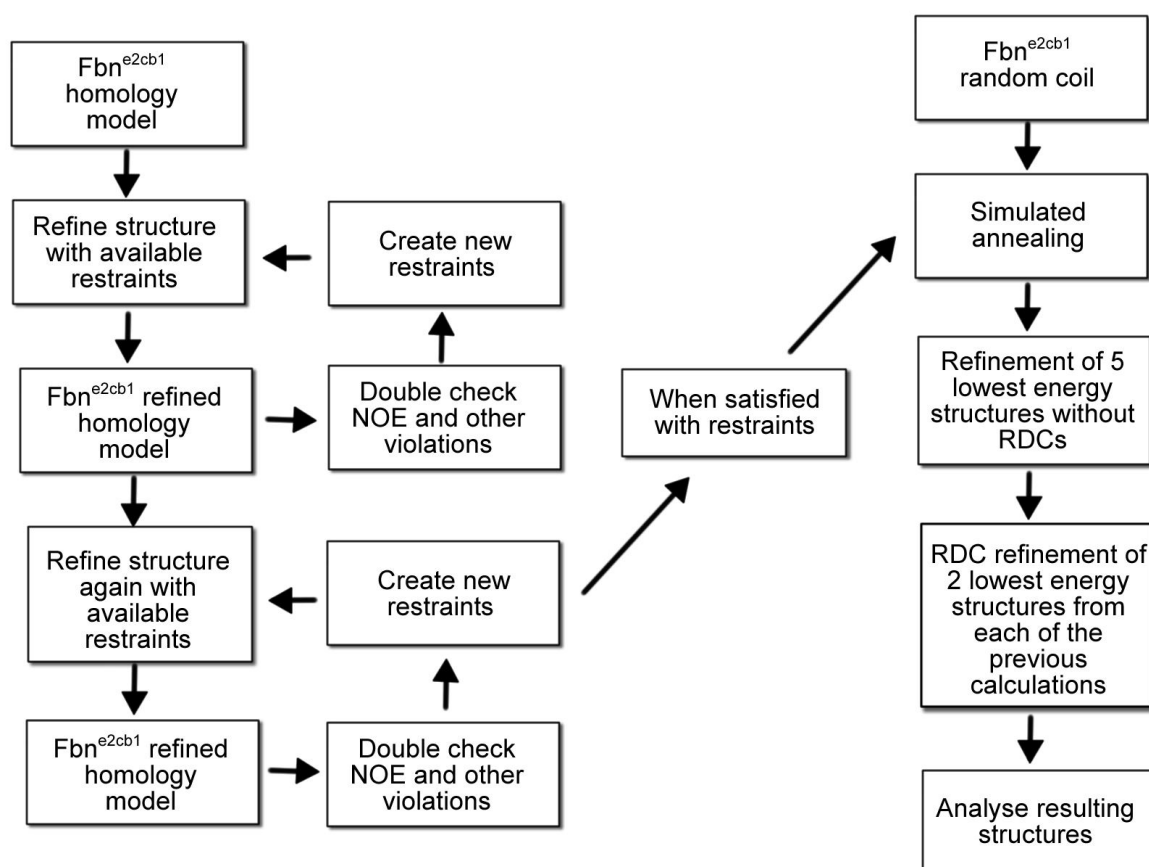


Figure 3.13 – Flow diagram of structure determination process employed for FBN^{e2cb1}

3.9 Chemical shift information

Some structural information can be obtained from the chemical shifts of the backbone assignments, although attempted structure determination using chemical shifts alone using the CS23D2.0 server [9] was unsuccessful. However chemical shifts can also be used as a predictor of secondary structure, with the “motif identification from chemical shifts” (MICS) server [10], which gave relatively good agreement with key secondary structure features in the homology model (Figure 3.14). $\text{C}\alpha$, $\text{C}\beta$, CO, $\text{H}\alpha$, and $\text{H}\beta$ Chemical shifts were also used to predict ϕ and ψ torsion angles using TALOS+ [11], shown in Figure 3.15. The bond angle predictions marked as “good” by the TALOS+ calculations, were used later in structure determination.

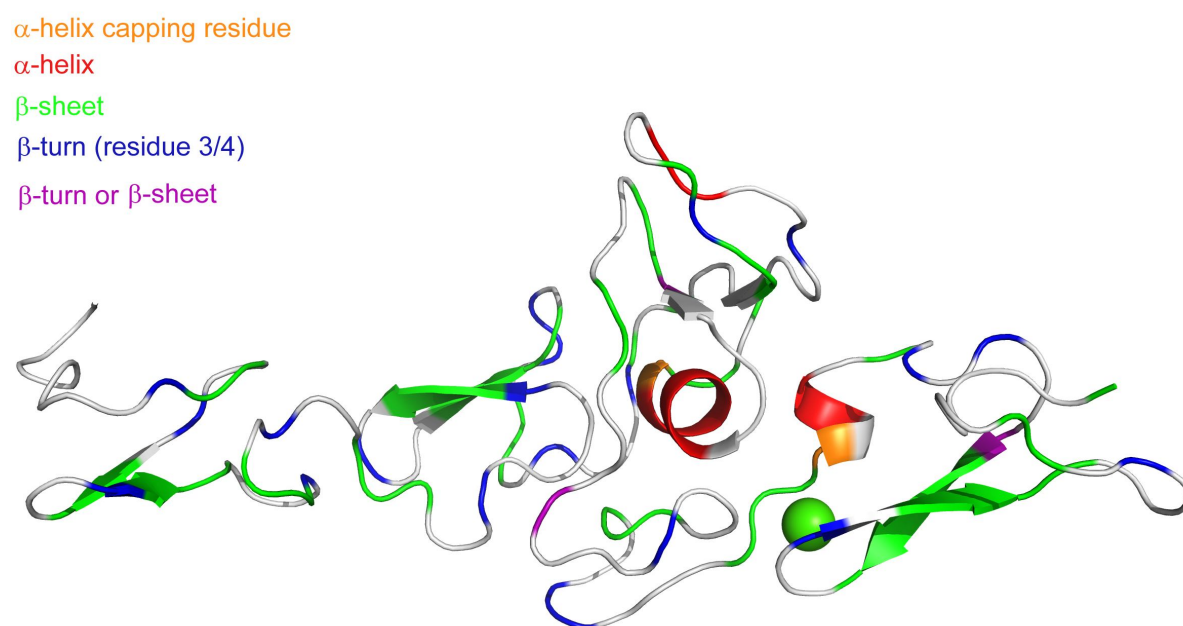


Figure 3.14 – Secondary structure predictions based on chemical shift information, mapped onto FBN^{e2cb1} homology model

Secondary structure predictions obtained using the “motif identification from chemical shifts” (MICS) server [10], key features such as regions of α -helix and large β -strands in the homology model appear to agree quite well with MICS predictions. Calcium included as a green sphere.

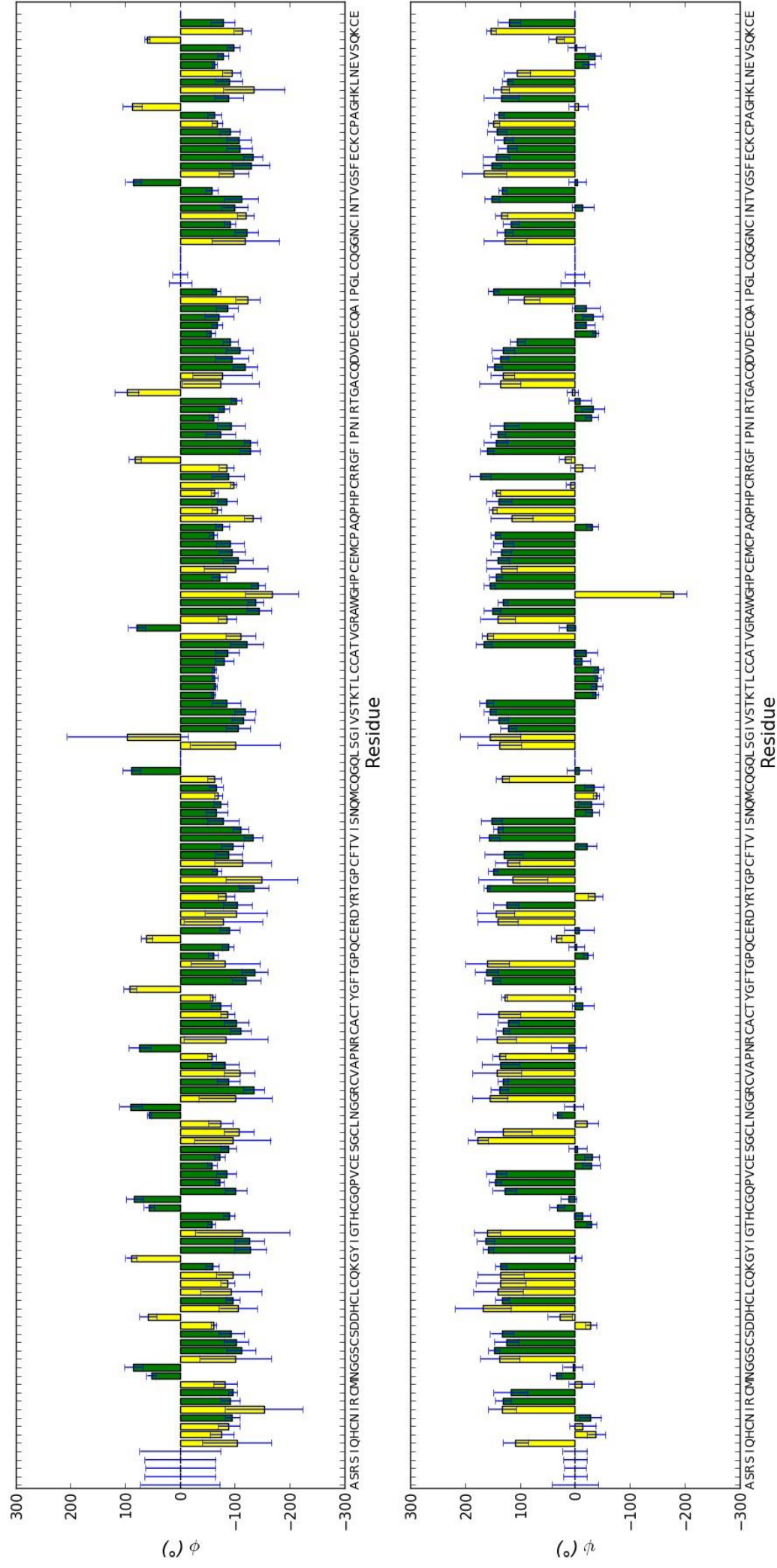


Figure 3.15 - TALOS+ predictions of ϕ and ψ torsion angles

Chemical shifts were used to predict ϕ and ψ torsion angles using the TALOS+ server [11]. Angle predictions considered "good" are coloured green and more ambiguous predictions are coloured yellow (representation generated with help from D. Yadin).

3.10 NOE assignment of FBN^{e2cb1}

For full scale structure determination of FBN^{e2cb1}, NOE assignment was carried out on a number of spectra, so that distance restraints could be generated for structure calculations. All spectra were assigned manually, as weak signal from some peaks, and strong noise surrounding other peaks, made automated peak picking and assignment unrealistic. A total of five different spectra were used to produce restraints, two ¹³C edited NOESY spectra were collected, one in D₂O on a Bruker 500 MHz instrument equipped with a cryoprobe and one in H₂O on an Omega 950 MHz instrument². The ¹³C NOESY collected in H₂O allowed assignment of NH - CH NOEs, while the one collected in D₂O allowed assignment of NOEs to H^α atoms that would otherwise be masked by the water signal. A mixing time of 75 ms was used for both ¹³C NOESY spectra to minimise spin diffusion. Two ¹⁵N NOESY spectra were also assigned, one with a mixing time of 75 ms, and the other with a mixing time of 150 ms (initially collected for backbone assignment), to look at shorter range and longer range NH – CH and NH - NH NOEs. A 2D NOESY collected in D₂O was also used, primarily for assignment of aromatic residue NOEs, as the peak resolution and signal to noise in this experiment was superior to that in the 3D NOESY spectra.

As many atoms in FBN^{e2cb1} have similar chemical shifts, and not all side chain assignments were available, care was needed to validate each NOE assignment, and the homology model was used to assist the process. In total 2991 inter-residue NOE connections were determined across all five spectra, the details of which are summarised in Table 3.1.

² Different spectrometers were used for these experiments based on availability.

Table 3.1 – Numbers of assigned NOE connections in each spectrum

Experiment	Total assigned NOEs	Intra residue NOEs	Inter residue NOEs	Sequential NOEs	Short range (Within 4)	Short range non - sequential	Long range
All spectra	5666	2675	2991	1460	2030	570	961
¹³ C-NOESY H ₂ O	2216	1293	923	403	565	162	358
¹³ C-NOESY D ₂ O	537	277	260	42	110	68	150
¹⁵ N-NOESY 75 ms	1355	512	843	545	703	158	140
¹⁵ N-NOESY 150 ms	1214	463	751	452	603	151	148
2D NOESY D ₂ O	341	130	211	17	48	31	163

Figure 3.16 shows a 2D residue interaction matrix summarising all the inter-residue NOE connections from every assigned NOESY spectra. In this matrix several long-range NOE connections can be seen between the EGF2 and EGF3 domains, and also between the hyb1 and cbEGF1 domains, consistent with their predicted interdomain packing. However connections between EGF3 and hyb1 are more limited, and mostly involve the immediate C and N-termini of each domain. Assignment of more NOEs connecting EGF3 and hyb1 was hindered by the weak intensities of many peaks in the interface region, and also numerous overlapping peaks preventing unambiguous assignment of NOEs from other residues. The presence of NOEs between EGF3 and hyb1, does support interface formation, especially between EGF3 and the N-terminal residues of hyb1. However only NOEs between tyrosine 181 and threonine 205 / 207 connect the N-terminus of the hyb1 domain with the rest of this domain, which limits the structural insights NOEs provide for this interface.

In order to convert these NOE connections into distance restraints for use in structure calculations, the relationship between peak intensity and inter-atomic distance was examined. For ¹³C edited NOESY spectra, the intensity of NOEs between proline side chain protons were used (Figure 3.17), while for the ¹⁵N NOESY spectra NH-NH and

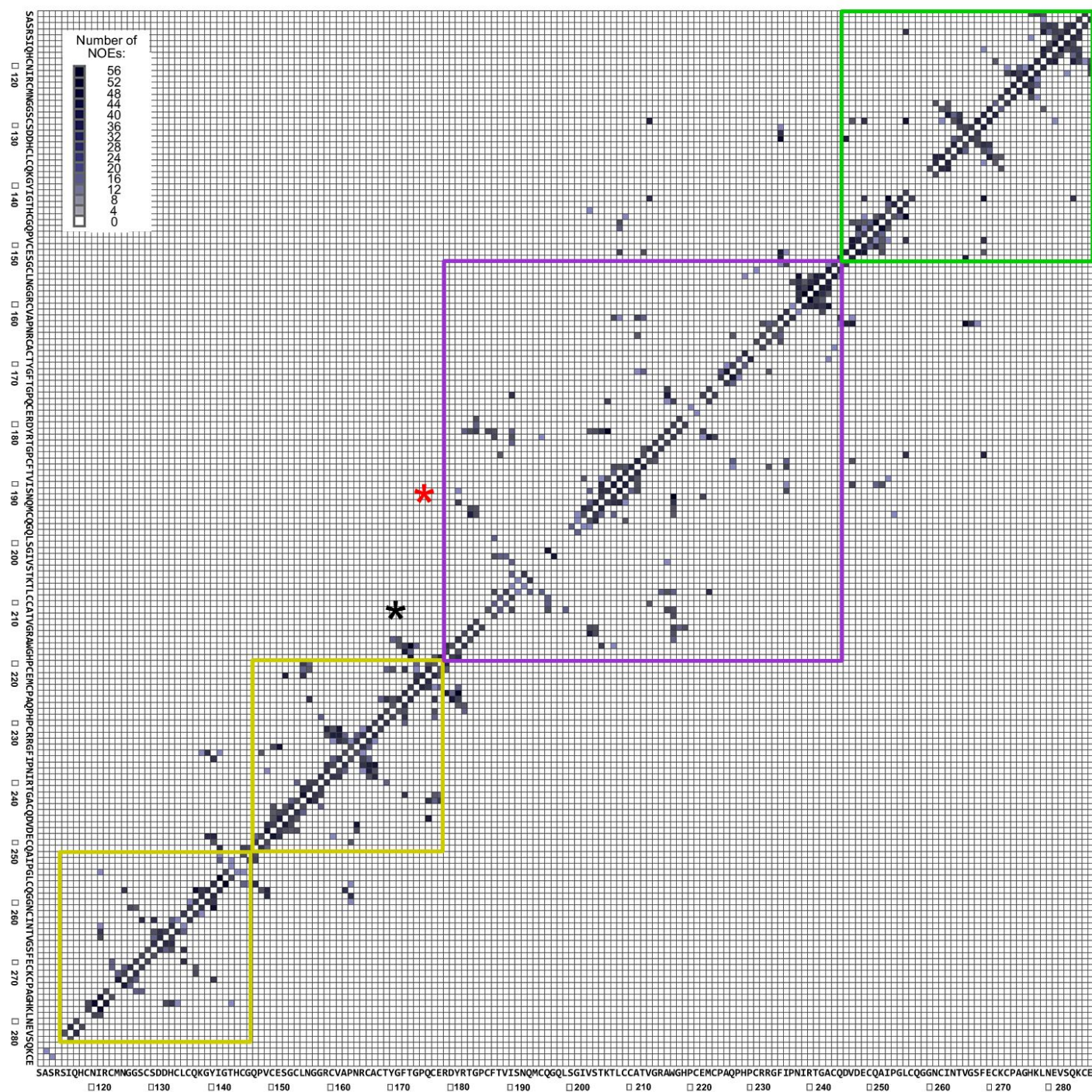


Figure 3.16 - 2D interaction matrix demonstrating the NOE connections between various residues in FBN^{e2cb1}

Coloured squares on the grid represent NOEs between atoms assigned to those specific residues in one or more of the NOESY spectra. Intra-domain NOEs are surrounded by a box coloured according to that domain, and NOEs outside of these boxes represent inter-domain NOEs. Colour intensity of squares represents the number of NOEs between those two residues (legend top right) and does not relate to the intensity of those NOEs. Asterisks highlight specific NOE's discussed in the text.

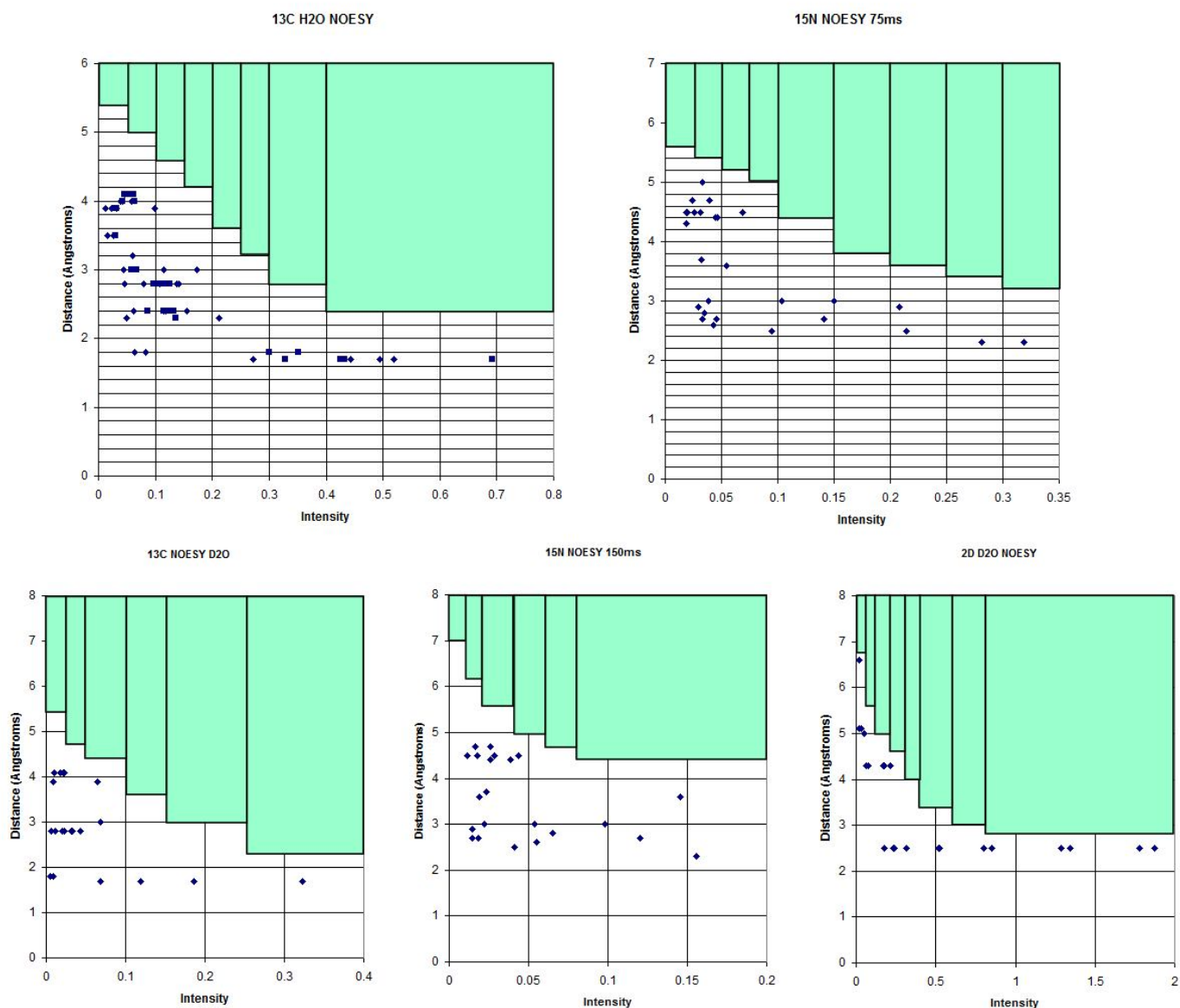


Figure 3.17 - NOE distance restraint calibration

Graphs of peak intensities used to calibrate distance restraints for each spectrum (as described in Section 3.10). The distance bins used to convert NOE peak intensities are shown, with blue boxes representing the limits of each bin.

NH-H^α NOEs in the predicted β-sheet region of cbEGF1 were used. For calibration of the 2D D₂O NOESY distance bins, NOE intensities between side chain aromatic protons were measured.

Distance restraints were then generated from these NOE connections using CCPN analysis software. For assignments of resonances from stereo-specific protons, prochiral swapping was used in initial refinement calculations [12] (Section 8.31). These calculations were then used to assign the chirality of specific methylene and methyl groups, and to update distance restraints where possible, or mark them as ambiguous depending on the energy of the resulting structures (Section 8.31).

3.11 Residual Dipolar Couplings

While NOE restraints were able to provide a significant amount of structural information, they are limited to short range information on the relative position of residues. NOE restraints can only provide information on large-scale structure, for example on the relative orientation of domains, if sufficient NOE's can be assigned from key regions, such as at domain interfaces. As previously discussed only limited NOE information was available to connect the very N-terminus of hyb1 with the rest of the domain (Figure 3.16), making any additional data that can inform on overall domain orientations, highly valuable.

Residual dipolar coupling (RDC) data can be collected from a protein suspended in an alignment medium, and can inform on the orientation of specific nuclei pairs relative to the overall alignment of the protein within the medium, (which is itself aligned with the magnetic field). RDC data can then provide long-range restraints that can describe the relative orientation of domains and residues, and can then be used in structure determination.

An RDC data set was collected for the amides of FBN^{e2cb1} using a 2.2% mix of PEG–hexanol in H₂O:D₂O as an alignment medium [13] (Section 8.30). IPAP experiments were collected in this alignment medium, and the peak splitting measured and compared to the splitting of peaks in an “isotropic” IPAP experiment (without alignment medium) (Figure 3.18). The differences in splitting between peaks in the alignment and isotropic media then gives the RDC value for each amide (displayed in Hz for selected peaks in Figure 3.18). RDC values measured for FBN^{e2cb1} amides are summarised in Appendix 1 Table A1.2. It was not possible to assign RDC values to many peaks due to several factors, including poor signal to noise, poor peak shape, and peaks that had their maxima distorted by overlapping peaks³.

The number and size distribution of these RDC values can inform on the overall alignment of protein within the medium, which is defined by D_a and R (rhombicity) values [14]. Histograms of RDC values from the whole FBN^{e2cb1} protein, as well as specific sets of domains, are shown in Figure 3.19. These histograms show that FBN^{e2cb1} has a number of large positive and negative RDCs (the corresponding peaks for some of these large RDCs are also shown in Figure 3.18 ii). This suggests strong alignment with the 2.2% PEG-hexanol medium, which tends to be seen with long elongated proteins (C. Redfield and P. Weissbuh, personal communication). An unusual feature of these histograms however is that the most populated range of RDC values is negative, and the largest magnitude RDC value is also negative. With negative maxima, the largest RDC would be expected to be positive [14], but it is possible that the N-H vectors sampled here did not adequately cover all possible angles, perhaps due to some weak signal intensity for some peaks.

³ Initially only the best 62 RDC values were used in structure refinement, but fitting all measurable RDC's to the preliminary structures was then used to select a further 22 RDC's for inclusion in structure calculations, which are also included in the analysis here.

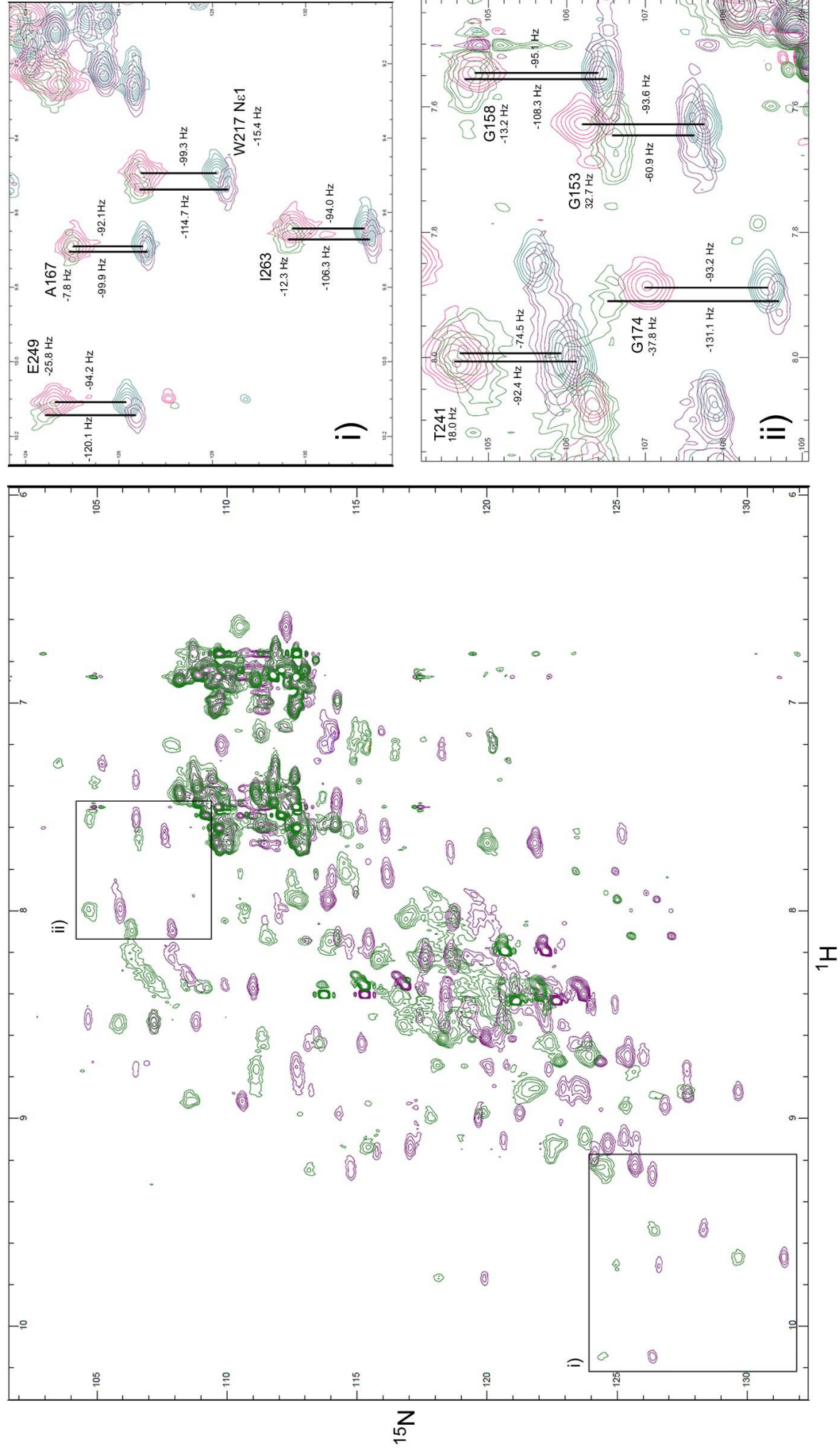


Figure 3.18 - IPAP spectra collected in 2.2% PEG-hexanol for RDC analysis

Spectra collected in 2.2% PEG-hexanol are shown in the main panel with selected regions shown in detail on the right, and overlaid with isotropic experiments (collected without alignment medium). Isotropic peaks are coloured pink and teal, while those in alignment media are coloured purple and green. In i) and ii) lines highlight the measured splitting for both sets of peaks and the resulting RDC is given below the peak assignments. G174 and G153 give particularly large RDCs suggesting strong alignment in the medium despite a relatively low concentration of PEG-hexanol.

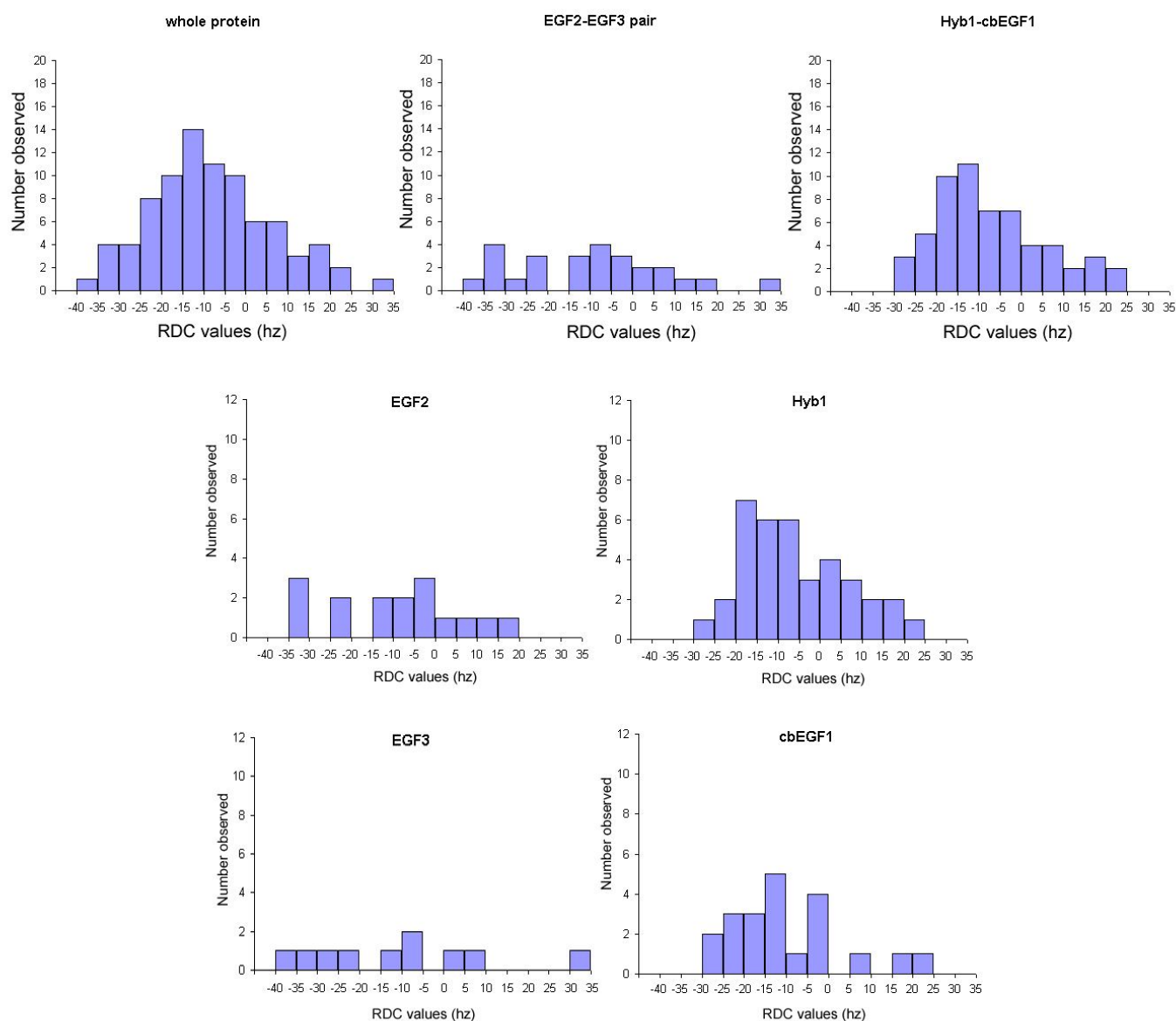


Figure 3.19 - Histograms of measurable FBN^{e2cb1} RDCs

Distance bins of 5 used to generate all histograms using Microsoft Excel. Demonstrating similar distribution of RDCs in whole protein and hyb1-cbEGF1 domain pair, but a limited number or RDCs from EGF2-3 that still exhibit a wide distribution.

3.12 Structure calculations

As summarised in Figure 3.13 an iterative process was used to improve the restraints used in structure calculations, initially using a homology model of FBN^{e2cb1} in XPLOR refinement calculations (Section 8.31), along with the following restraints:

- NOE derived distance restraints from the 5 spectra discussed in Section 3.10
- Torsion angle restraints derived from TALOS+ predictions as described in Section 3.9
- RDC restraints as described in Section 3.11
- Hydrogen bond restraints, predicted from amide D₂O exchange experiments, with hydrogen bonding partners predicted from the homology model (included as ambiguous restraints where necessary)
- Calcium ligand restraints based on calcium binding residues in homologous cbEGF structures.
- NOE restraints for EGF2 and the first part of EGF3 collected from the FBN^{NE3} protein (data kindly provided by D. Yadin) to improve the structures of these domains.

After multiple cycles of structure calculations and NOE assignment, structures were generated which converged well and gave low energies in the XPLOR calculations (~220 - 300 kcal mol⁻¹). The 10 lowest energy structures from a refinement run of 100 structures (with pro-chiral swapping) are summarised in Figure 3.20.

However the fact that these structures were based on refinement of a homology model has the potential to bias the resulting structures towards those that more closely resemble the starting model. To fully sample all possible structures that satisfy the

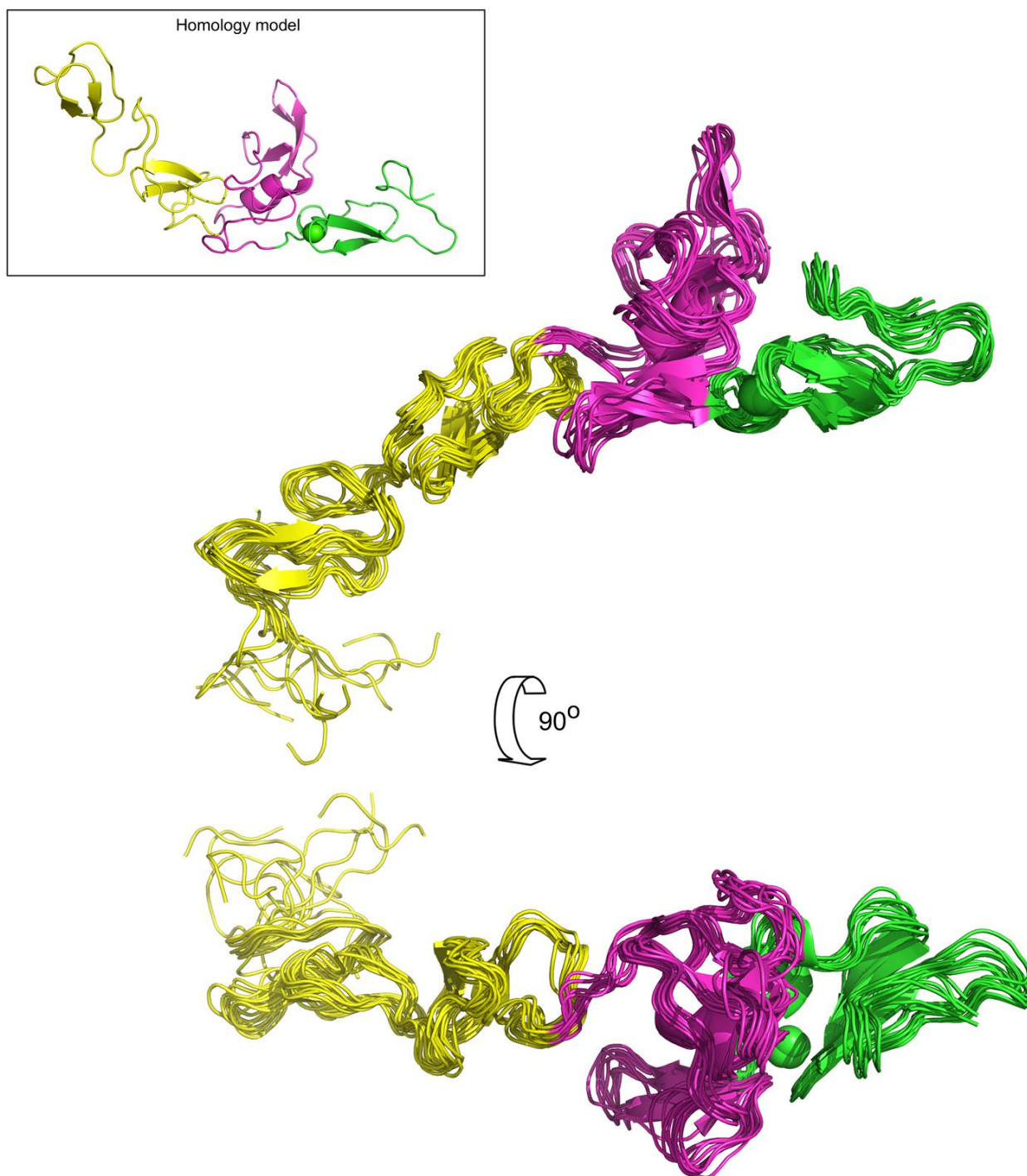


Figure 3.20 - Preliminary structures obtained from RDC refinement of FBN^{e2cb1} homology model
 The homology model starting structure is shown in the box at the top left, with the hyb1-cbEGF1 domain pair of the homology model aligned to the hyb1-cbEGF1 domain pair of the structure in the main figure below. This structure was refined with the XPLOR program (Section 8.31) using various structural restraints, including NOE based distance restraints and RDCs. The 10 lowest energy structures from a run of 100 structures are shown here, aligned using pymol. (The starting structure for this refinement was the lowest energy structure of 20 initially generated from the homology model refinement). EGF2-3 domains are coloured yellow, while the hyb1 domain is coloured magenta and cbEGF coloured green. (Pro-chiral swapping was used to generate these structures, but no Ramachandran potential was included.)

available restraints, simulated annealing calculations were carried out (Section 8.31), which start from a randomly generated arrangement of the protein backbone. After running these simulated annealing calculations to generate 100 structures, the 5 lowest energy structures were then further refined, using NOE and other restraints (but not RDC restraints). Then from the 20 refined structures generated from each starting structure, the two lowest energy structures were then taken forward for further RDC refinement. The low energy structures generated at each calculation stage are summarised in Figure 3.21.

As can be seen in Figure 3.21, and in an overlaid summary of the final RDC refined structures (Figure 3.22), the structures refined from simulated annealing calculations do not converge as well as those refined directly from the homology model of FBN^{e2cb1}. While the individual domains adopt similar folds in all structures (Figure 3.22), the major difference arises from the relative orientation of the EGF2-3 domain pair to the hyb1-cbEGF1 domain pair, appearing to hinge around the EGF3 - hyb1 interface. Structures appear to adopt either a “bent” or “straight” conformation, both of which give similar fitting of the restraints, and the energies and violations of example bent and straight structures are summarised in Table 3.2. The RDC data should theoretically help with questions over relative domain orientation, but in this case it appears that both bent and straight orientations satisfy the RDC values, as both give similar fitting of the RDCs and low “sani” energies in XPLOR⁴.

The NMR results presented here do not therefore represent a definitive structure of the FBN^{e2cb1} fragment, but a preliminary insight into its potential domain arrangement. Further experimental evidence may be required to distinguish between the bent and straight conformations suggested by the data used here. Such evidence might take

⁴ A similar mix of bent and straight structures were also obtained when the RDC alignment parameters, D_a and R , were varied.

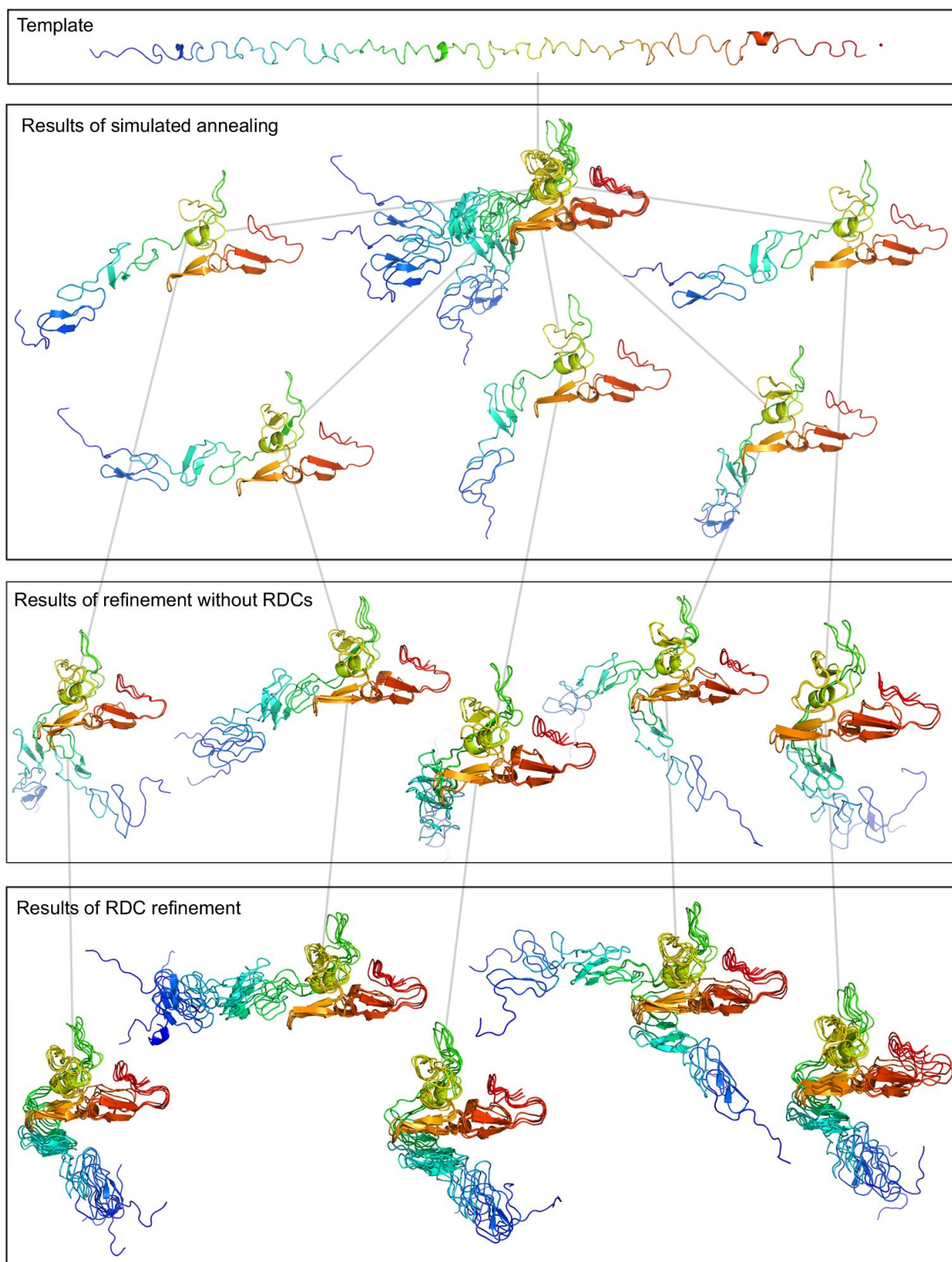


Figure 3.21 - Structures obtained at various stages of FBN^{e2cb1} structure calculation starting from a random template

Summary of de novo structure determination for FBN^{e2cb1}. Chain is rainbow coloured starting from dark blue at the N-terminus through to red at the C-terminus. Initial random template is shown at the top of the figure which is used as an input into the simulated annealing calculations (Section 8.31). The 5 lowest energy structures of the 100 calculated are shown in the next panel, overlaid, and also separately. These were then used as starting structures for further refinement without RDCs. The lowest energy results of this first refinement are shown in the next panel down (best 2 of 20 structures). Finally each of these structures was then separately refined using RDC data. The lowest energy structures from each RDC refinement are shown in the last panel (best 2 of each 20 structure refinement). The initial refinement structures can be seen to sample a broad range of domain orientations.

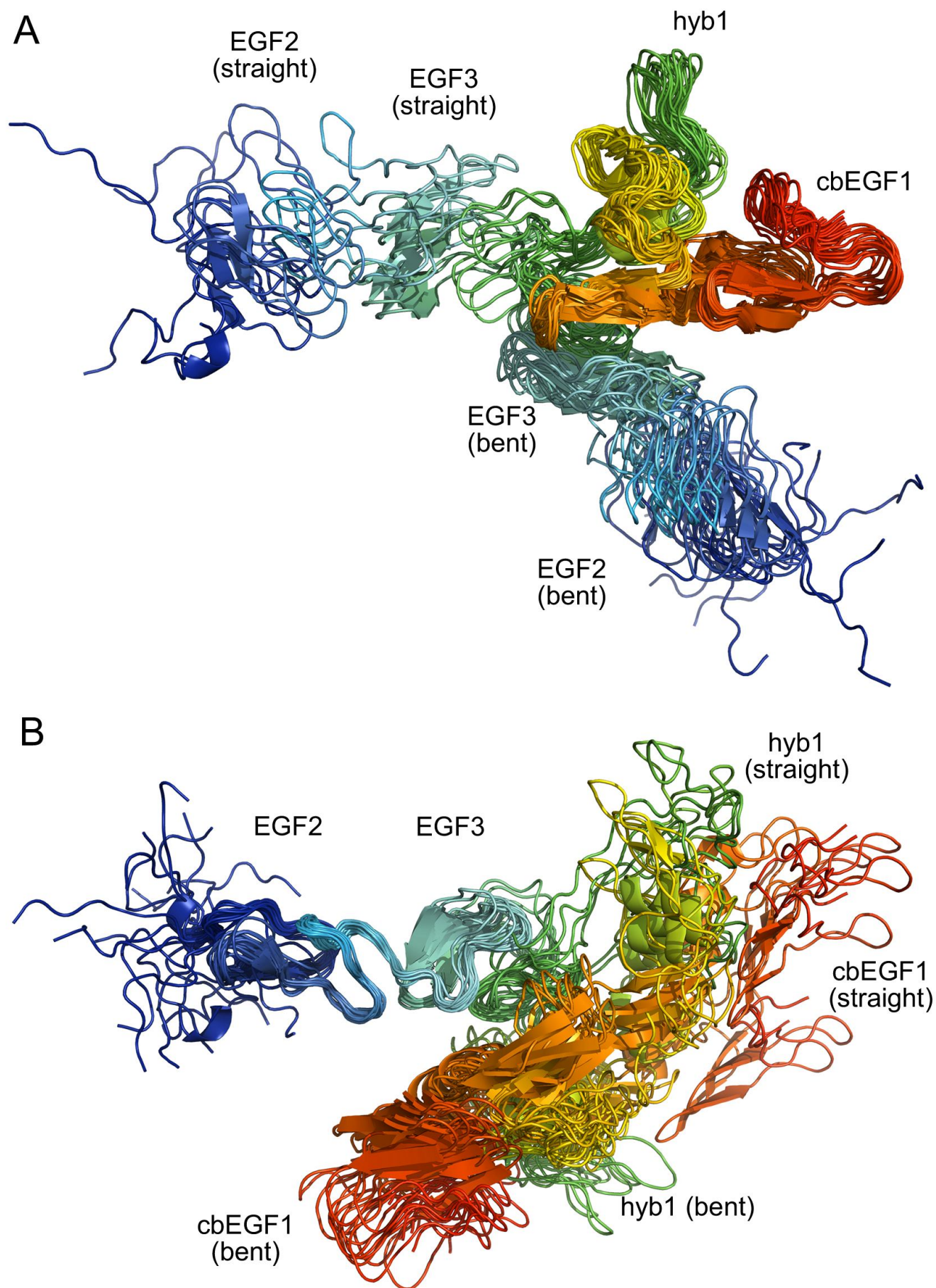


Figure 3.22 - Overlay of final, RDC refined, $\text{FBN}^{\text{e2cb1}}$ structures

A) 20 lowest energy final structures (from Figure 3.21), aligned by the hyb1-cbEGF1 pair, demonstrating two potential domain orientations (6 straight structures, 14 bent). The chain is rainbow coloured with the N-terminus in dark blue and the C-terminus in red. **B)** The same structures in A but aligned by the EGF2-EGF3 pair.

the form of an additional set of RDCs collected in an alternative alignment medium, or other biophysical data collected on the overall shape of the fragment, for example using SAXS experiments. However these experiments fell outside of the timescale of this thesis.

Table 3.2 – Energies and violations of example bent and straight structures
(Backbone RMSD values given for 20 lowest energy structures aligned by specific domains)

	Straight structure (40-3-3)	Bent structure (41-3-14)
Total energy	200.9	194.7
Bond length violation energies	8.85	9.29
Bond Angle violation energies	80.7	80.9
Improper peptide bond energies?	20.4	18.7
Van Der Walls violation energies	37.5	32.4
NOE/ distance restraint energies	43.3	40.5
Dihedral angle energies	3.01	3.80
RDC violation energies	7.24	9.05
Ramachandran potential (not included in total energy)	-1038.5	-1046.9
Number of distance violations greater than 0.2 Å	4 of 5039	2 of 5039
Number of dihedral angle violations > 3	1 of 161	3 of 161
RMSD EGF2-EGF3 region	1.47 +/- 0.97	
RMSD hyb1-cbEGF1 region	2.44 +/- 0.64	

3.13 Structural features of FBN^{e2cb1}

Despite the limitations of the structures discussed above, they still reveal some interesting structural insights into the FBN^{e2cb1} protein. While it cannot be clearly established whether the bent or straight conformation of FBN^{e2cb1} is the most accurate representation of its structure, both orientations are distinct from domain orientations seen in cbEGF9-hyb2-cbEGF10 and cbEGF22-TB4-cbEGF23 fragments (Figure 3.23). Also some novel features of the EGF3 - hyb1 interface can be clearly

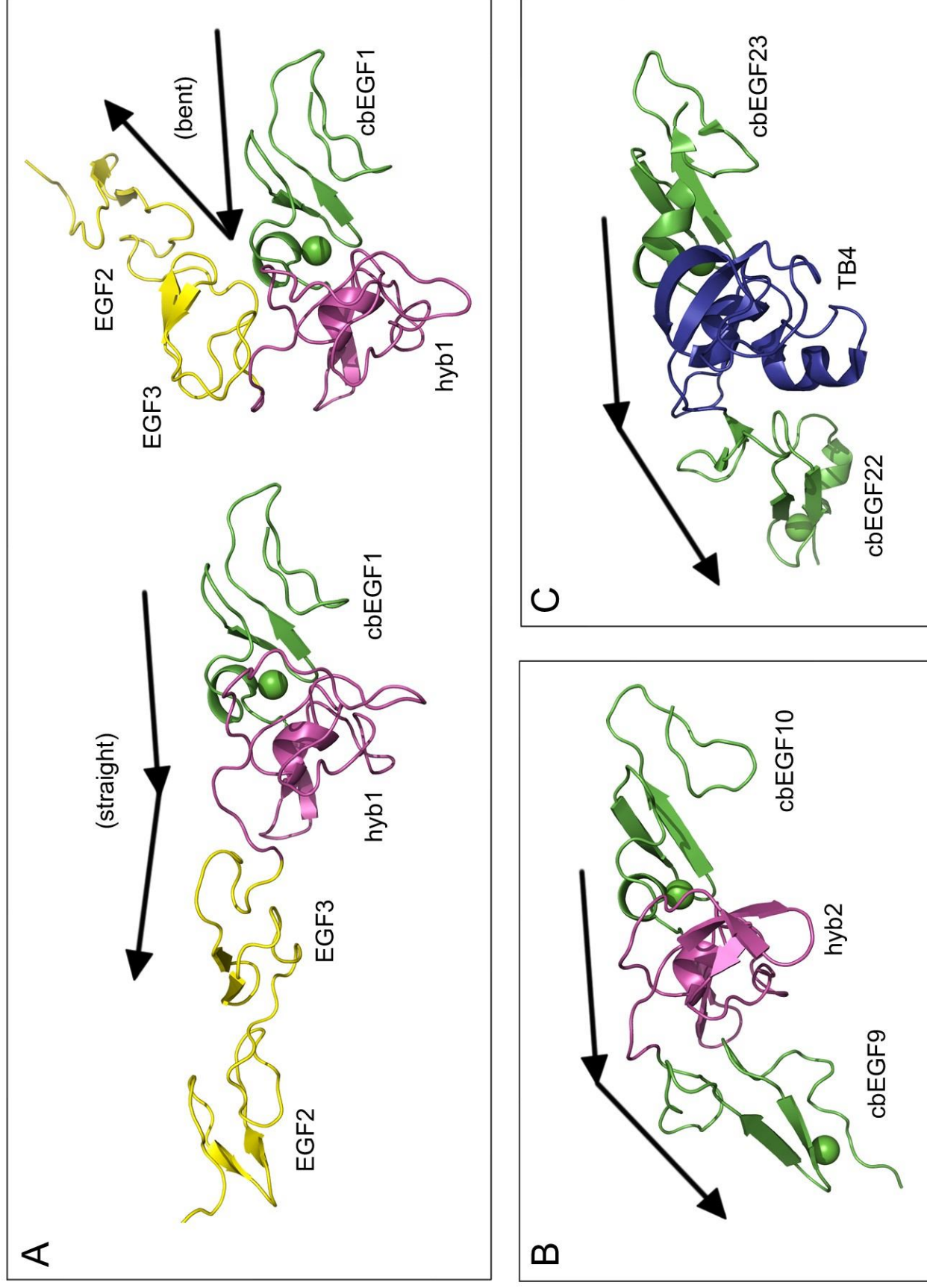


Figure 3.23 - Comparison of domain orientations in $\text{FBN}^{\text{e2cb1}}$ structures with those in cbEGF9-hyb2-cbEGF10 and cbEGF22-TB4-cbEGF23

A) Example structures of both straight and bent $\text{FBN}^{\text{e2cb1}}$ potential conformations. **B)** Crystal structure of cbEGF9-hyb2-cbEGF10 [4] (PDB 2W86). **C)** Crystal structure of cbEGF22-TB4-cbEGF23 [8] (PDB 1UZJ). All structures in this figure are aligned by their last two domains using the pymol structural alignment tool. Arrows summarise the orientation of the upstream EGF/cbEGF domain relative to the downstream hyb/TB - cbEGF domain pair.

distinguished in all structures (Figure 3.24), even though these features cannot be used to precisely define the correct domain orientations. Specifically a number of aromatic packing interactions can be seen to occur between the C-terminus of EGF3 and the N-terminus of hyb1, which are supported by NOEs. Tyrosine 181 at the start of hyb1 seems to be particularly central to the interdomain interaction, packing with the methyl group of threonine 173, the side chain of arginine 179, and also displaying an NOE connection with threonine 205, in the core of the hybrid domain (Figure 3.24 B). Phenylalanine 172 in the EGF3 “GF” packing motif also appears to pack with the side chain of aspartate 180 at the start of the hybrid domain, and also the side-chain of asparagine 156 in EGF3. This packing with asparagine 156 is consistent with the up-field shift seen for the H^β resonances of this residue when compared with FBN^{NE3} spectra (Figure 3.8). The fact that phenylalanine 172 is also present in FBN^{NE3} , but does not appear to pack with asparagine 156 in this construct, suggests that the additional interactions with the start of the hyb1 domain are required to stabilise the packing of phenylalanine 172 and asparagine 156.

While the cbEGF9-hyb2-cbEGF10 crystal structure has very few side chain interactions that may stabilise an interface between cbEGF9 and hyb1 (Section 1.2.3), the crystal structure of cbEGF22-TB4-cbEGF23 has a number of residues in cbEGF22 that interact with residues in TB4 [8] (Figure 3.24 C). However the cbEGF22-TB4 interaction appears different from the interaction of EGF3 and hyb1 in a number of ways, for example instead of the “GF” packing motif found in EGF3, cbEGF22 has a “DF” motif, from which aspartate 1516 appears to form a salt bridge with arginine 1530 in TB4 (equivalent to arginine 182 in hybrid1). In cbEGF22 phenylalanine 1517 in the DF motif does appear to pack with the side chain of aspartate 1528 at the start of the TB4 domain (equivalent to D180 in hyb1). However unlike phenylalanine 172 in EGF3, phenylalanine 1517 does not appear to pack with the equivalent residue to asparagine

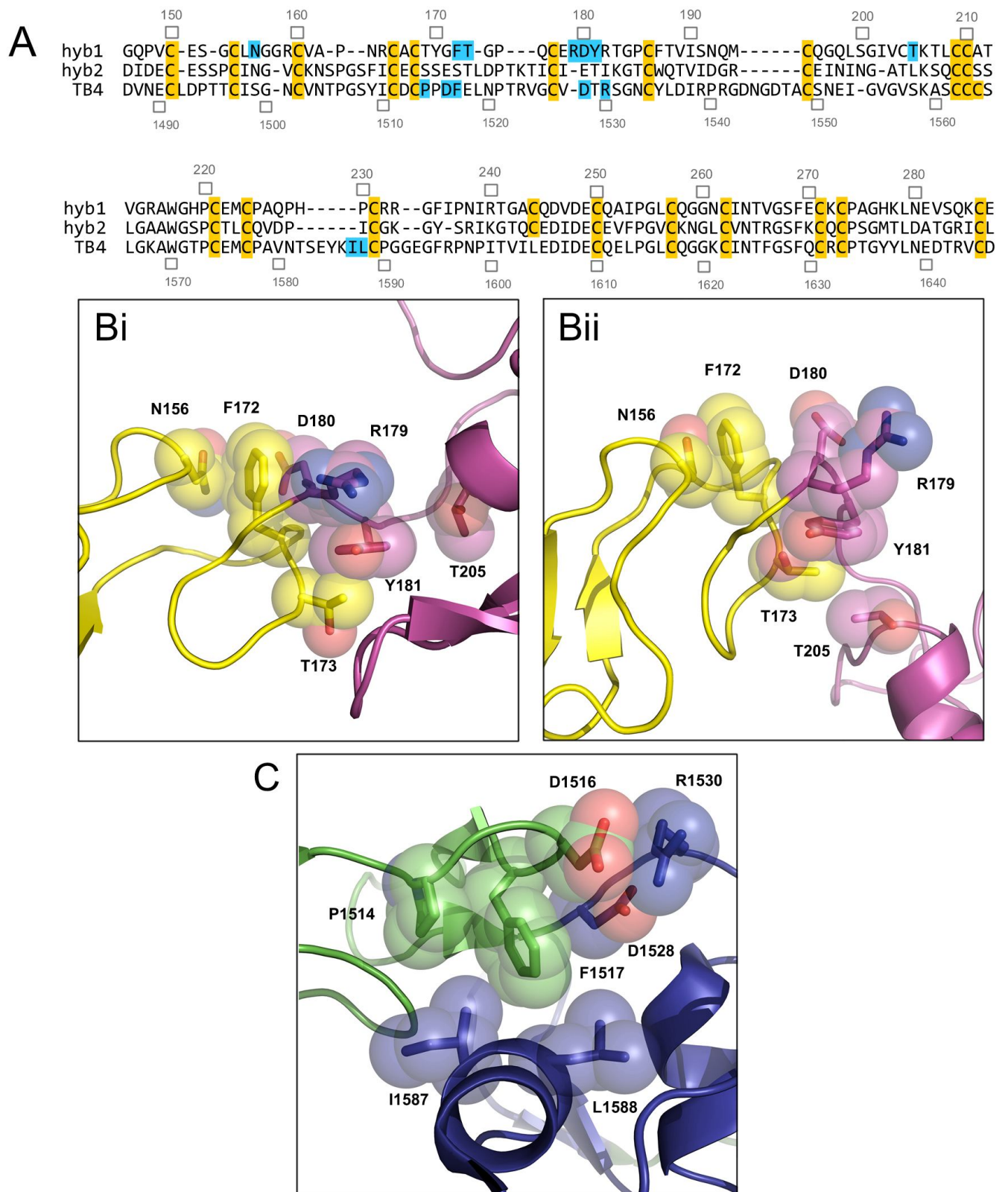


Figure 3.24 - Detailed look at features of EGF3 - hyb1 interface

A) Alignment of the sequences of EGF3-hyb1-cbEGF1, cbEGF9-hyb2-cbEGF10, and cbEGF22-TB4-cbEGF23. Cysteine residues are highlighted in yellow and packing residues shown in B and C are highlighted in blue. Numbers above the alignment correspond to residues in EGF3-hyb1-cbEGF1, while those below correspond to cbEGF22-TB4-cbEGF23. **Bi)** Straight and **Bii)** bent interfaces between EGF3 and hyb1 in selected FBN^{e2cb1} structures. The side chains of key packing residues are shown as sticks and semi transparent spheres. Many of the specific packing interactions shown are supported by NOESY data (Figure 3.16) such as NOEs between F172 and N156/D180, and between Y181 and R179/T173/T205. Residues in EGF3 are coloured yellow, while those of hyb1 are coloured magenta. **C)** cbEGF22-TB4 packing interaction seen in the crystal structure of these domains. Selected packing residues are represented as sticks, with those from cbEGF22 coloured green, and those from are TB4 coloured blue.

156 on the first turn of the EGF, but instead with proline 1514 on the C-terminal loop of cbEGF22. Phenylalanine 1517 also packs with residues on the second helix of TB4, and the equivalent helix is absent from hybrid domains.

The interdomain interactions between EGF2 and EGF3, and those between hyb1 and cbEGF1 are more consistently defined in the structures generated. The interaction between EGF2 and EGF3 had previously been studied in the FBN^{NE3} structure (D. Yadin, personal communication) and is consistent with the interface observed in FBN^{e2cb1} structures (Figure 3.25 A). The interface between hyb1 and cbEGF1 is also similar in all structures (Figure 3.25 B), and involves extensive hydrophobic contacts between residues of both domains. These interdomain contacts are quite similar to those seen in the crystal structures of the hyb2-cbEGF10 and TB4-cbEGF23 domains (Figure 3.26). In all three proteins the G(Y/F) motif near the C-terminus of the hybrid or TB domain, packs with the “XG” motif on the main β -turn of the down-stream cbEGF. Other potential hydrophobic contacts can be seen between residues at the end of the core α -helix of the hybrid/TB domain and the cbEGF in all three structures. Also the second β -strand of the hybrid/TB domain forms several contacts with the N-terminal α -helical region of the cbEGF. Here all three domains seem to differ slightly in terms of which residues on the second β -strand of hybrid/TB domain contact the cbEGF α -helical region; in hyb1 isoleucine 202, near the end of the second β -strand, packs with the helix region in cbEGF1, whereas in hyb2 and TB4 residues closer to the middle of this β -strand region pack with the downstream cbEGF.

As well as the interdomain interfaces, another feature of the FBN^{e2cb1} structures that is of particular interest is the positioning of the serine 204 residue, which in WT protein is a cysteine residue. This residue adds a 9th cysteine to the hyb1 domain, which is highly conserved (Figure 2.8), but is not present in other hybrid or TB domains. This

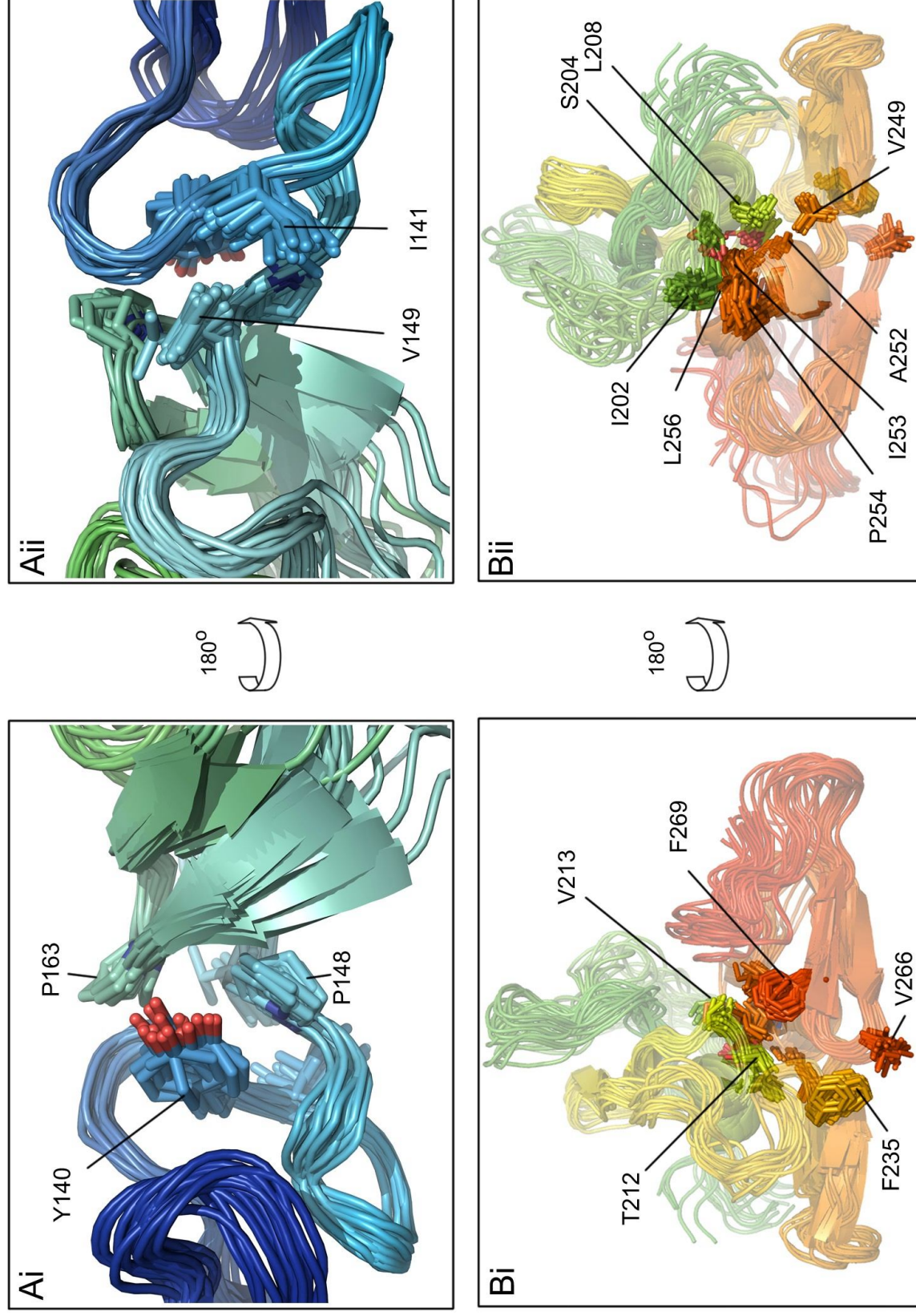


Figure 3.25 – Consistent interface formation between EGF2 and EGF3, and between hyb1 and cbEGF1 in structure calculations

The same 20 overlaid structures from Figure 3.22 are shown, and using the same colour scheme, but focusing on: **A)** The EGF2-EGF3 interface: Packing between tyrosine 140 and prolines 148 and 163 can be observed and is consistent with that observed previously (D. Yadin, personal communication). ii) Highlighting valine 149 and isoleucine 141 which also appear to pack together on the opposite side of the domain pair. **B)** Just the structures of the hyb1 and cbEGF1 domains are shown for simplicity, and rendered semi transparently so that the interface residues stand out. A number of hydrophobic contacts between hyb1 and cbEGF1 can be seen consistently in all structures and the residues involved are highlighted in stick representation. ii) Opposite side of the interface with different residues highlighted.

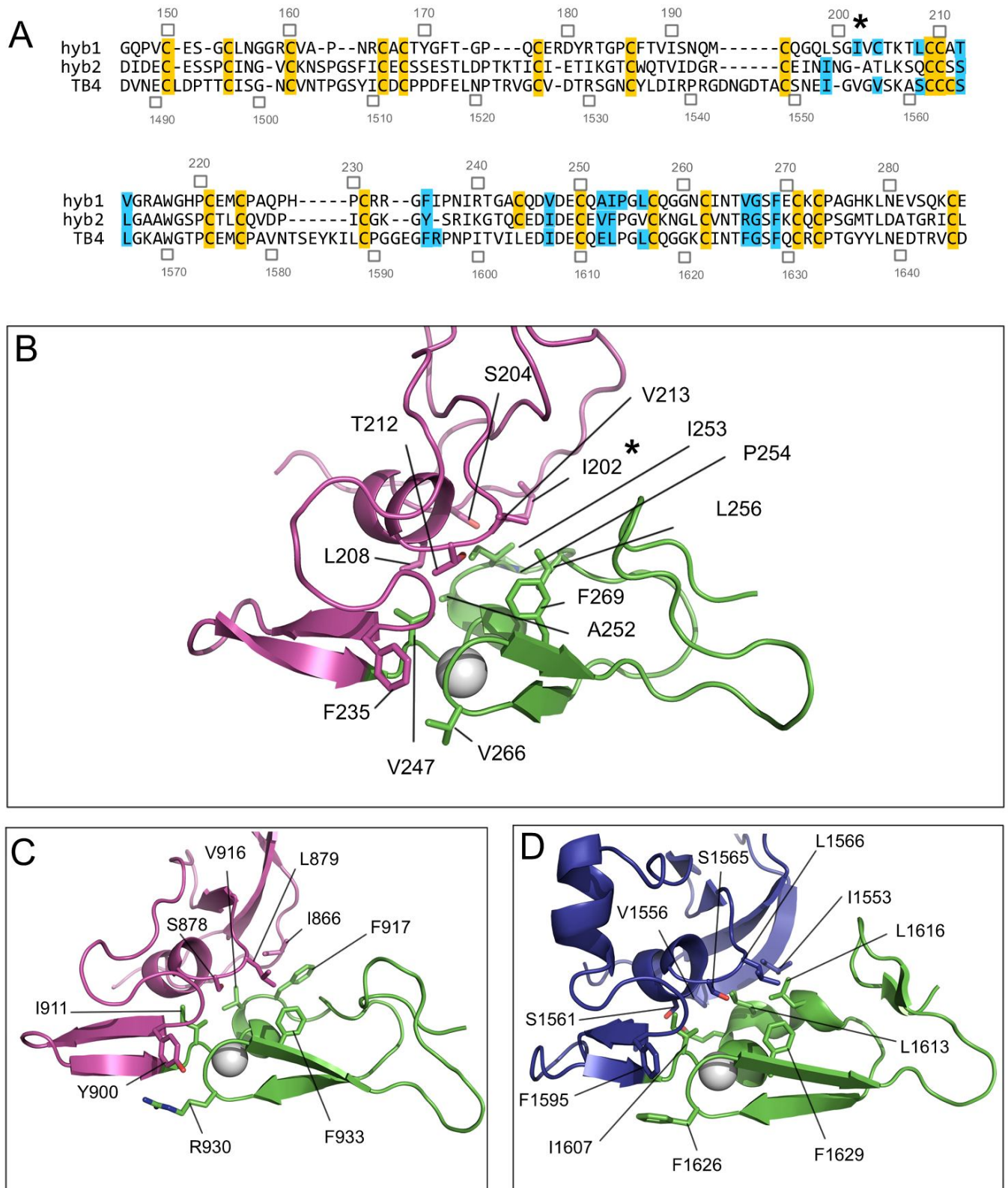


Figure 3.26 - Comparison of hyb1-cb1 interface with hyb2-cb10 and TB4-cb23 interfaces

A) Alignment of the sequences of EGF3-hyb1-cbEGF1, cbEGF9-hyb2-cbEGF10, and cbEGF22-TB4-cbEGF23. Cysteine residues are highlighted in yellow, and packing residues represented as sticks below in B, C and D are highlighted with a blue background. Numbers above the alignment correspond to residues in EGF3-hyb1-cbEGF1, while those below correspond to cbEGF22-TB4-cbEGF23. **B)** Structure of hyb1-cbEGF1 interface, where hyb1 is coloured magenta, cbEGF1 green and calcium ion light grey. Interface residues are highlighted as sticks and labelled accordingly. **C)** Structure of hyb2-cbEGF10 interface, coloured as in B. **D)** Structure of TB4-cbEGF23 interface coloured as in B except the TB domain is shown in dark blue. Asterisk highlights I202, which defines a novel feature of the hyb1-cbEGF1 interaction.

residue is present as a free thiol when expressed by eukaryotic cells [5], and so may be involved in forming inter-molecular disulphides and directing microfibril assembly. Previous modelling of hyb1 based on the structure of hyb2 suggested that this residue may be solvent exposed [4], but in the NMR structures presented here S204 appears to be buried (Figure 3.27), with a surface exposure of only 1.3% (in structure 40-3-3, calculated by Naccess). While serine 204 still appears close to the surface of the protein, the side chains of isoleucine 202, valine 203 and proline 254 protect it from full solvent exposure. Small molecules, such as thiol labelling reagents, may still be able to access the side chain by routes highlighted with arrows in Figure 3.27 E, but for this side chain to interact with a protein and form an inter-molecular disulphide bond, some structural rearrangement of hyb1 would be necessary. It is possible that the replacement of cysteine with serine alone causes some structural alterations to this region, although the side chain hydroxyl group of serine 204 does not appear to form any putative hydrogen bonds in the NMR structures, and there is nothing else to suggest a cysteine side chain would behave differently in this position.

3.14 Conclusions

These structural studies of the fibrillin1 EGF2-EGF3-hyb1-cbEGF1 domains demonstrate that they can be folded successfully *in vitro*, and that the individual domains appear to all adopt folds that are similar to those seen with homologous domains. A heteronuclear NOE experiment looking at fast time scale dynamics demonstrates that the protein is generally quite rigid, only displaying flexible residues at the very N-terminus, and in the C-terminal β -turn of hyb1. No residues appear to have the potential to behave as flexible inter-domain linkers, and NOE assignment demonstrates interfaces forming between all four domains.

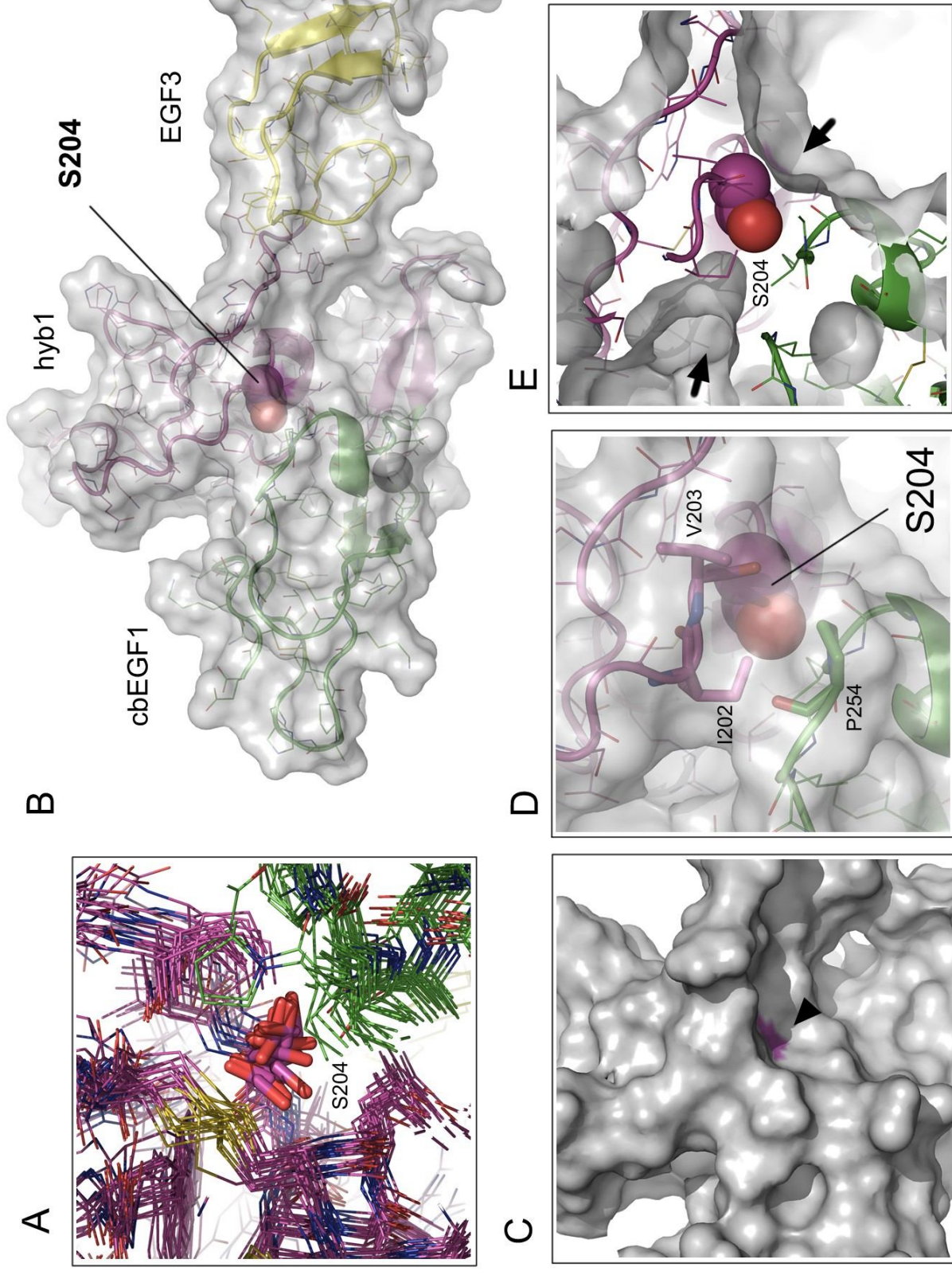


Figure 3.27 - Buried position of serine 204 in hyb1, which corresponds to a free cysteine in WT fibrillin

A) Consistent position of serine 204 (stick representation) in all low energy NMR structures (same structures used as shown in Figure 3.22). **B)** Position of serine 204 in relation to the overall structure of FBN^{e2cb1} (straight structure 40-3-3 shown here). Protein surface shown with 40% transparency, and the serine 204 side chain is highlighted by magenta (carbon) and red (oxygen) spheres. **C)** Solid surface representation showing only a very small patch in magenta that is defined by atoms of the serine 204 residue. **D)** Residues shielding S204 from surface exposure are shown represented as sticks. **E)** Cut away section showing the position of S204 and potential routes of access for small molecules (arrows).

Accurate structure determination of FBN^{e2cb1} was frustrated by a shortage of NOE restraints connecting the very N-terminus of hyb1 with the rest of the domain⁵, and RDCs that appear to support both a bent and straight arrangement of the EGF2-3 pair relative to the hyb1-cbEGF1 pair. The large size of some the RDCs collected in 2.2% PEG-hexanol, (a relatively low concentration for RDC studies), may favour a straighter shape, which could give stronger alignment with the medium, but further experimental information is still required to resolve which domain orientations give an accurate representation of FBN^{e2cb1}. There is also the possibility that slow time scale movements may occur between the N-terminus of hyb1 and the rest of the domain, which would make determination of a single structure impossible. Further RDC experiments could potentially rule out or confirm this, depending on whether the values obtained are consistent with either the bent or straight orientations, or neither.

Despite the remaining questions over the structure of FBN^{e2cb1}, the structural information obtained in this chapter allows identification of a number of novel features in the EGF3 – hyb1 interdomain interface, and also in the hyb1 – cbEGF1 interface, when compared with the structures of cbEGF9-hyb2-cbEGF10 and cbEGF22-TB4-cbEGF23 domain fragments.

This structural information on FBN^{e2cb1} also has implications for the numerous protein-protein interactions involving this region of fibrillin. The positioning of serine 204 in the NMR structures suggests that if the same structure is adopted in WT protein, cysteine 204 would be protected from solvent exposure. This means that if it is involved in microfibril assembly, and inter-molecular disulphide formation, some structural rearrangement of the hyb1 domain would be necessary. This could perhaps be achieved by altered packing of the second β -hairpin of hyb1 with cbEGF1, so that it

⁵ Lack of NOE restraints here was mainly due to a mixture of weak and overlapping peaks from key residues in this region.

more closely resembled the packing of hyb2, which would displace cysteine/serine 204 out into a more solvent exposed position.

As well as potential roles in microfibril assembly, this region of fibrillin has been implicated in numerous interactions with other important extracellular proteins (Figure 1.7), such as the LTBP's and fibulins. The structural insights gained in this chapter provide a foundation for mapping these binding sites, and building a molecular understanding of the role of fibrillins in directing important protein interactions in the ECM.

References

1. Isogai, Z., et al., *Latent transforming growth factor beta-binding protein 1 interacts with fibrillin and is a microfibril-associated protein*. J Biol Chem, 2003. **278**(4): p. 2750-7.
2. Massam-Wu, T., et al., *Assembly of fibrillin microfibrils governs extracellular deposition of latent TGF beta*. J Cell Sci. **123**(Pt 17): p. 3006-18.
3. Ono, R.N., et al., *Latent transforming growth factor beta-binding proteins and fibulins compete for fibrillin-1 and exhibit exquisite specificities in binding sites*. J Biol Chem, 2009. **284**(25): p. 16872-81.
4. Jensen, S.A., et al., *Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins*. Structure, 2009. **17**(5): p. 759-68.
5. Reinhardt, D.P., et al., *Initial steps in assembly of microfibrils. Formation of disulfide-cross-linked multimers containing fibrillin-1*. J Biol Chem, 2000. **275**(3): p. 2205-10.
6. Smallridge, R.S., et al., *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, 2003. **278**(14): p. 12199-206.
7. Downing, A.K., et al., *Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders*. Cell, 1996. **85**(4): p. 597-605.
8. Lee, S.S., et al., *Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization*. Structure, 2004. **12**(4): p. 717-29.
9. Wishart, D.S., et al., *CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data*. Nucleic Acids Res, 2008. **36**(Web Server issue): p. W496-502.
10. Shen, Y. and A. Bax, *Identification of helix capping and b-turn motifs from NMR chemical shifts*. J Biomol NMR, 2012. **52**(3): p. 211-32.
11. Shen, Y., et al., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts*. J Biomol NMR, 2009. **44**(4): p. 213-23.
12. Folmer, R.H., et al., *Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data*. J Biomol NMR, 1997. **9**(3): p. 245-58.
13. Otting, M.R.a.G., *Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments*. J. Am. Chem. Soc, 2000. **122**(32): p. 7793-7797.
14. Clore, G.M., A.M. Gronenborn, and A. Bax, *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, 1998. **133**(1): p. 216-21.
15. Kay, L.E., D.A. Torchia, and A. Bax, *Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease*. Biochemistry, 1989. **28**(23): p. 8972-9.

Chapter 4- Structural studies of the LTBP1 C-terminus

4.1 Introduction

The fibrillin-1 binding site in LTBP1 has been localised to the C-terminal region of the protein by a number of biochemical and biophysical studies [1-3]. The minimum fragment that binds fibrillin1 in these studies consists of the third TB domain, EGF3, and C-terminal cbEGF15 domain (* in Figure 4.1) [1]. As well as these domains, there is also a region of 47 amino acids between the last cysteine of TB3 and first cysteine of EGF3 that contains no recognised domains, and a further 16 amino acids at the C-terminus following cbEGF15 with no clear structural features in its sequence.

While homologous structures exist that can be used to model the LTBP1 TB3 domain or EGF3-cbEGF15 pair in isolation, it is not possible to predict how the C-terminus of LTBP1 is structured as a whole, especially with large regions of sequence present with no known homology. The poor conservation of these regions may suggest they are unstructured (discussed in Section 2.4, Figure 2.7), but this requires verification.

To develop a molecular understanding of how LTBP1 interacts with fibrillin-1 the structure of the C-terminal region of LTBP1 must first be analysed. In this chapter NMR is used to verify the accuracy of homology models of the individual domains, and also to investigate the overall dynamics of the LTBP1 C-terminus.

4.2 Production of recombinant LTBP1 proteins

To investigate this region, a number of overlapping constructs of the LTBP1 C-terminus were cloned into the prokaryotic expression construct pQE30, from human LTBP1L cDNA (kindly provided by Professors Daniel Rifkin and Jorma Keski-oja) (see materials and methods Section 8.7). The three constructs used here are summarised in Figure 4.1. LTBP1^{cb14TB3} is a domain pair consisting of cbEGF14 and TB3, LTBP1^{TB3E3}

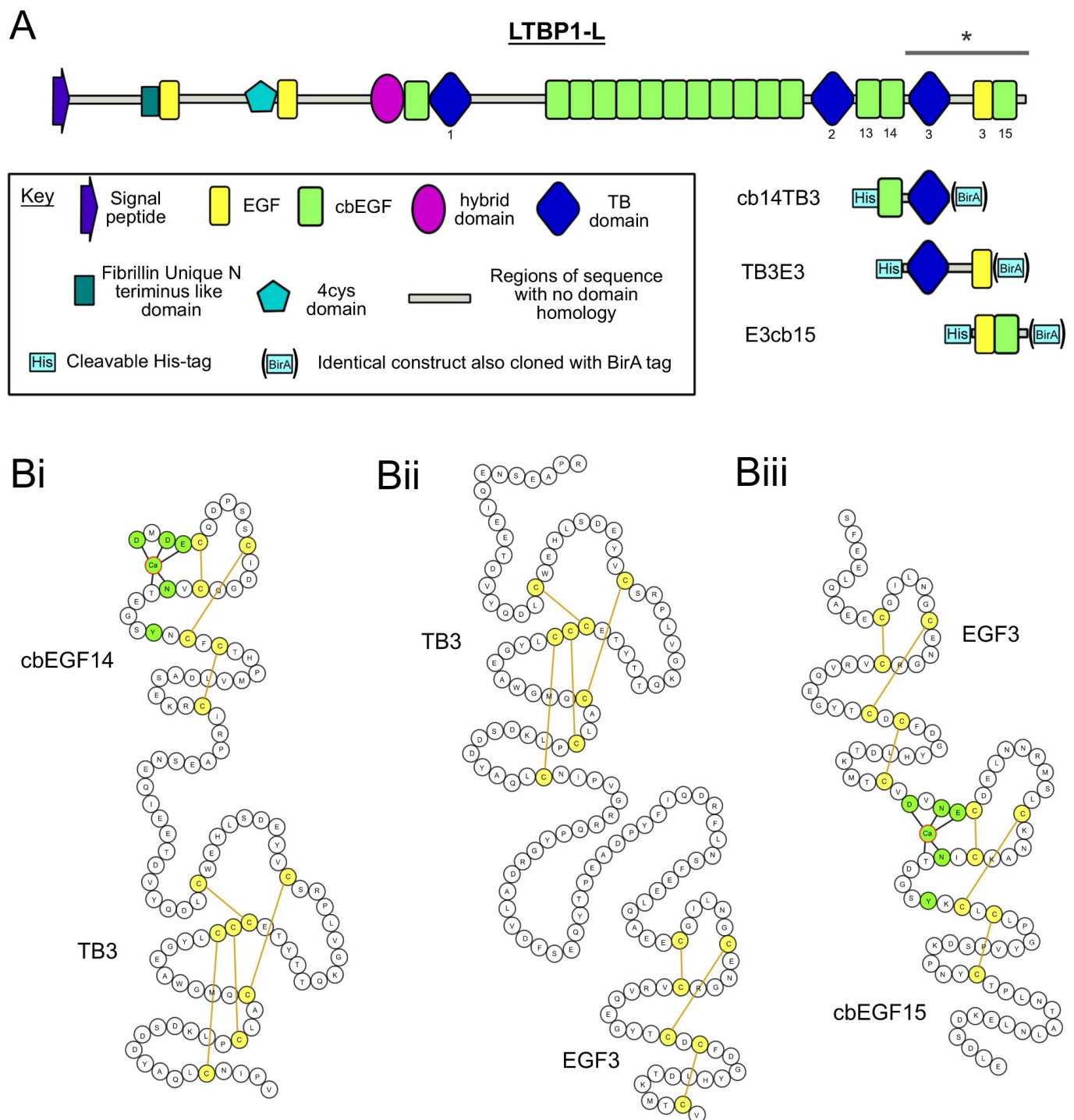


Figure 4.1 - LTBP1 constructs

A) Three overlapping constructs, cb14TB3 (LTBP1L residues 1468-1582), TB3E3 (LTBP1L residues 1509-1662) and E3cb15 (LTBP1L residues 1617-1722), were generated, which span the fibrillin-1 binding region (highlighted with a line and asterisk). Fragments were expressed prokaryotically with and without BirA tags and refolded. The selected domains are also numbered for clarity. **B)** Schematic representations of the primary sequence of (i) cb14TB3 (ii) TB3E3 and (iii) E3cb15 constructs. Calcium binding consensus residues are coloured green and cysteines are coloured yellow.

consists of TB3, the interdomain linker, and EGF3, while LTBP^{E3cb15} consists of the EGF3-cbEGF15 domain pair, and includes a short sequence of conserved amino acids preceding EGF3, as well as the 16 amino acids following cbEGF15 at the C-terminus of LTBP1. These fragments in particular were chosen because their individual sizes are all below 20 kDa, which should simplify NMR assignment. Their small size also allows a more precise dissection of domains involved in the interaction of LTBP1 with fibrillin-1 (see Chapter 5). Overlapping fragments were also chosen so that it would be possible to verify whether or not individual domains could fold independently *in vitro*. Fragments were cloned with and without BirA tags for biotinylation in case they were needed for interaction assays.

As with fibrillin fragments (Chapter 3), proteins were expressed in *E. coli* BL21 cells, in either rich or minimal media, depending on whether isotopic labelling was required (Sections 8.9 and 8.10). All LTBP1 proteins were then extracted by guanidine-HCl lysis, followed by nickel column purification (Section 8.11). Due to their low pI (~5) these LTBP proteins precipitated when guanidine was removed by dialysis against 0.1% TFA at pH 2. However proteins could then be re-dissolved and refolded directly from this precipitate (as described in Section 8.14 and 8.15). Reverse-phase HPLC was used to purify the *in vitro* refolded proteins (Section 8.13), which were then freeze dried. Proteins were then re-dissolved and subjected to factor Xa cleavage to remove their His-tags (Section 8.16), except in the case of LTBP1^{TB3E3}, which showed additional degradation when exposed to factor Xa. All proteins were further purified by FPLC using a MonoQ anion exchange column (Section 8.17), and a final round of HPLC followed by freeze drying. Figure 4.2 demonstrates example FPLC and final HPLC chromatograms for each LTBP construct. SDS-PAGE and ESI-MS (Section 8.19 and 8.20) were then used to confirm the mass and purity of the refolded proteins (Figure 4.3).

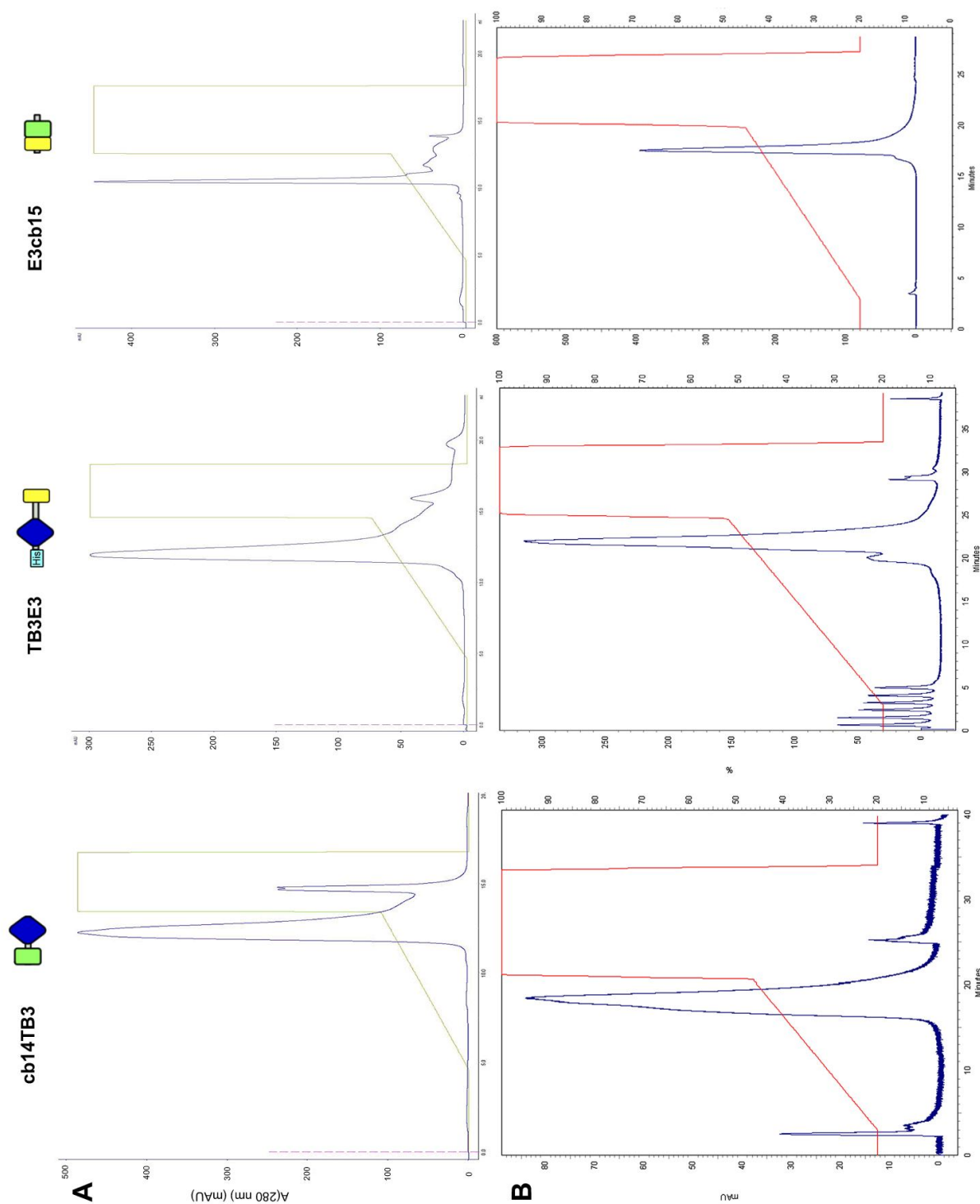
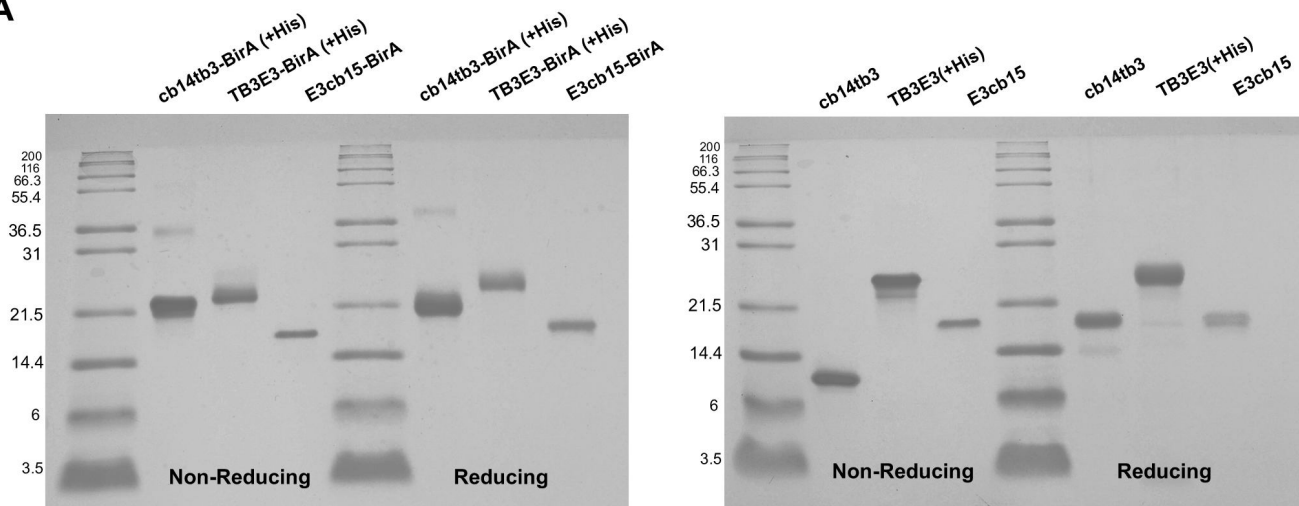


Figure 4.2 - LTP1 construct purification chromatograms

A) FPLC chromatograms of all three LTP1 constructs eluted from a MonoQ column by increasing NaCl concentration at pH 8.5 (Section 8.17). **B)** Final HPLC chromatograms of the same proteins after FPLC purification, eluted from a C8 reverse phase column by an increasing acetonitrile gradient (Section 8.13). Red and green lines represent the buffer A to B gradients. (Multiple peaks at the start of the HPLC trace of TB3E3 were due to sample injection)

A**B**

Protein	Measured Mass	Predicted Mass
cb14TB3BirA (His)	17094	17093
TB3E3BirA (His)	21734	21734.1
E3cb15BirA (His)	15910	15902
cb14TB3 (His)	15000	14998.6
TB3E3 (His)	19640	19639.6
E3cb15	11954	11954.3

Figure 4.3 - Mass and purity of LTBP1 fragments

A) Protein products analysed by SDS-PAGE and coomassie staining, under both non-reducing and reducing conditions. The preparations of BirA tagged constructs are shown on the left, and preparations of constructs lacking BirA tags are shown on the right. All the samples used in these gels had been stored at -20 °C for > 6 months, demonstrating their stability at this temperature. **B)** Table of the masses of the protein fragments, as measured by ESI-Mass spectrometry. All masses correspond to the predicted masses for each protein (within 9 Da).

4.3 Initial assessment of fold

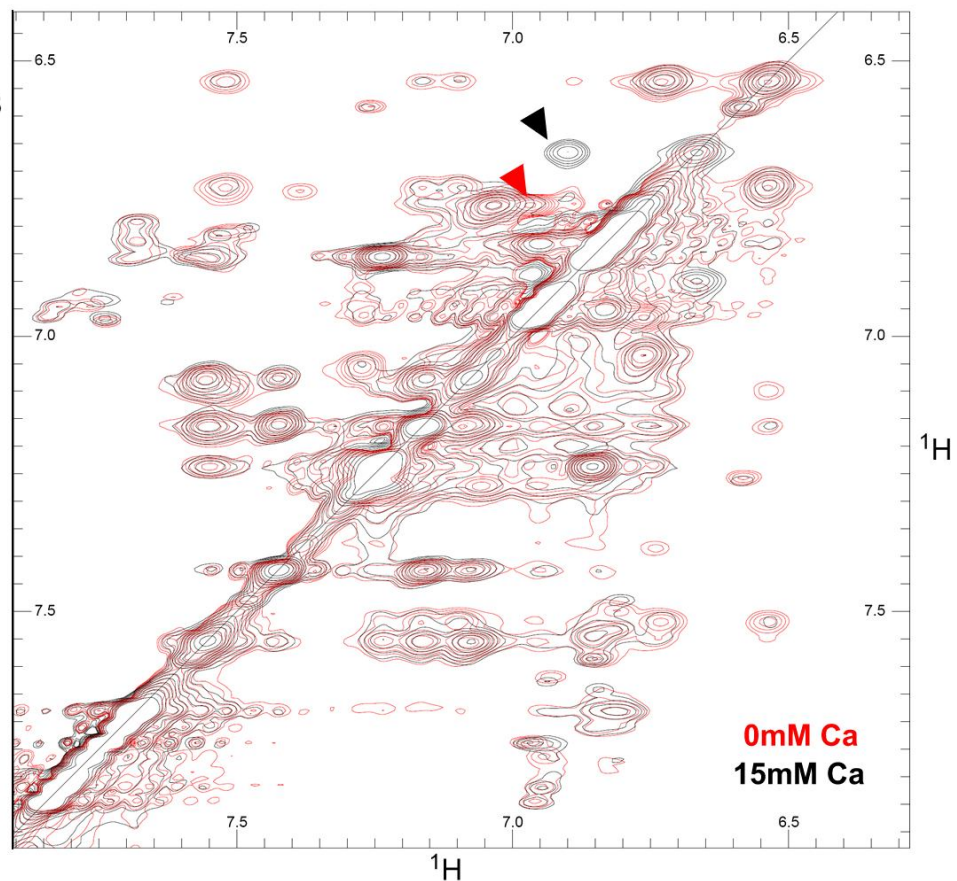
To ensure oxido-shuffling and purification had produced correctly folded LTBP1 fragments, the fold of the proteins was assessed using 2D NOESY NMR experiments. All proteins showed spectra with well-dispersed peaks, characteristic of structured proteins, as well as specific features, such as downfield shifted H^α resonances, characteristic of β sheet regions, which would be expected from the domain types present (Section 1.2).

With the cbEGF containing proteins LTBP1^{cb14tb3} and LTBP^{E3cb15}, 2D NOESY spectra were collected with and without calcium present. A sub-population peaks in the spectra of both proteins exhibited perturbations, including peaks from aromatic side chains (Figure 4.4A+B), indicative of specific calcium-binding by these proteins, and consistent with correctly folded cbEGF domains in these constructs¹.

To assay the fold of specific domains in the LTBP1^{TB3E3} construct, which contains no calcium-binding domains, spectra were compared with the overlapping LTBP1^{cb14TB3} and LTBP1^{E3cb15} constructs. Figure 4.5 compares the aromatic region of NOESY spectra from all three overlapping proteins, with shared resonances highlighted. Various peaks can be seen in spectra from LTBP1^{TB3E3}, in similar positions to those seen in LTBP1^{cb14TB3} and LTBP1^{E3cb15}. Homology models of each construct are shown above each spectrum, with aromatic residues highlighted as cyan spheres. The presence of peaks in similar positions in overlapping constructs suggests a consistent pattern of folding in their shared domains.

¹ Although non-calcium binding proteins may exhibit peak perturbations due to changes in ionic strength, these are not normally seen at these low calcium concentrations (Personal communication Prof Christina Redfield). Later assignment verified that these perturbed peaks were specific to canonical cbEGF calcium binding residues.

A
cb14TB3



B
E3cb15

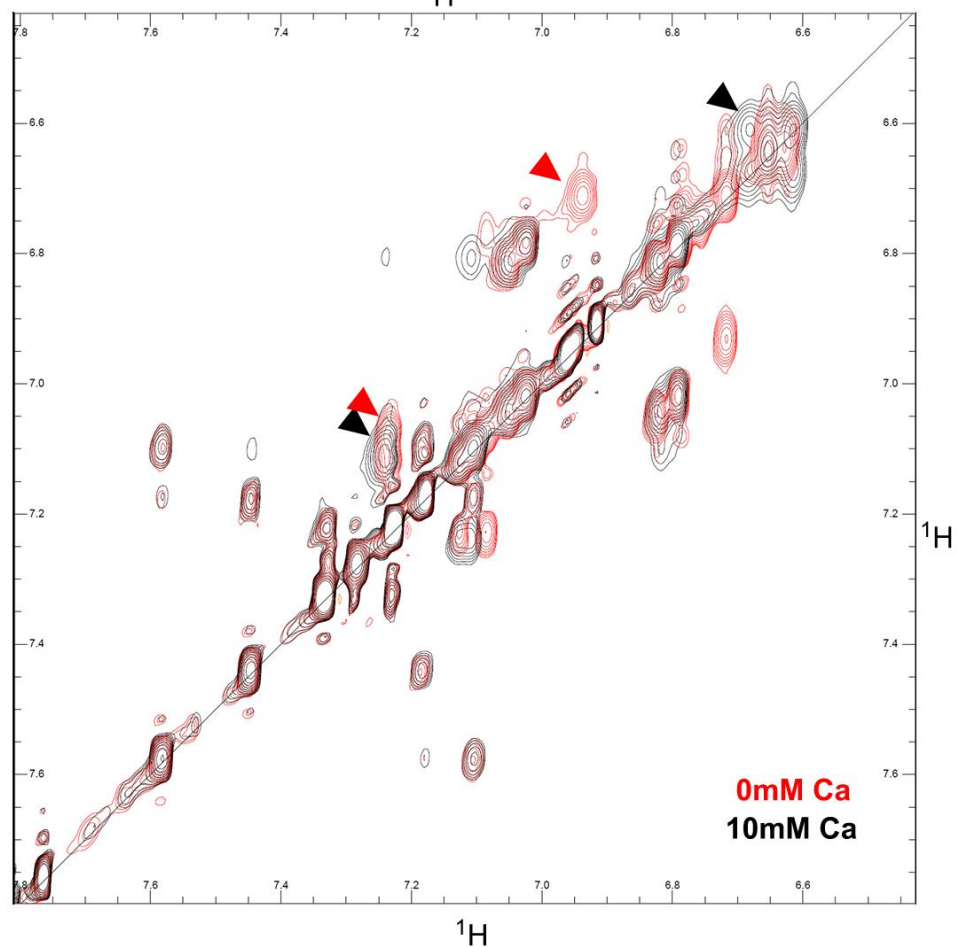


Figure 4.4 - Calcium sensitivity of 2D NOESY spectra of LTBP1^{cb14TB3} and LTBP1^{E3cb15}

750 MHz 2D NOESY spectra collected in D₂O at pH 7.7, showing just the aromatic region for simplicity. Red and black arrowheads highlight calcium-sensitive peaks. A) Overlaid spectra for BirA-tagged LTBP1^{cb14TB3} shows one clearly calcium-sensitive peak, which was later assigned to tyrosine 1490 in cbEGF14. Both spectra were collected in the same way and are contoured to the same intensity levels, (although some general loss of signal to noise seems to occur with 15mM calcium.) B) Overlaid spectra of BirA-tagged LTBP1^{E3cb15}, displaying a number of calcium-sensitive peaks, one of which was later assigned to tyrosine 1689 in the core of cbEGF15.

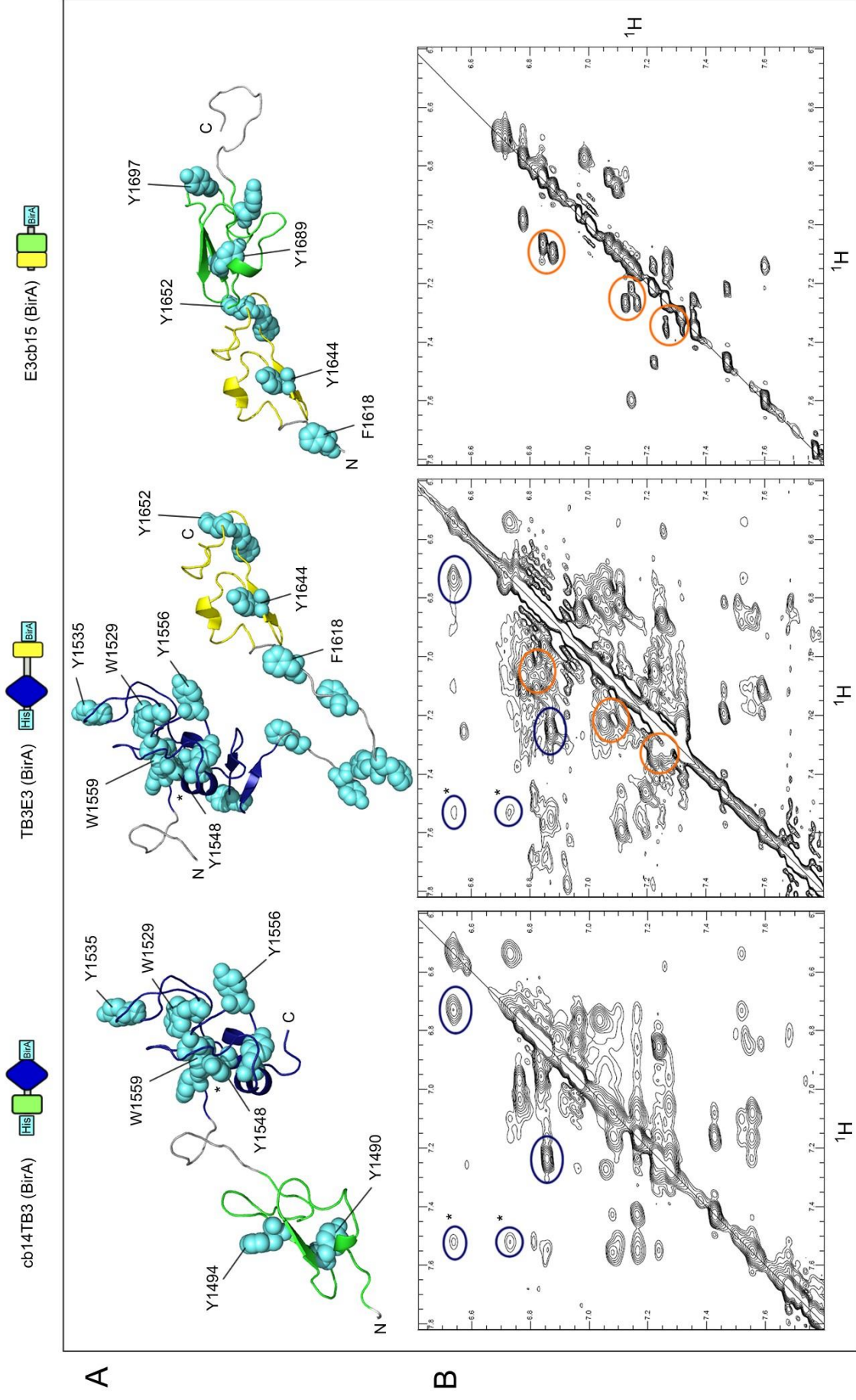


Figure 4.5 - Consistent folding of overlapping domains in LTP1 fragments

A) Homology models of LTP1 fragments used, with aromatics highlighted as cyan spheres. Ribbon coloured according to domain type (see Figure 4.1). The asterisk in cb14TB3 and TB3E3 panels highlights the tyrosine 1548 ring, which is predicted to be in close proximity to the tryptophan 1559 ring. **B)** Aromatic region of 750 MHz 2D ^1H NOESY spectra of each of the refolded BirA-tagged LTP1 fragments, focusing on the aromatic region. Blue circles highlight peaks found in both cb14TB3 and TB3E3, and likely corresponding to aromatics in the TB3 domain. Orange circles highlight peaks found in both TB3E3 and E3cb15, and potentially correspond to aromatics in the EGF3 region. The weaker peaks with asterisks next to them in cb14TB3 and TB3E3 appear to represent NOEs between tyrosine 1548 and tryptophan 1559, consistent with the homology model of the TB3 domain. (Some peaks shared in all three spectra correspond to protons from tryptophan in the BirA tag, and are not encircled). All spectra shown here were collected in D_2O at pH 7.7 without calcium present.

4.4 Backbone resonance assignment

To gain more detailed structural insights into these fragments, and for later interaction studies, backbone resonance assignment was carried out on all three LTBP1 fragments.

Xa-cleaved LTBP1^{cb14TB3} contains 117 amino acids, and it was possible to assign 88% of the LTBP1^{cb14TB3} backbone using ¹⁵N-labelled protein and ¹⁵N-edited NOESY and TOCSY experiments alone. Assignment was difficult for the final loop of cbEGF14, as the peaks were extremely weak for this region. However the number of backbone amides assigned was thought reasonable for the analyses employed here. An assigned HSQC spectrum is shown in Figure 4.6, and Table A1.3 in Appendix 1 summarises the full assignments for this protein.

LTBP1^{E3cb15} contains 108 amino acids when Xa-cleaved. For this protein backbone protons were assigned using both ¹⁵N-edited NOESY and TOCSY experiments, and also triple resonance HNCA, CBCANH, CBCACONH, HNCOC, and HNCACOC experiments (Section 8.29), as weak peaks from EGF3 made assignment difficult with double resonance experiments alone. For this protein it was possible to assign 90% of the backbone amides, and an annotated HSQC is shown in Figure 4.7, with assigned resonances summarised in Appendix 1 Table A1.4. Like cbEGF14 in LTBP1^{cb14TB3} the C-terminal loop of EGF3 also gave extremely weak peaks, preventing complete assignment of this protein

LTBP1^{TB3E3} was the largest fragment assigned with 172 amino acids (including a His tag). Because this was the last fragment to be assigned, it was possible to use peaks shared with overlapping LTBP1^{cb14TB3} and LTBP1^{E3cb15} constructs to assist with the assignment. 78% of backbone amides could be assigned using ¹⁵N-edited TOCSY and NOESY experiments alone. Figure 4.8 shows an annotated HSQC for this

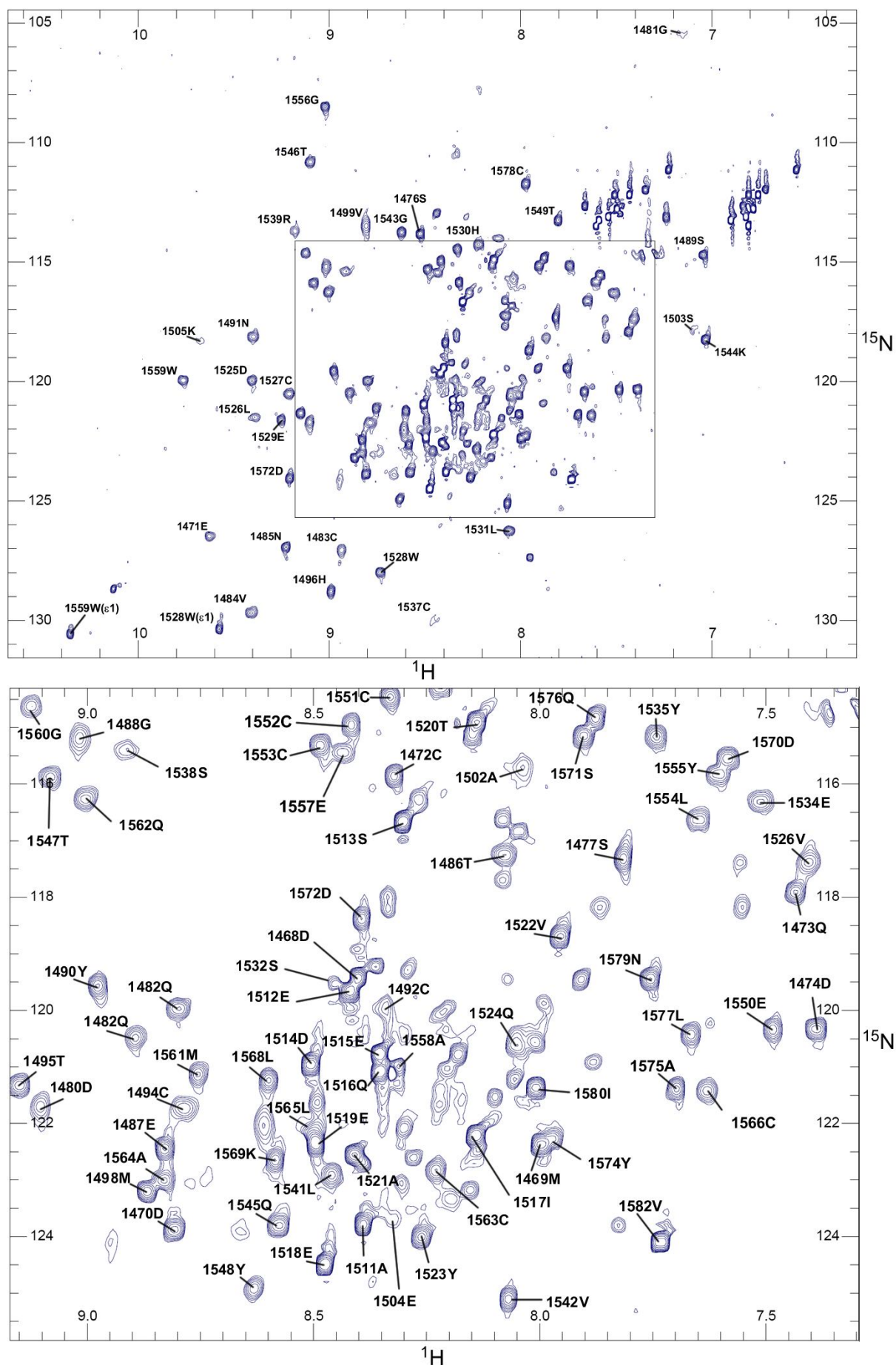


Figure 4.6 - 500 MHz HSQC of LTPB1^{cb14TB3} with peak assignments

Sample is Xa cleaved LTPB1^{cb14TB3} without a BirA tag present, at pH 5.9, with 20 mM calcium present in 95% H₂O : 5% D₂O. Peak assignments are given to the top left of each peak, except where specifically indicated by a line. Peaks without labels were not assigned and mostly displayed low or negative hetero-nuclear NOE ratios, which may be due to a small amount of misfolded material.

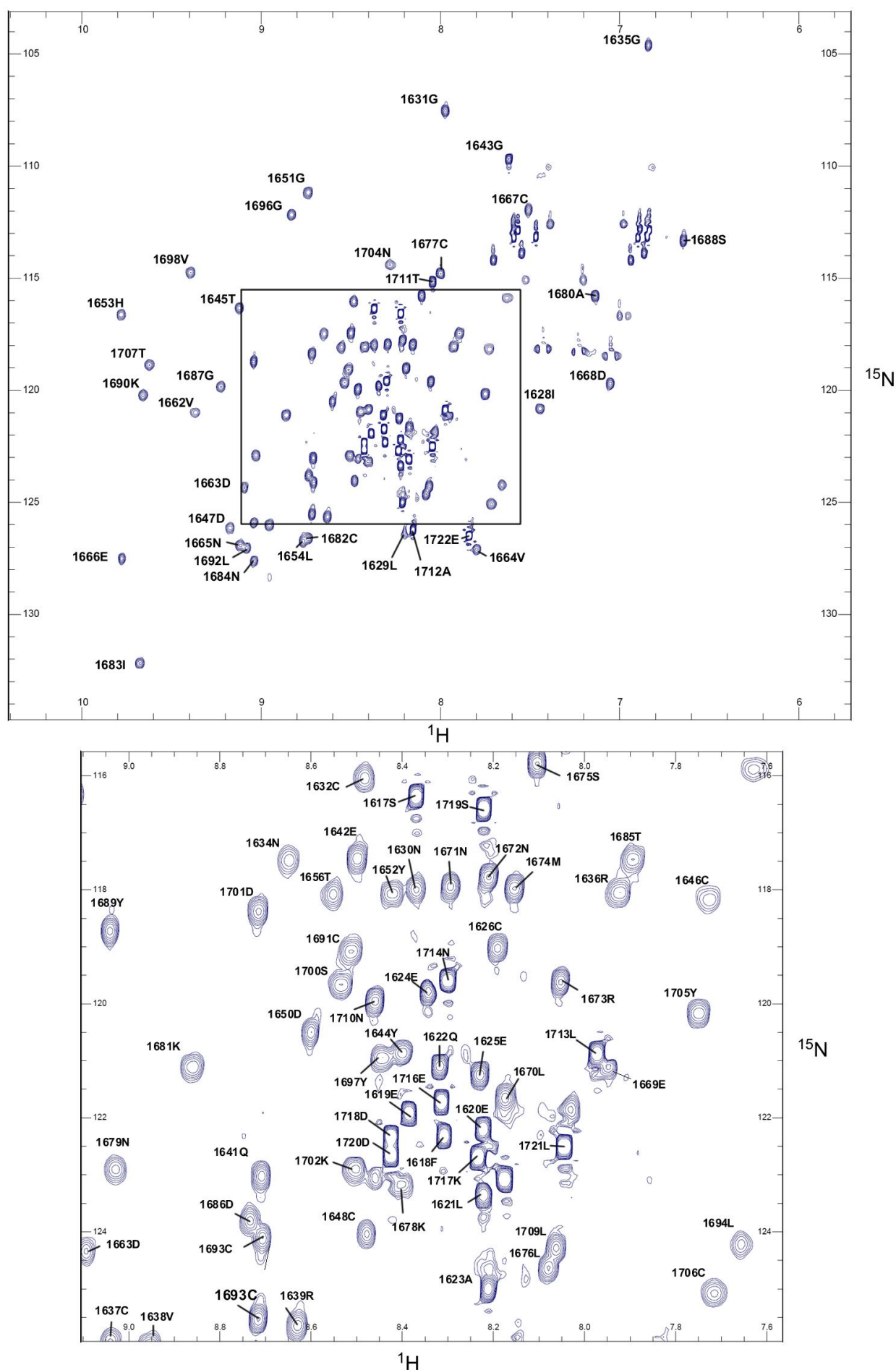


Figure 4.7 - 600 MHz HSQC of LTBP1^{E3cb15} with peak assignments

Sample is Xa cleaved LTBP1^{E3cb15} without a BirA tag present, at pH 5.4, with 10 mM calcium present in 95% H₂O : 5% D₂O. Peak assignments are given to the top left of each peak, except where specifically indicated by a line.

construct, and Table A1.5 in Appendix 1 summarises the assignments. While there are many dispersed peaks in the spectrum shown in Figure 4.8, there are also a large number of overlapping peaks in the centre of the spectrum, and these hindered complete backbone assignment for this protein.

Figure 4.9 compares HSQC spectra from all three LTBP1 fragments, collected at pH 6². Peaks assigned to the overlapping domains are consistently seen in both relevant sets of spectra (some examples of which are encircled), confirming that the TB3 domain adopts the same fold in both LTBP1^{cb14TB3} and LTBP1^{TB3E3}, and that the EGF3 domain adopts the same fold in both LTBP1^{TB3E3} and LTBP1^{E3cb15}. These data indicate that neither domain requires specific adjacent domains to be present for folding *in vitro*. The chemical shift differences between overlapping constructs (Figure 4.9B) demonstrate very little change in chemical shift between constructs. An exception is in the “GY” packing motif of EGF3, most likely as a result of a canonical packing interaction with cbEGF15 (Section 1.2.4).

4.5 Modelling the LTBP1 C-terminus

In order to facilitate the interpretation of NMR data collected for the C-terminal fragments of LTBP1, homology models were generated for LTBP1^{cb14TB3} and LTBP1^{E3cb15}, using the python based “Modeller” program (Section 8.6) [4]. The model of cbEGF14-TB3 was based on the structure of fibrillin cbEGF22-TB4 [5] (PDB 1UZJ), although, as there was no evidence found for an interface between LTBP1 cbEGF14 and TB3 domains (Section 4.8), for illustrative purposes the two domains in this model were moved apart using the manual sculpting function in pymol. The EGF3-cbEGF15

² pH 6 was chosen as all proteins were soluble at this pH. Assignments were carried out at different pH's initially in order to maintain protein solubility, while at the same time getting the pH as low as possible to aid with assignment in later fibrillin interaction experiments.

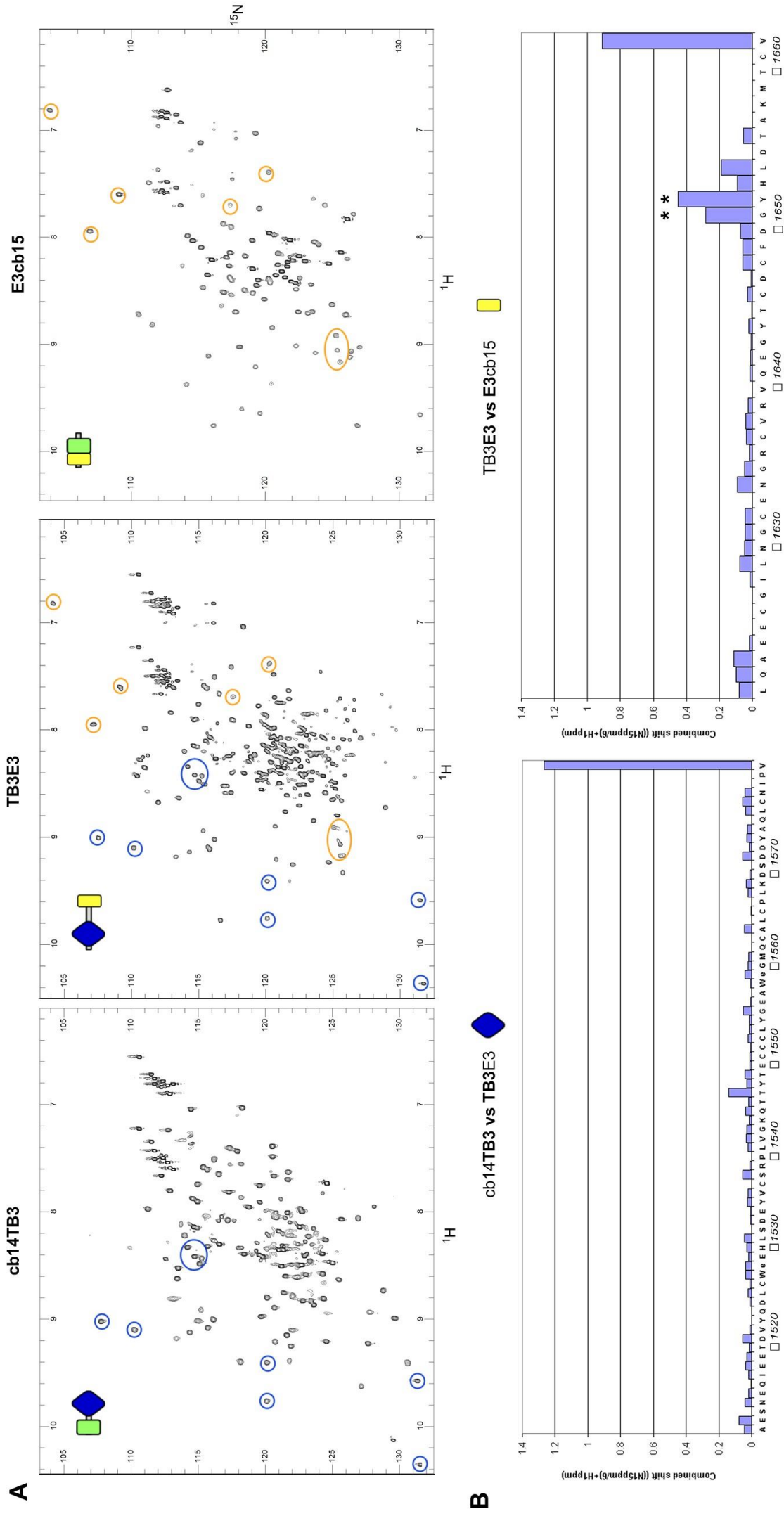


Figure 4.9 - Comparison of shared HSQC peaks in overlapping LTBP1 constructs.

A) 500 MHz HSQCs of all three constructs corresponding to the TB3 domain are highlighted with blue circles in cb14TB3, and TB3E3 spectra, while some peaks corresponding to EGF3 residues are highlighted with orange circles in TB3E3 and E3cb15 spectra. **B)** Differences in peak position between constructs are plotted for TB3 and EGF3 domains. Only small differences in peak position are seen, except in the GY packing motif of EGF3 (highlighted with **), and in the very C-terminal residues, (which are likely to be different due to the presence of a carboxylate group in one construct and a peptide bond in the other).

domain pair was modelled based on the fibrillin-1 cbEGF12-cbEGF13 domain pair [6] (PDB 1LMJ), as this structure had the closest sequence homology, including the presence of a GY interdomain packing motif in the upstream domain of both constructs. However the calcium atom in fibrillin-1 cbEGF12 was not modelled into LTBP1 EGF3.

4.6 Calcium-binding

Both LTBP1^{cb14TB3} and LTBP1^{E3cb15} constructs contain calcium-binding EGF domains, and display calcium-sensitive peaks in 2D NOESY spectra. In all cbEGF structures determined so far calcium is bound at the N-terminus of the domain by side chain and backbone carbonyl ligands, which are present at the very start of the domains, and in the first β hairpin. To see if the cbEGF14 and cbEGF15 domains in the C-terminus of LTBP1 bind calcium via the canonical residues, and to further confirm the presence of a cbEGF type fold in these domains, HSQC spectra of LTBP1^{cb14TB3} and LTBP1^{E3cb15} were collected in the presence and absence of calcium.

For LTBP1^{cb14TB3} an HSQC calcium titration was carried out (Figure 4.10A), with an HSQC collected without calcium, and with five further HSQCs collected at different calcium concentrations up to 20 mM. A number of peaks exhibited significant shifts in response to calcium, and the corresponding residues are highlighted on the homology model (Figure 4.10B). The peaks exhibiting the most significant shifts (residues highlighted on the structure in dark blue) correspond to residues at the N-terminus of the domain and the main β -hairpin, and so suggest that the homology model provides an accurate representation of the calcium-binding site.

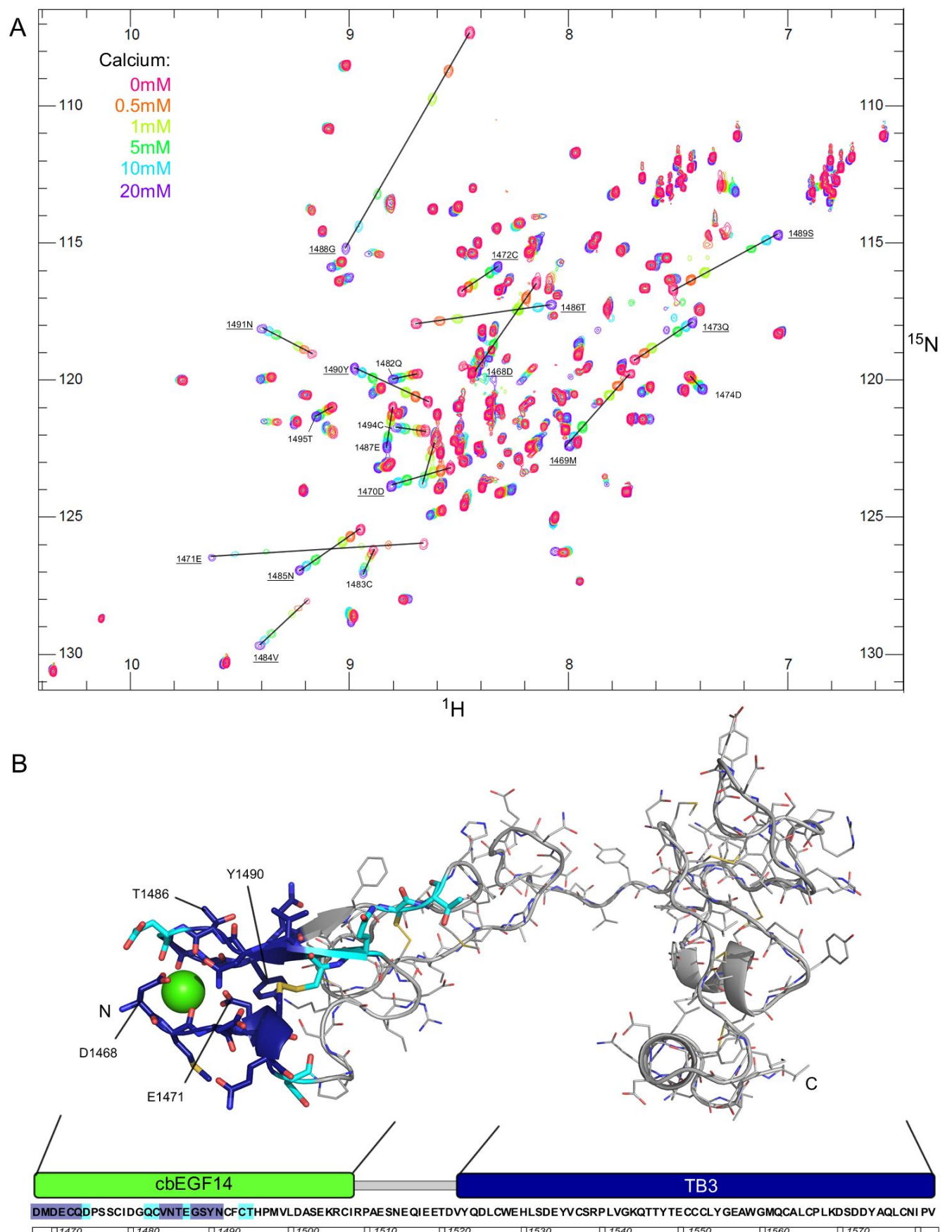


Figure 4.10 - Residues in LTBP1^{cb14TB3} affected by calcium-binding

A) Overlaid 500 MHz HSQC spectra collected with different concentrations of calcium, coloured according to the legend in the top left. Sample was prepared at pH 5.9, in 95% H₂O : 5% D₂O without any additional salt. Assignments for peaks showing significant shifts are highlighted, and those with the largest shifts are underlined. **B)** Homology model of cbEGF14TB3, with calcium shown as a green sphere and residues giving shifting peaks in A) are highlighted as stick representation, coloured cyan in the case of peaks with smaller shifts, and dark blue with larger shifts (underlined assignments). Below the homology model the calcium-sensitive peaks are highlighted in the protein sequence.

In the LTBP1^{cb14TB3} calcium titration there is still significant movement of peaks between 10 and 20 mM calcium³, suggesting the construct has not reached saturation by 10 mM, and that it has a low calcium-binding affinity ($K_d > 1$ mM). However this cbEGF domain is at the N-terminus of the LTBP1^{cb14TB3} construct, and the N-terminal domain in the native protein (cbEGF13) is not present to provide packing interactions with it. These packing interactions are important for enhancing calcium affinity in cbEGFs [6, 7], and so it is not surprising that the calcium affinity appears to be weak in this construct.

In LTBP1^{E3cb15} on the other hand, cbEGF15 is preceded by EGF3, which should be able to provide a packing interaction with cbEGF15 to stabilise the calcium-binding pocket. A preliminary 1D calcium titration of this construct demonstrated peaks appearing and strengthening, while others weakened and disappeared (Figure 4.11), demonstrating slow exchange between the calcium free and calcium bound states. Saturation also appeared to be achieved with 2 mM calcium, and suggests a moderate to high calcium affinity ($K_d < 1$ mM), (a 2 mM protein concentration used)

As the LTBP1^{E3cb15} construct exhibits slow exchange with calcium it was not possible to carry out an HSQC titration where the shifts of the peaks, and their assignments, could be directly followed. However spectra collected in the presence and absence of calcium can be compared to look at the appearance and disappearance of specific peaks. Assignment of LTBP1^{E3cb15} was carried out with 10 mM calcium present, and Figure 4.12A shows HSQC spectra with either no calcium or the addition of 10 mM calcium. Assignments are highlighted for peaks that are absent without calcium

³ The titration was not taken beyond 20 mM calcium as protein precipitation began to occur.

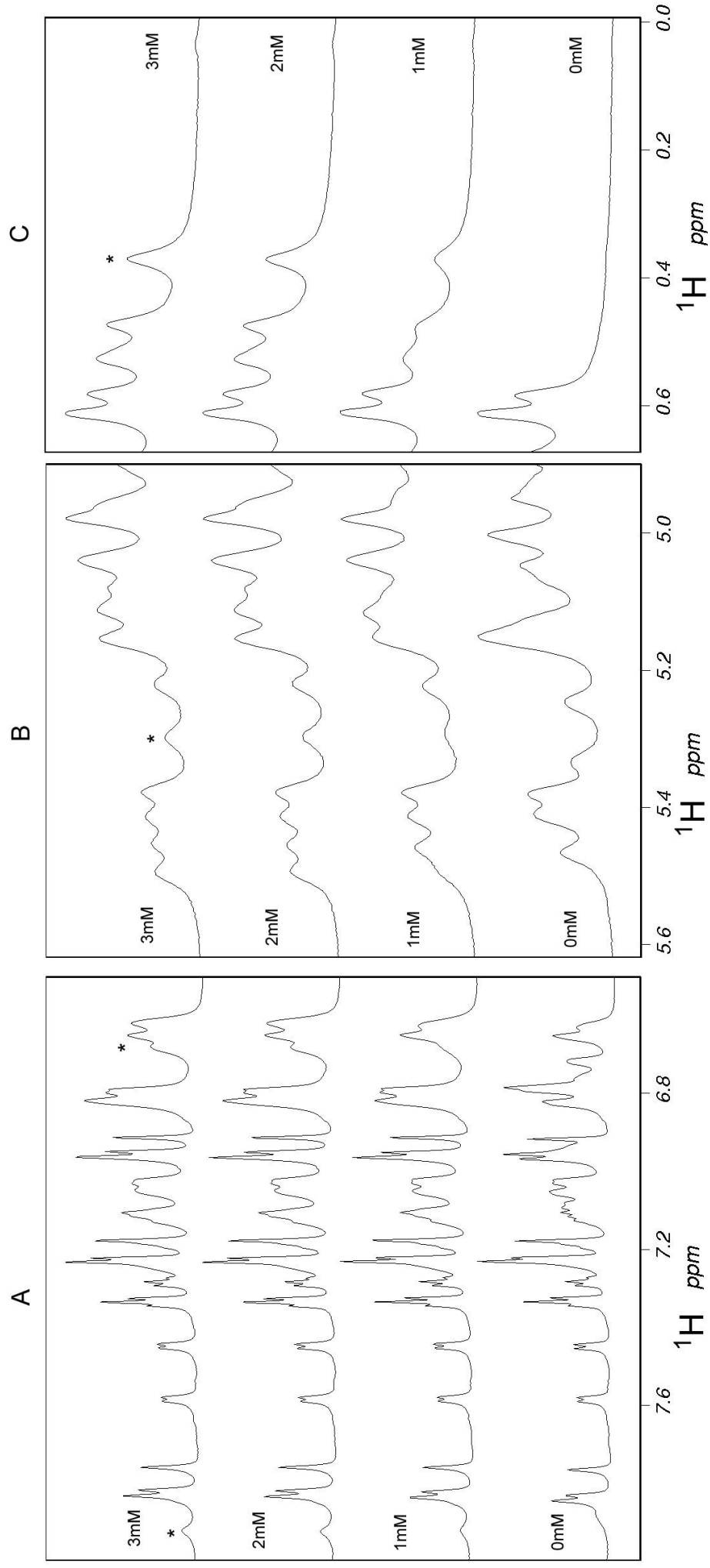


Figure 4.11 - 1D calcium titrations of LTBP1^{E3cb15}

Specific regions of the same 1D spectra of unlabelled LTBP1^{E3cb15} collected in D_2O at pH 7.7 with different concentrations of calcium in a 750 MHz spectrometer. **A)** The downfield region including aromatic peaks. **B)** The up-field shifted methyl region. **C)** The downfield shifted methyl region. In each region peaks can be seen appearing and disappearing with increased calcium. Asterisks highlight examples of peaks that appear with the addition of calcium.

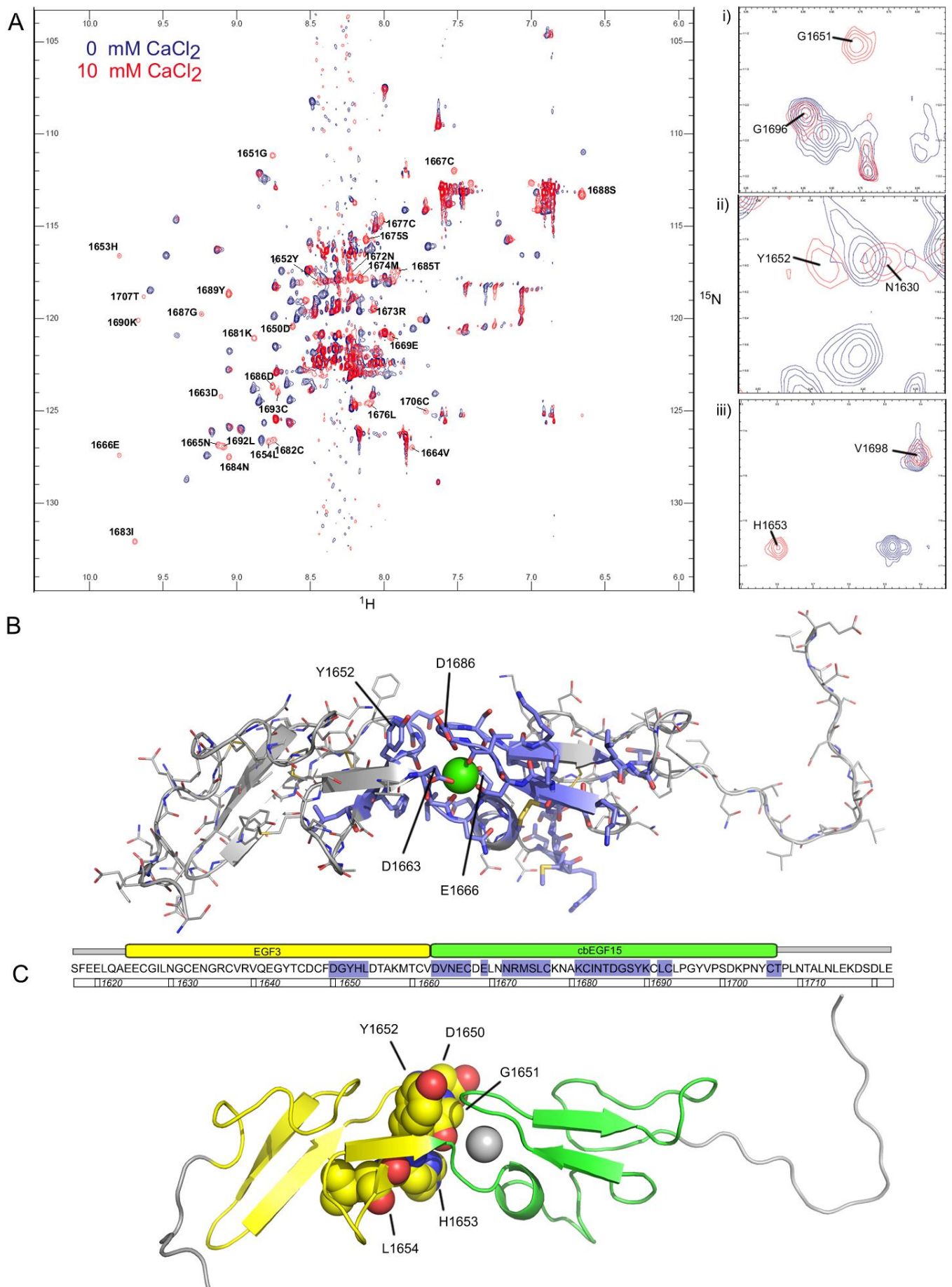


Figure 4.12 - Residues in LTBP1^{E3cb15} affected by calcium-binding

A) 750 MHz HSQC spectra recorded with and without calcium at pH 5.4. Assignments of calcium dependent peaks are highlighted and selected peaks are shown in detail on the right. **B)** Calcium dependant assignments mapped onto the homology model of EGF3-cbEGF15, shown in stick representation and coloured blue. Calcium is also shown as a green sphere, and the protein sequence with calcium dependant residues highlighted shown below. **C)** Residues in EGF3 with peaks that appear calcium-dependent and may form an interdomain packing interface with cbEGF15 are highlighted and represented as spheres. Calcium is shown here as a light grey sphere.

present. When these calcium-sensitive residues are highlighted on the homology model (Figure 4.12B) they again highlight residues that would be expected to be near the calcium-binding site of a canonical cbEGF. As well as calcium-sensitive residues in the N-terminal region and main β hairpin of cbEGF15, there is also a stretch of residues at the C-terminus of EGF3 with calcium-sensitive HSQC peaks. More detailed images of three of these peaks are shown in panels i-iii of fig 4.12A. These residues comprise the loop of EGF3 containing the “GY” interdomain packing motif, which is predicted in the homology model to pack with the β -hairpin of cbEGF15 (Figure 4.12C). The calcium sensitivity of this DGYHL loop is therefore consistent with an interaction between this region of EGF3 and the calcium-binding region of cbEGF15.

In summary, the calcium-sensitive peaks for the LTBP1^{E3cb15} construct provide both validation of the homology-modelled structure of cbEGF15, and support for the predicted packing interaction of the C-terminus of EGF3 with cbEGF15. Domain cbEGF14 in the LTBP^{cb14TB3} construct does not have an N-terminal domain to pack against, weakening its calcium affinity, but the calcium-sensitive peaks, and their corresponding residue assignments, demonstrate that cbEGF14 still adopts a standard cbEGF-like fold.

4.7 Solvent accessibility and hydrogen bonding

To further validate the homology models of the C-terminal domains of LTBP1, especially those of the TB3 and EGF3 domains that could not be directly validated by calcium-binding experiments, D₂O exchange experiments were carried out. Lyophilised ¹⁵N labelled proteins were dissolved in D₂O (described in detail in Section 8.32), and the exchange of backbone amide protons with solvent deuterium was

followed by the loss of their corresponding peaks in successive HSQC spectra. The HSQC peaks that persist indicate residues that are protected from this exchange with solvent, either through the formation of hydrogen bonds, or by being buried within a hydrophobic core. HSQC assignments can then be used to determine whether potential hydrogen bonded and core residues are consistent with the homology models proposed.

A D₂O exchange experiment with the LTBP1^{cb14TB3} construct (Figure 4.13A) demonstrates a number of slowly exchanging amide protons. Of these, only three slowly exchanging amides are found in the cbEGF14 domain, and in the homology model these are predicted to participate in hydrogen bonding in the main β -hairpin and C-terminal loop of the domain (Figure 4.13B), further supporting the homology model. No slowly exchanging amides are present in the short linker region between cbEGF14 and TB3. The TB3 domain on the other hand contains numerous slowly exchanging amides, and these are consistent with the predicted hydrogen bonds in α -helical and anti-parallel β -sheet regions. Some amides show particularly slow exchange with solvent, with their HSQC peaks persisting after more than 15 hours in D₂O (coloured dark blue in fig4.13B). These residues correspond to those predicted to be within the hydrophobic core of the TB3 domain, which likely affords them additional protection from exchange. Close inspection of amide arrangement in the homology model confirms that highlighted residues are in close proximity to potential hydrogen bonding partners, except for the backbone amides of leucine 1565 and valine 1582, and the side chain indole NH of tryptophan 1528. However, minor rearrangements in the secondary structure of the homology model could accommodate hydrogen bonding of these amide protons as well, and the rest of the D₂O exchange data strongly supports the homology model being an accurate representation of the overall TB3 domain fold.

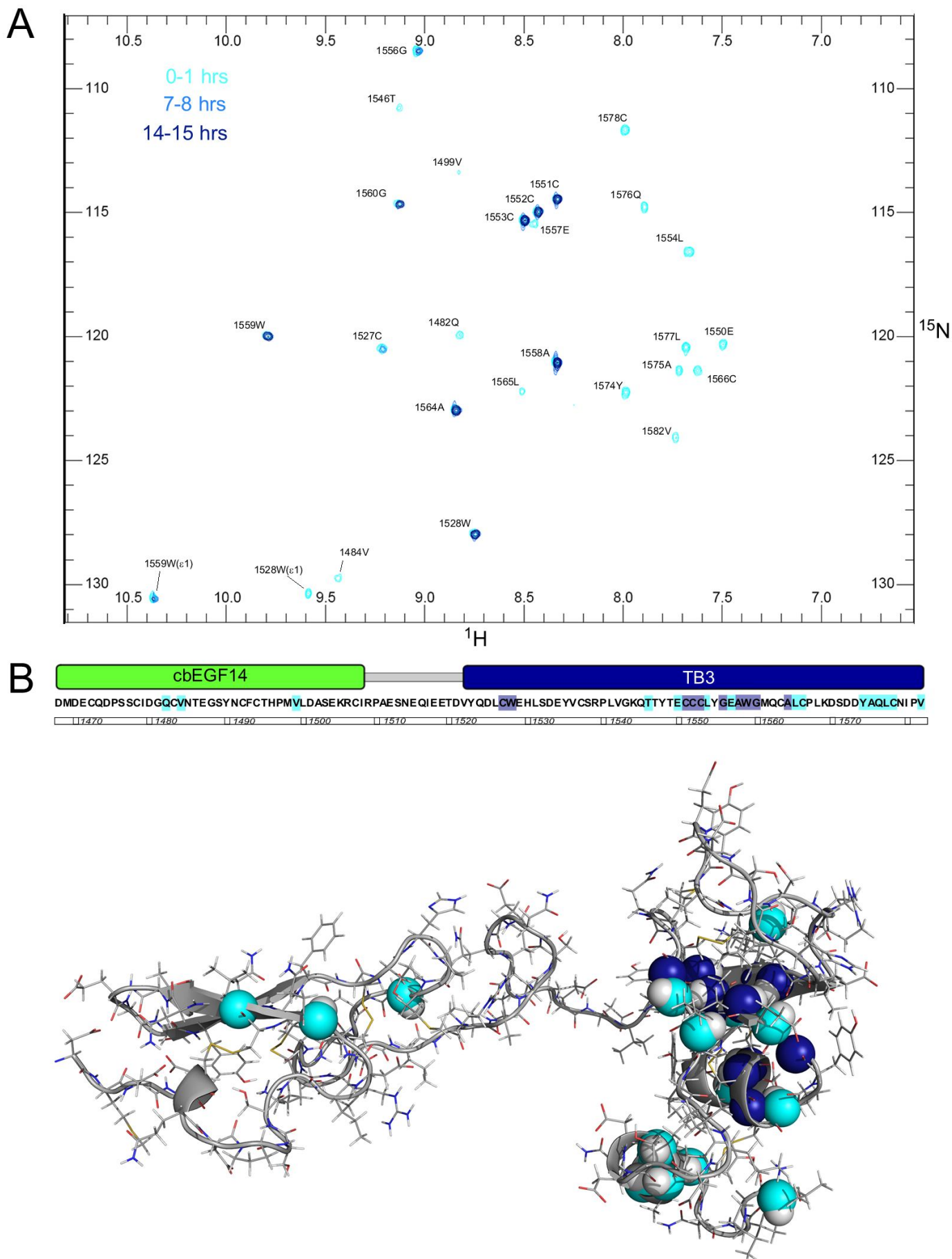


Figure 4.13 - Slowly exchanging amides in LTB1^{cb14}TB3

A) 500 MHz HSQCs collected at different time points after dissolving protein in D₂O at pH 5.9 with 20 mM calcium. Assignments are highlighted for those peaks that are still visible in D₂O. **B)** Slow exchanging amides highlighted in the protein sequence, and as spheres within a homology model, with nitrogen atoms coloured cyan if visible in the first 1 hour HSQC, and dark blue if they are still visible after 14 hours.

A D₂O exchange experiment was also carried out with a sample of LTBP^{E3cb15} (Figure 4.14), and again a number of slowly exchanging amides are seen, which persist for >15 hours in some cases. These slow exchanging amides are found in EGF3 and in cbEGF15, and in both domains they correspond well with the expected pattern of hydrogen bonding predicted by the homology model. Hydrogen bonding is especially obvious in the main β -hairpins of both EGF3 and cbEGF15, where many of the slowest exchanging amides can be found. The only slowly exchanging amides for which the hydrogen bonding partners are not immediately clear from the homology model are those of tyrosine 1689 and lysine 1702, although the peak for lysine 1702 is very weak and disappears quickly.

These two D₂O exchange experiments support the predicted secondary structure elements of all four recognised domains in the LTBP1 C-terminus, indicating that the homology models provide a good representation of overall domain fold. These models should be sufficiently accurate for use in later protein docking simulations, where various minor structural rearrangements can be tested during the calculations.

4.8 Fast time scale dynamics

To understand the structure of the LTBP1 C-terminus as a whole it is necessary to look at the protein dynamics. This is especially the case with the inter-domain linking regions, in order to determine whether they act as flexible linkers that allow the domains to function independently, or adopt specific structures, which would impose a more rigid overall structure on the region.

To investigate the sub-nanosecond dynamics of the LTBP1 C-terminus, heteronuclear NOE experiments (Section 8.33) were carried out on all three overlapping C-terminal constructs. Figure 4.15A shows the heteronuclear NOE intensity ratios for C-terminus

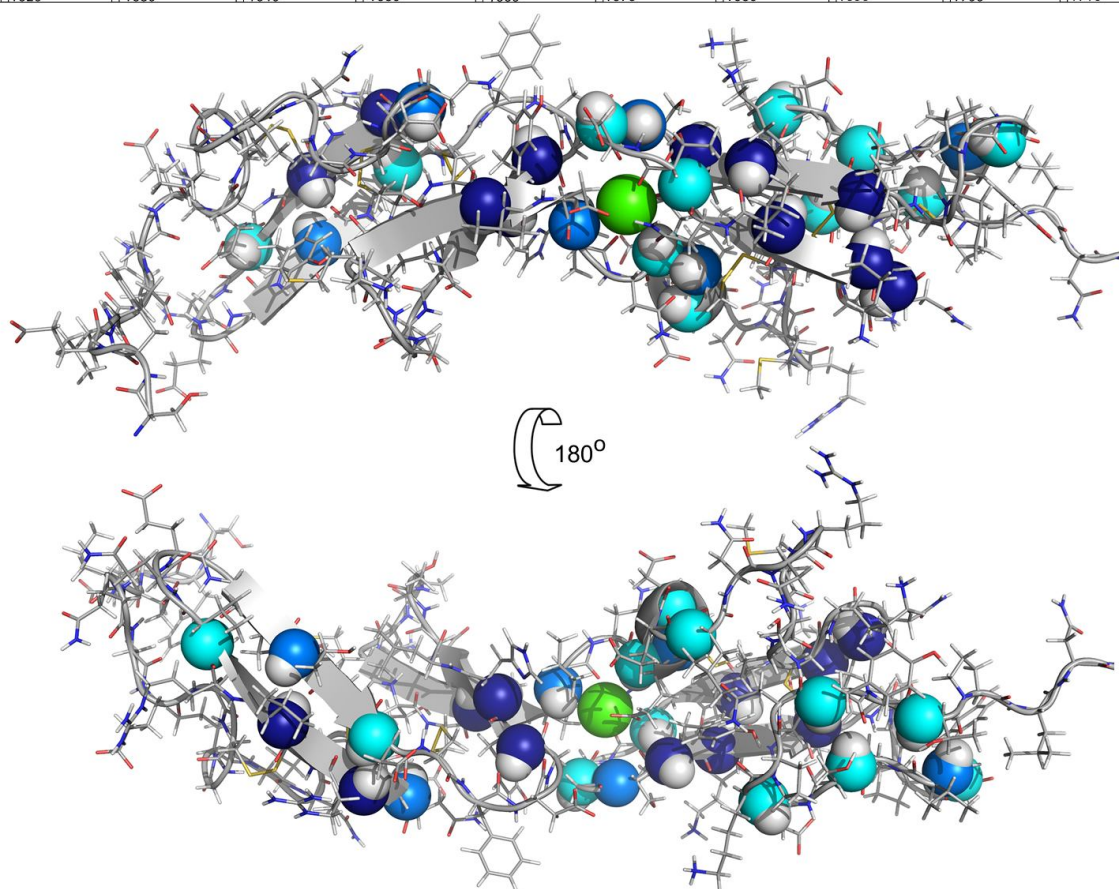
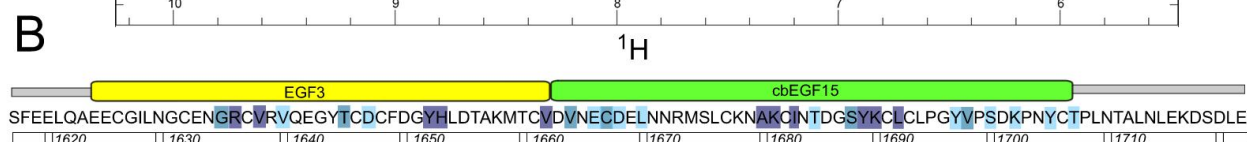
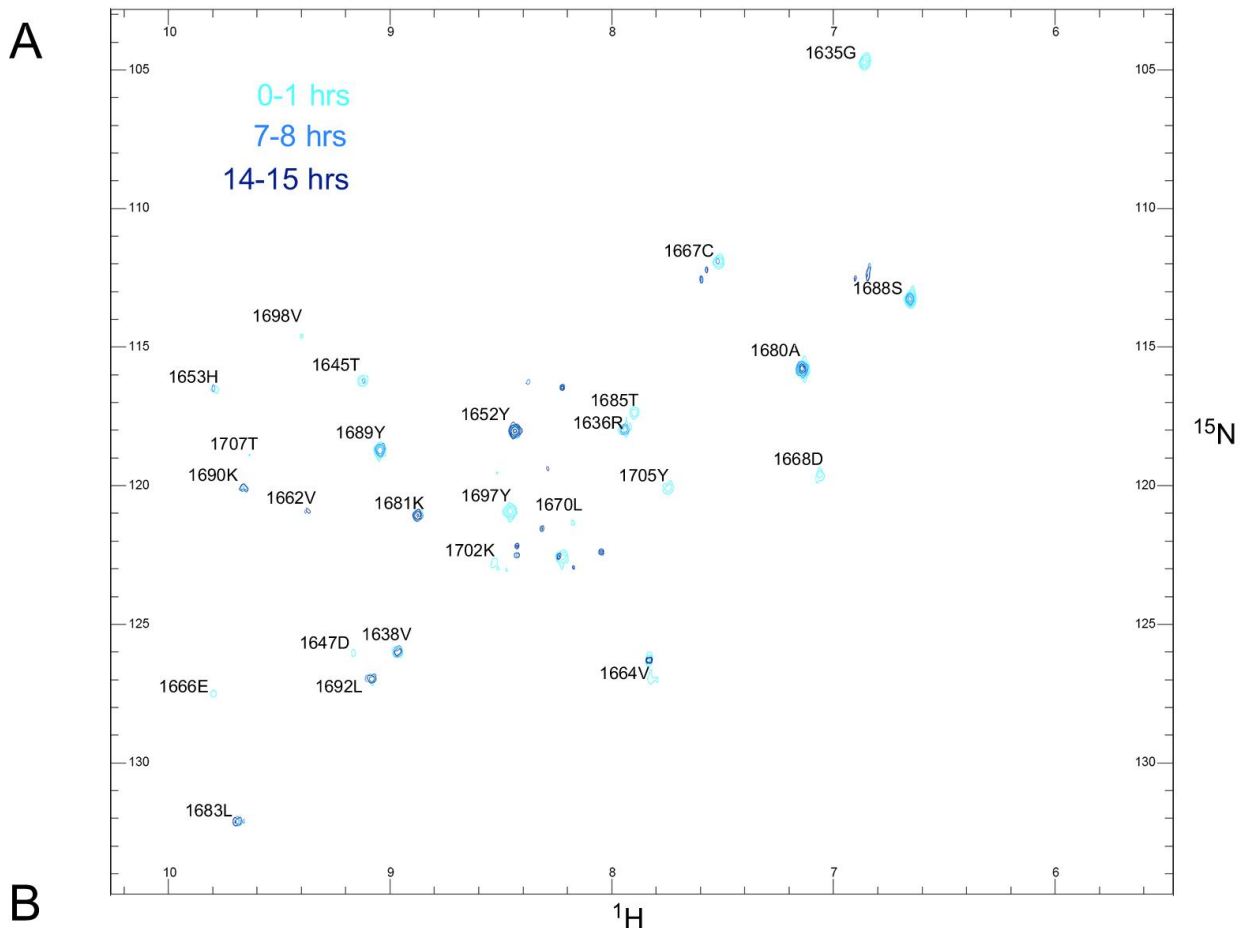


Figure 4.14 - Slowly exchanging amides in LTBP1E3cb15

A) 500 MHz HSQCs collected at different time points after dissolving protein in D_2O at pH 5.4 with 10 mM calcium. (Some weak but sharp peaks are visible at similar intensity in all D_2O HSQC spectra, and correspond to flexible N and C-terminal residues that give extremely intense peaks, visible here due to residual amounts of H_2O present). **B)** Slow exchanging amides highlighted in the protein sequence, and as spheres within the homology model. Nitrogen atoms are coloured according to the last spectrum in which their HSQC peaks are visible.

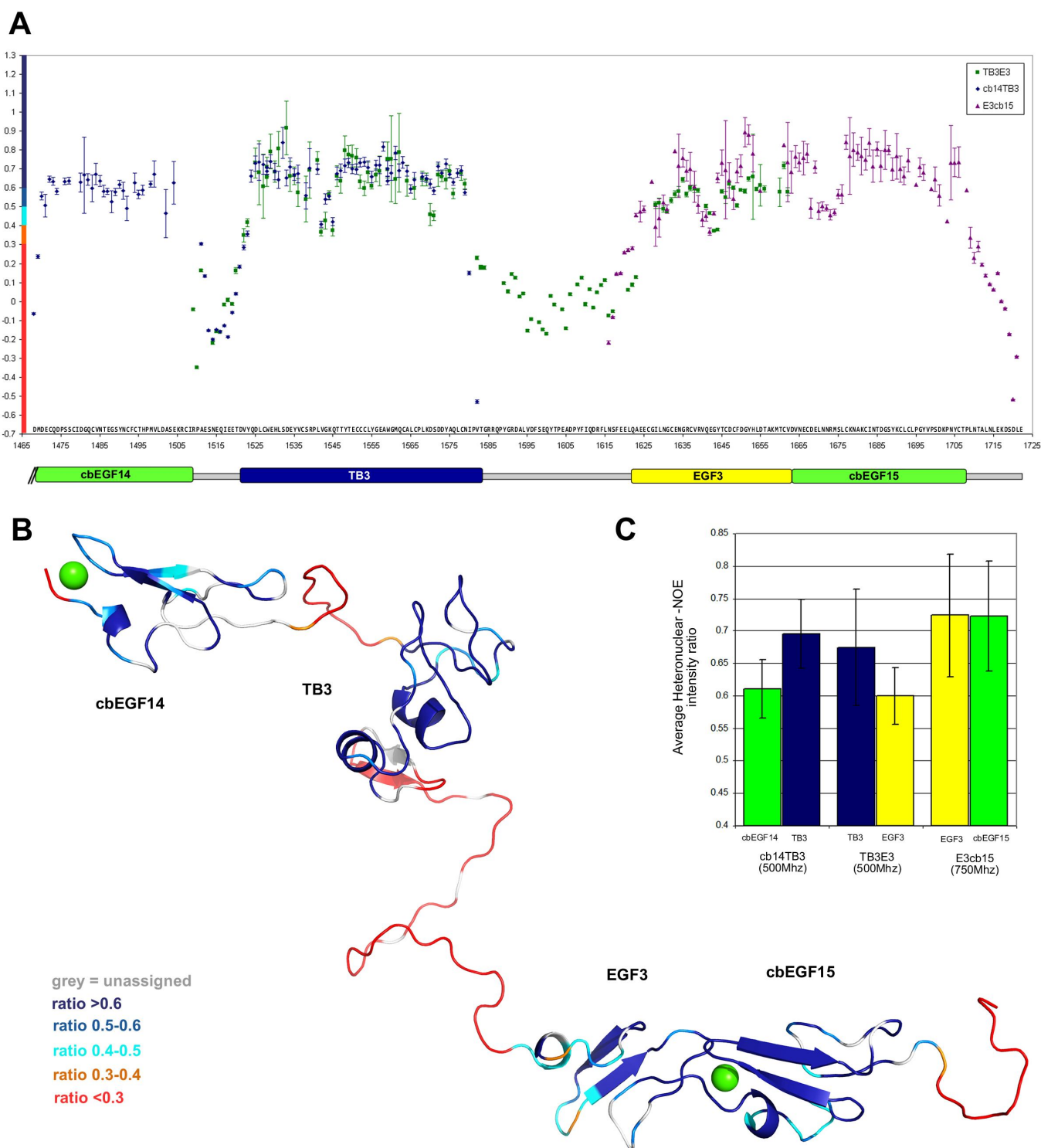


Figure 4.15 - Fast timescale dynamics of the LTBP1 C-terminus, as determined by heteronuclear NOE experiments.

A) Heteronuclear NOE ratios observed for residues in the LTBP1^{cb14TB3} (dark blue) LTBP1^{TB3E3} (green) and LTBP1^{E3cb15} (purple) constructs. A summary of the domain organisation is shown below. Error bars give the standard error for each value based on peak intensity relative to background noise (Section 8.33) **B)** Heteronuclear NOE values from (A) plotted on to a homology model of the complete LTBP1 C-terminus, with linker regions placed randomly using the pymol manual sculpting tool. Colour code legend provided at the bottom. For residues 1488-1578 values from LTBP1^{cb14TB3} are used, for residues 1579-1648 values from LTBP1^{TB3E3} are used, and for residues 1649-1722 values from LTBP1^{E3cb15} are used. **C)** Chart of the average heteronuclear NOE ratios for each domain, (values from flexible loop regions excluded). Error bars represent the standard deviation of Heteronuclear NOE ratios for that domain.

of LTBP1, collected from three different experiments on the three overlapping LTBP1 constructs, LTBP1^{cb14TB3}, LTBP1^{TB3E3}, and LTBP1^{E3cb15}. In these experiments, a lower value for the intensity ratio reflects greater flexibility (especially in the case of negative values), while a higher value suggests the residue is in a more structured region. Heteronuclear NOE values correspond well with each other in the overlapping regions of the three constructs tested, although the scale of the error on the measurements differs considerably due to different signal to noise levels with each sample.

It is clear from these data that the LTBP1 constructs examined exhibit both structured regions (heteronuclear ratios >0.6), and highly flexible unstructured regions (heteronuclear ratios <0.4). From the domain organisation summary (bottom of Figure 4.15A), it can be seen that the structured regions correspond closely with the predicted EGF, cbEGF and TB domains, while the interdomain linker regions between cbEGF14 and TB3, TB3 and EGF3, and the C-terminal tail following cbEGF15, all appear to be highly flexible and unstructured. This is consistent with their generally poor conservation (Figure 2.7 Chapter 2), although their presence may still be functionally significant by allowing free movement of the cbEGF14, TB3 and EGF3-cbEGF15 domains relative to each other.

In Figure 4.15B, a homology model of the LTBP1 C-terminus, with unstructured regions included and placed randomly, is coloured according to flexibility. As well as the large unstructured regions in the LTBP1 C-terminus, close inspection of the data reveals a number of other interesting details about the individual domains; cbEGF14 appears well structured for the most part, except for the two N-terminal residues, aspartate 1468 and methionine 1469. This is somewhat surprising, as the side chain carboxylate group of aspartate 1468 would be predicted to contribute to the calcium-binding site. However as previously discussed, this construct lacks the N-terminal

cbEGF13 domain that would normally provide packing interactions and enhance calcium affinity, and the flexibility may then be due to the behaviour of the residues in the calcium free state. Alternatively the preceding domain may provide critical structure to this region that could be required to correctly orient aspartate 1468 for calcium-binding. Either way, it is likely that the observed flexibility of these two residues is a product of the domain's behaviour in isolation, and does not reflect their true flexibility in the full-length protein.

The structured region of the TB domain appears to start from glutamine 1524, and finish with asparagine 1579, which follows immediately after the last cysteine of the TB domain. The residues following asparagine 1579, whose equivalents in the fibrillin cbEGF22-TB4-cbEGF23 crystal structure form a β -hairpin structure linking TB4 and cbEGF22 [5], appear dynamic and unstructured in LTBP1 TB3. This corresponds with observations from the analysis of isolated TB domains, such as fibrillin-1 TB6 and LTBP1 TB2, where it appears domain packing with a C-terminal cbEGF is necessary to stabilise this β -hairpin element [8, 9]. While most of the TB3 backbone appears structured and inflexible, there are four residues of the second predicted β -strand region, between the second and third cysteines of the TB domain, which have significantly increased flexibility. This region corresponds to valine 1542 - glutamine 1545, and forms a loop located at the surface of the TB domain (Figure 4.14 B highlighted in lighter blue). It is worth noting that the equivalent loop in the fibrillin1 TB4 domain forms interdomain contacts with the C-terminal cbEGF23, and so a packing interaction may be required for rigidification of these residues.

The flexible region linking TB3 and EGF3 spans 43 residues and includes the "SFEELQA" motif that immediately precedes the EGF3 domain. This motif, unlike the rest of the linker region, is well conserved (Figure 2.7B), and as it is also flexible, this

suggests significant but non-structural function for this sequence of conserved amino acids, perhaps playing a role in mediating protein interactions.

EGF3 itself is clearly structured, but does exhibit some regions of increased flexibility, most notably in the loop in between cysteine 1 and 2, and in the main β -hairpin. In a cbEGF this hairpin would be stabilised by calcium, but here, without the necessary residues to coordinate calcium-binding, it displays marked flexibility. In contrast, the main β -hairpin of cbEGF15 gives consistently high heteronuclear NOE values, and only the loop region between cysteine 1 and 2 exhibits significant flexibility in this domain. Also none of the five residues in between the last cysteine of EGF3 and the first cysteine of cbEGF15 appear flexible, which is consistent with a packing interaction between these two domains, as suggested by the calcium-binding data and comparison of peak positions (Figure 4.12 and 4.9).

The presence of flexible interdomain linkers between cbEGF14 and TB3, and between TB3 and EGF3, supports a model of a flexible LTBP1 C-terminus, with independent movement of the TB3 domain, and EGF3-cbEGF15 domain pair. However the presence of these flexible linkers by themselves is not conclusive evidence for this model, as interdomain interactions could still occur, with the intermediate linker regions acting as extremely large flexible loops protruding from interacting domains. Comparison of the average heteronuclear NOE values for the TB domain with those of the cbEGF14 domain or the EGF3 domain, excluding flexible loops (Figure 4.15C), hints that in both $\text{LTBP1}^{\text{cb14TB3}}$ and $\text{LTBP1}^{\text{TB3E3}}$, the residues in the larger TB domain display lower flexibility on average than those in the smaller cbEGF14 and EGF3 domains. This would be consistent with the larger domain tumbling more slowly in solution, independently of smaller domains. On the other hand in the EGF3-cbEGF15 pair, which is predicted to pack together and tumble as one unit, the average

heteronuclear-NOE values are very similar for both domains. Although the variability in the heteronuclear NOE values in all domains make these comparisons inconclusive.

Clearer evidence for the lack of interfaces between cbEGF14 - TB3 and TB3 - EGF3 comes from the earlier comparison of peak position in HSQCs from the three overlapping LTBP constructs (Figure 4.9). In the LTBP1^{cb14TB3} and LTBP1^{TB3E3} constructs the context of the TB3 domain is different, with the N-terminal domain absent in one, and the C-terminal domain absent in the other. The same is true for EGF3 in LTBP1^{TB3E3} and LTBP1^{E3cb15}. However in all overlapping regions the only residues where significant differences in peak position are seen between constructs is in the “GY” motif of EGF3, which is predicted to pack with cbEGF15. The lack of other changes in peak position demonstrates a lack of any significant interface between cbEGF14 and TB3, and between TB3 and EGF3. Therefore the unstructured inter-domain regions should act as flexible linkers rather than flexible loops.

4.9 Conclusions

These NMR studies have validated structural models for the four C-terminal domains of LTBP1, and also given significant insight into the overall dynamics and interdomain interactions of this region.

The numerous flexible regions would make many regions of the LTBP1 C-terminus unsuitable for crystallography, or structural investigations with techniques that rely on rigid body modelling. Structure determination of the C-terminal domains of LTBP1 using NMR data may be possible, based on the quality of the spectra obtained. However the lack of novel inter-domain interactions, and the relative accuracy of the homology models, validated by calcium-binding and D₂O exchange experiments, make this unnecessary for the purposes of this thesis.

From these studies a “double-jointed tail” model for the C-terminus of LTBP1 can be put forward, in which the linkers either side of the TB3 domain act as the “joints” around which the other domains can move freely. These studies of limited fragments cannot rule out the possibility that novel long-range inter-domain interactions may occur in larger fragments, for example between cbEGF14 and EGF3/cbEGF15, which would not be present in the limited fragments studied. This possibility may be worth examining further if any interesting interdomain synergy emerges in experiments with larger constructs. However the experiments in the rest of this thesis continue to utilise these double-domain fragments, and so the double-jointed tail model of the LTBP1 C-terminus should be sufficient for their interpretation.

References

1. Isogai, Z., et al., *Latent transforming growth factor beta-binding protein 1 interacts with fibrillin and is a microfibril-associated protein*. J Biol Chem, 2003. **278**(4): p. 2750-7.
2. Massam-Wu, T., et al., *Assembly of fibrillin microfibrils governs extracellular deposition of latent TGF beta*. J Cell Sci. **123**(Pt 17): p. 3006-18.
3. Ono, R.N., et al., *Latent transforming growth factor beta-binding proteins and fibulins compete for fibrillin-1 and exhibit exquisite specificities in binding sites*. J Biol Chem, 2009. **284**(25): p. 16872-81.
4. Eswar, N., et al., *Comparative protein structure modelling using MODELLER*. Curr Protoc Protein Sci, 2007. **Chapter 2**: p. Unit 2 9.
5. Lee, S.S., et al., *Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization*. Structure, 2004. **12**(4): p. 717-29.
6. Smallridge, R.S., et al., *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, 2003. **278**(14): p. 12199-206.
7. Downing, A.K., et al., *Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders*. Cell, 1996. **85**(4): p. 597-605.
8. Yuan, X., et al., *Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils*. EMBO J, 1997. **16**(22): p. 6659-66.
9. Lack, J., et al., *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, 2003. **334**(2): p. 281-91.

Chapter 5- Studies of the interactions between the Fibrillin1 N-terminus and LTBP1 C-terminus

5.1 Introduction

There is a clear connection between fibrillin microfibrils and TGF β signalling (Chapter 1, Section 1.3.4), and the observation of an interaction between LTBP1 and fibrillin1 may be significant for this biological link between fibrillin and TGF β [1-3]. At present however, the mechanism for this functional linkage is unclear.

Previous studies of the LTBP1 - fibrillin1 interaction have narrowed the binding regions to a small group of domains at the N-terminus of fibrillin and the C-terminus of LTBP1 [1-3], but they have not determined the precise binding sites, or specific amino acids involved. The aim of this chapter is to determine those binding sites, using the constructs described in chapters three and four, and to use the data obtained to generate a structural model for the complex formed by the interaction of the LTBP1 C-terminus and fibrillin N-terminus.

The fibrillin1 – LTBP1 complex likely represents a novel protein-protein interaction, with no structures of any homologous complexes available. A detailed understanding of the LTBP1 – fibrillin1 interaction may provide an important piece of the puzzle in dissecting the connection between fibrillin and TGF β signalling. Once this mechanism is more clearly established, a structural understanding of the interaction may also be relevant to the development of therapeutics to specifically target the regulation of TGF β signalling by fibrillin1.

5.2- LTBP1 and fibrillin1 protein fragments

The protein constructs used for studies of the LTBP1 - fibrillin1 interaction (Figure 5.1) have been described previously (Section 3.2 and 4.2), except for the BirA – tagged

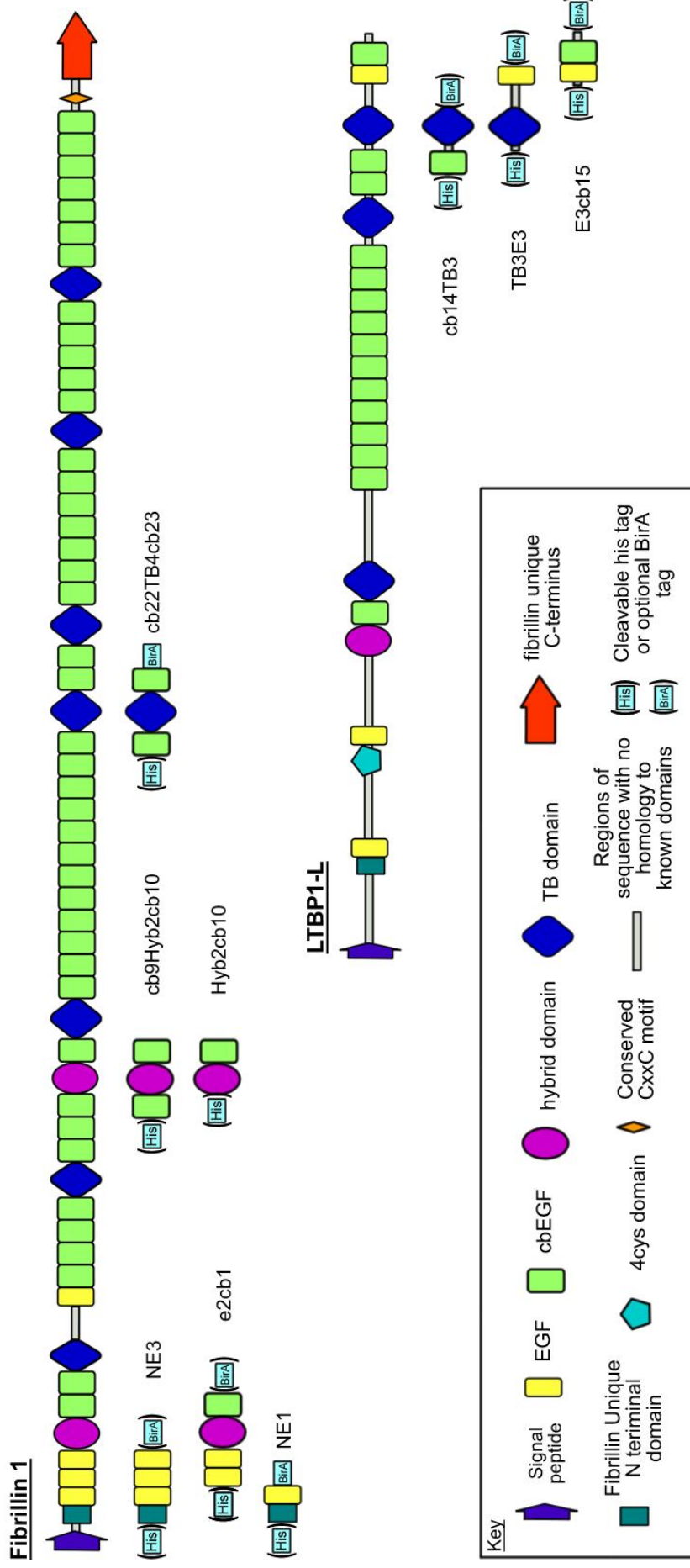


Figure 5.1 - Constructs used in interaction experiments

Summary of the fibrillin 1 and LTBP1 constructs used in this chapter. NE1 and hyb2cb10 constructs were used as negative controls in the pull down assays, while cb9hyb2cb10 and cb22TB4cb23 constructs were used as controls in the SPR experiments. Hyb2cb10, cb9hyb2cb10, and cb22TB4cb23 proteins were kindly provided by S. Iqbal and S. Jensen and were produced as previously described [6, 7]. Proteins were expressed with cleavable His-tags at their N-termini, and in some cases could also be expressed with a C-terminal BirA tag.

version of FBN^{e2cb1}, which was produced separately for pull down studies (See Materials and Methods, Figure 8.1).

Additionally, several other fragments of fibrillin were used in these experiments including the FBN^{NE3} fragment, which consists of the first four domains of fibrillin1: the unique N-terminal domain, and EGF domains 1, 2, and 3. This fragment was used to examine the requirement for the hybrid1-cbEGF1 domains in specific elements of the LTBP1 interaction. This FBN^{NE3} protein was produced as described in Section 8.9, and has been extensively characterised, including determination of its structure by NMR (D. Yadin personal communication). The other fibrillin fragments shown in Figure 5.1 were used as controls in various interaction experiments were kindly provided by Dr S. Jensen and Dr S. Iqbal.

5.3 Pull down studies

For preliminary confirmation that LTBP1 and fibrillin1 fragments interacted, pull down experiments were performed using His-tagged LTBP fragments bound to Nickel-NTA beads, to pull down biotinylated fibrillin fragments, which were then detected by Western blotting using a streptavidin-HRP conjugate (details in Section 8.24).

In the streptavidin blot, strong bands corresponding to pulled down biotinylated FBN^{e2cb1} were visible where beads were incubated with two specific LTBP fragments, LTBP^{TB3E3} and LTBP^{cb14TB3} (Figure 5.2A), suggesting an interaction between these LTBP fragments and FBN^{e2cb1}. No band corresponding to the negative control biotinylated FBN^{NE1} fragment was pulled down by the LTBP coated beads, but a weak FBN^{e2cb1} band was seen from pull downs with control FBN^{hyb2cb10} protein on the beads, and beads alone. This suggests some non-specific binding of FBN^{e2cb1} to the beads, but attempts to optimise conditions by varying levels of salt, imidazole, and detergent, did not succeed in eliminating this band. In the LTBP^{TB3E3} lanes an additional band

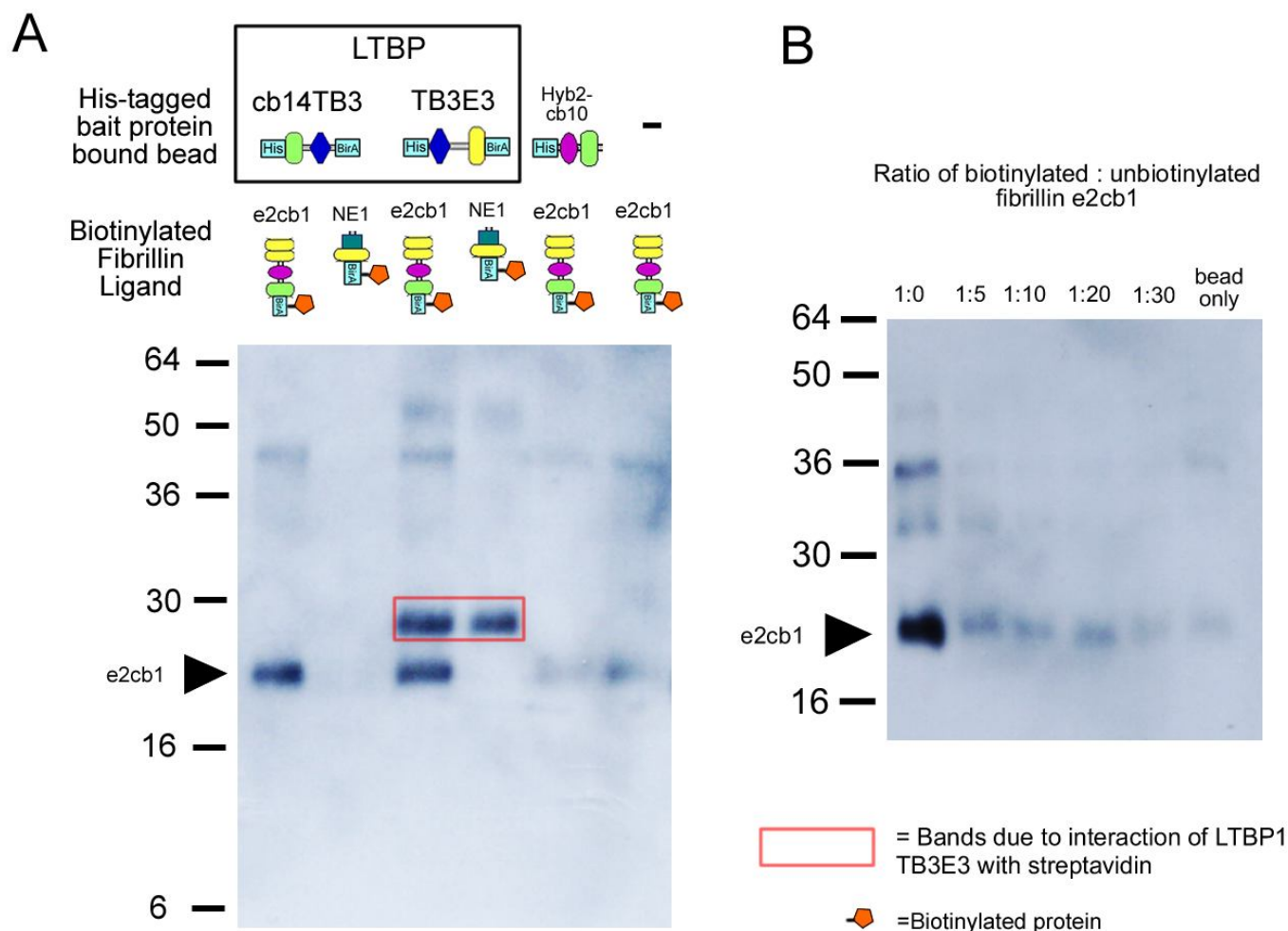


Figure 5.2 - Pull down experiments to verify the interaction between LTBP1 and fibrillin1 fragments.

A) Results of pull downs performed with Ni-NTA beads (Section 8.24), Western blots shown were developed with a streptavidin-HRP conjugate, detecting biotinylated proteins pulled down by His-tagged proteins bound to beads. A separate blot with antibodies specific to His-tagged proteins confirmed initial binding of bait proteins to beads (not shown). A dark band corresponding to biotinylated FBN^{e2cb1} can be seen pulled down by beads incubated with both LTBP1 proteins (highlighted by an arrow on the left), but is significantly weaker when a control protein or no protein is on the beads. The bands boxed in red are running at the same position as LTBP1^{TB3E3} and are an artefact due to the interaction of this LTBP fragment with streptavidin (it is also visible on control western of LTBP1^{TB3E3} BirA alone (not shown)). **B)** Competition for LTBP1 binding by addition of un-biotinylated FBN^{e2cb1}. Bands in this streptavidin blot corresponding to biotinylated FBN^{e2cb1} can be seen getting weaker as increasing amounts of unbiotinylated material are added. A separate blot with an anti-His tag antibody revealed equal loading of the beads with LTBP fragments in all competition pull downs (not shown). Contrast in these images has been adjusted to improve clarity. Data for this figure were generated by part II student Isabelle Osuch under candidate's supervision.

was also present, migrating at the expected position for LTBP^{TB3E3}. As the LTBP^{TB3E3} protein used in this experiment also had a BirA tag, this band could be due to a small fraction of the LTBP being biotinylated when expressed in *E. coli*, or another specific interaction between this region of LTBP and streptavidin.

To further confirm that the pull down interaction between LTBP-beads and fibrillin was specific, a competition experiment was performed, in which un-biotinylated FBN^{e2cb1} was also added at increasing concentration, to compete with biotinylated FBN^{e2cb1} (Figure 5.2B). A clear reduction of the biotinylated fibrillin band is seen in these experiments, confirming a specific interaction between LTBP^{cb14TB3} and FBN^{e2cb1}, which justified further examination with other biophysical techniques.

5.4 Surface plasmon resonance studies

Surface plasmon resonance (SPR) was used to examine the binding of all three overlapping LTBP1 protein fragments to FBN^{e2cb1} and FBN^{NE3} fragments. Due to the high pI values of the fibrillin constructs, which could potentially affect their solubility in buffers in the range of pH 7-8, the fibrillin constructs were chosen for amine coupling to the CM5 sensor chip (Section 8.26). The sensor chips used had four separate flow cells, one of which was coupled to FBN^{e2cb1}, one with FBN^{NE3}, and one with FBN^{cb22TB4cb23} as a negative control. The last cell was treated with coupling and blocking reagents alone, and used for blank subtraction of the sensorgram data. In all flow cells, coupling was successfully achieved to 1000 RU, and stable after regeneration with 50 mM HCl.

The LTBP fragments were then flowed over at various concentrations, using the “single cycle” biacore program (GE Healthcare) (Section 8.26), to allow for more rapid data collection. Figure 5.3 summarises the blank subtracted sensorgrams obtained, in which clear differences in the interaction behaviour of the three overlapping LTBP1

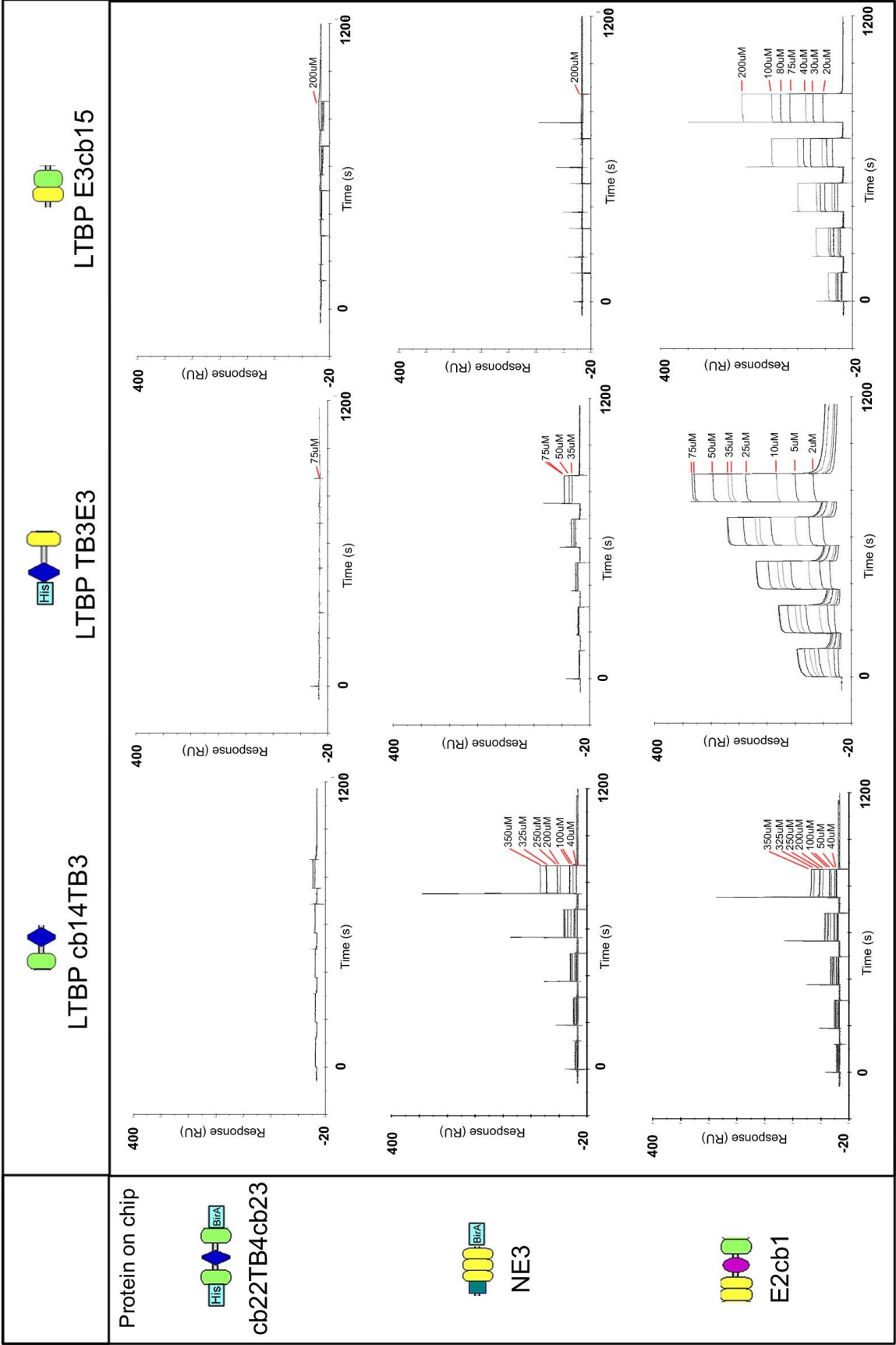


Figure 5.3 - SPR interaction studies between fibrillin1 and LTBp1 fragments.

Sensorgrams were collected using the single cycle Biacore method (Section 8.26), in which five samples of increasing concentration were injected sequentially. For each injection run the highest concentration sample is labelled in the figure, (other samples are a two fold serial dilution of this sample (see Section 8.26, Figure 8.4)). Sensorgrams are arranged in a table pattern, with the protein coupled to each flow cell of the chip in the left hand column, and the analytes flowed over these cells shown in the top row. Response from the blank flow cell subtracted from all sensorgrams. Analyte proteins were injected over all flow cells simultaneously, exposing them to identical analyte concentrations.

constructs can be observed. None of the LTBP1^{cb14TB3}, LTBP1^{TB3E3}, or LTBP1^{E3cb15} constructs showed any significant interaction with the negative control flow cell (fibrillin cb22TB4cb23, top row), additionally when the fibrillin cb9hyb2cb10 protein fragment was flowed over as a negative control analyte, (at several concentrations up to 50 μ M), no interaction with any of the fibrillin fragments on the chip was seen (data not shown).

All three overlapping LTBP1 constructs can be seen to interact with the FBN^{e2cb1} construct (bottom row). The LTBP1^{TB3E3} fragment gives considerably higher responses at lower concentrations than the other two constructs, especially LTBP1^{cb14TB3}, which only gives a measurable binding response with protein concentrations over 5 μ M. The flow cell coupled with FBN^{NE3} on the other hand only displays an interaction with the LTBP1^{TB3E3} and LTBP1^{cb14TB3} fragments, and LTBP1^{TB3E3} gives a considerably weaker response binding FBN^{NE3} than in binding the FBN^{e2cb1} flow cell. However with LTBP1^{cb14TB3} a similar response was seen with both FBN^{e2cb1} and FBN^{NE3}. These results were reproducible with two different chips, and separate protein preparations. Preliminary experiments with BirA tagged proteins also gave similar results.

To more easily compare the binding levels between various fragments, and to attempt to quantify the affinity of the interactions, plots of response versus concentration were generated (Figure 5.4). In these plots the difference between the LTBP1 constructs binding FBN^{e2cb1} is clearly apparent. Precise determination of binding affinity was complicated however by difficulties reaching saturated binding. With the weakly binding LTBP1^{E3cb15} and LTBP1^{cb14tb3} fragments this was a limitation brought about by the high K_d and solubility limits of the proteins (also sensorgrams had insufficient curvature for K_d determination by kinetic methods). However for LTBP1^{TB3E3} the relationship between concentration and sensorgram response is not what would be

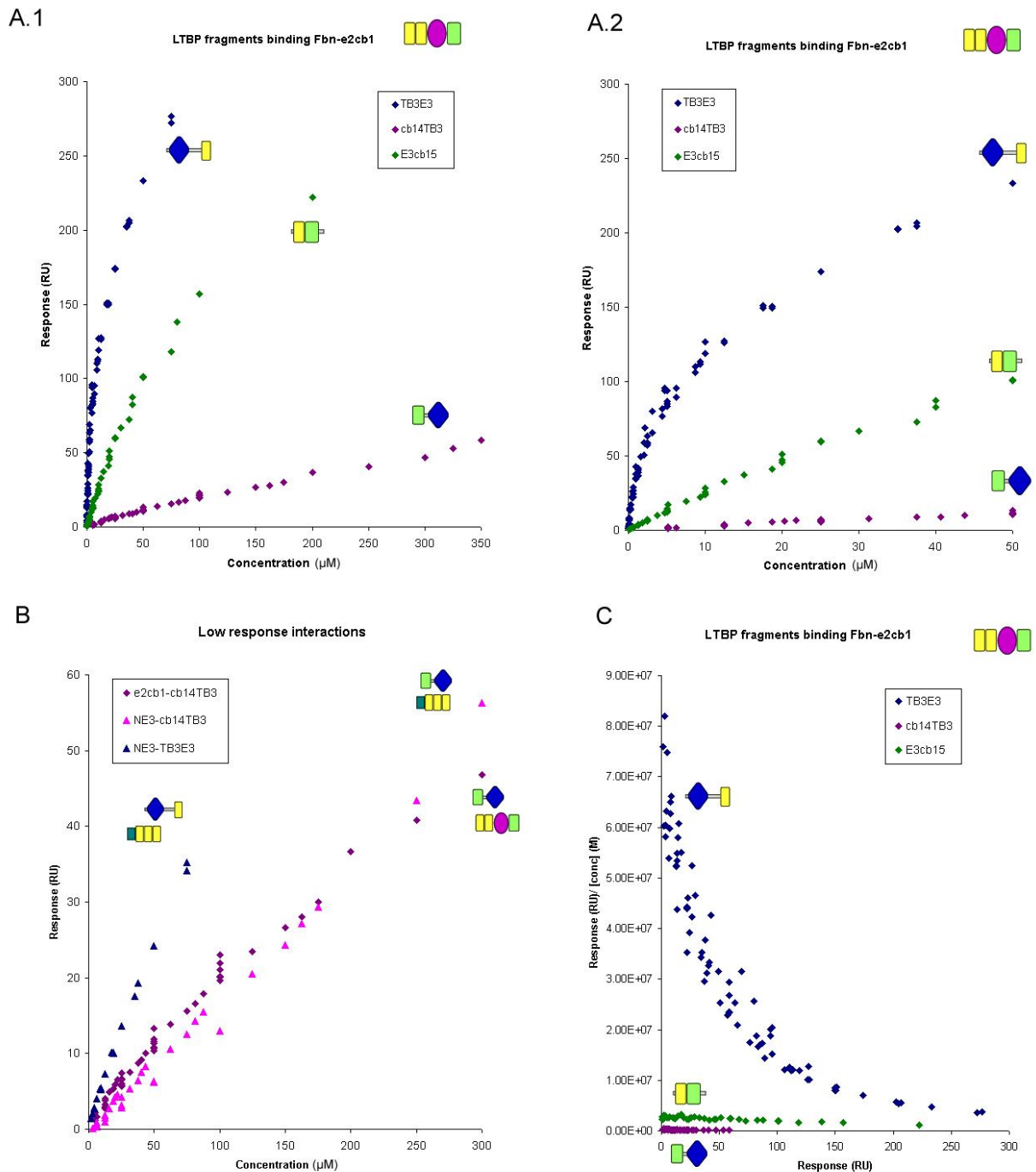


Figure 5.4 - Plots of SPR data

A) Plots of response vs concentration for the three LTBP1 fragments binding $\text{FBN}^{\text{e2cb1}}$ (not adjusted for molecular weight, which may affect relative response levels. For each fragment Molecular weights (Da) are: TB3E3= 19653.6, E3cb15= 11966.3, cb14TB3 = 13158.6). A.1 and A.2 show the same data, but A.2 focuses on the responses to lower concentrations of analyte ($\leq 50 \mu\text{M}$). For the LTBP1^{TB3E3} fragment a steep curve can be seen initially, which continues into a more shallow curve at higher concentrations. **B)** Plots of response vs concentration for weakly interacting fragments, specifically the interaction of LTBP1^{TB3E3} with FBN^{NE3} , and the interaction of LTBP1^{cb14TB3} with both $\text{FBN}^{\text{e2cb1}}$ and FBN^{NE3} , which show similar levels of binding for all three interactions. (Differences in mass and shape alone may account for the increased response with LTBP1^{TB3E3}.) **C)** Shows a Scatchard plot of (response / analyte concentration) vs (response), which is non linear for the LTBP1^{TB3E3} - $\text{FBN}^{\text{e2cb1}}$ interaction, meaning it does not conform to the Langmuir model (see Figure 5.5). Data shown are a combination of data collected from numerous single cycle experiments, performed with two or more preparations of each protein fragment on two separately prepared SPR chips.

expected for typical one to one Langmuir binding, as plots of response versus concentration (Figure 5.4 A and B) do not appear to reach saturation, despite the curve appearing to have quite a tight curvature at low concentrations. This non-Langmuir binding is confirmed in a Scatchard plot (Figure 5.4 D) where a non-linear relationship between $RU/[conc]$ and RU is observed.

In an ideal 1:1 interaction (Figure 5.5 A) a linear Scatchard plot should be obtained, but the non-linear relationship seen here could be due to a number of factors, such as heterogeneous exposure of FBN^{e2cb1} binding sites on the chip, where the CM5 matrix may partially occlude some binding sites in selected fibrillin molecules (Figure 5.5 B). Also it is possible that with two separate LTBP binding sites on FBN^{e2cb1} , some 2:1 interactions may occur between $LTBP1^{TB3E3}$ and FBN^{e2cb1} molecules at the chip surface (Figure 5.5C), which would further complicate the data. Estimates of K_d based on non-linear regression (Section 8.26) from the available data are shown in Table 5.1, but, due to the issues with $LTBP1^{TB3E3}$ binding, different values are obtained depending on the range of concentrations sampled, and these affinity estimates are intended solely for qualitative comparison.

Table 5.1 – Affinity estimates for different LTBP1 fragments Binding FBN^{e2cb1}

Affinity estimates from single cycle data, based on non-linear regression from RU – concentration values using a 1:1 binding model, which, as discussed in the text, is likely to be inaccurate. Values in red are also based on SPR data that is far from reaching saturation and therefore highly unreliable. Maximum analyte concentration used is given in each cell with K_d values below.

	$LTBP1^{TB3E3}$	$LTBP1^{E3cb15}$	$LTBP1^{cb14TB3}$
FBN^{E2cb1} Affinity	(75 μ M max) K_d 20.7 μM Rmax 355.2 χ^2 94.7	(200 μ M max) K_d 145 μM Rmax 379.5 χ^2 9.9	(350 μ M max) K_d 1.29 mM Rmax 264.2 χ^2 1.72

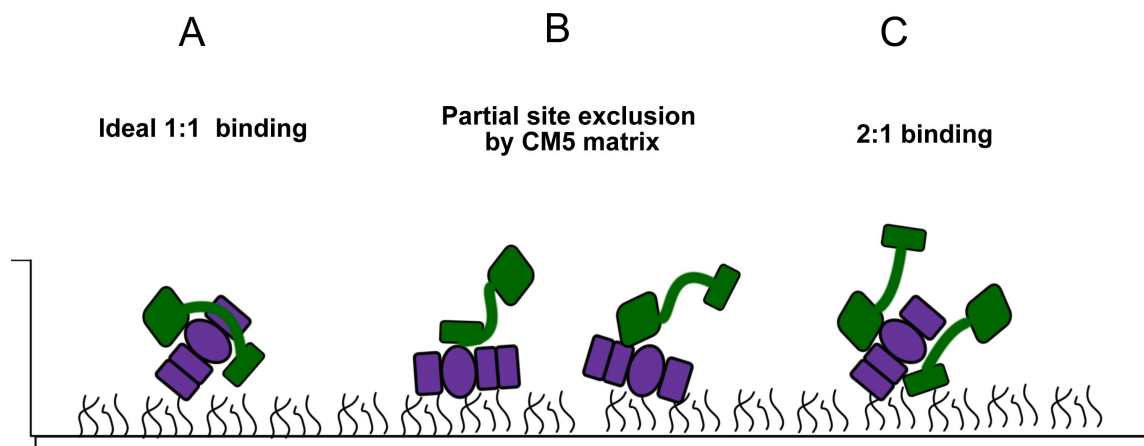


Figure 5.5 – Potential explanations for non-Langmuir binding between LTBP1^{1B3E3} and FBN^{e2cb1}

A) Ideal one to one binding at the chip surface, which would conform to the Langmuir model. **B)** Partial exclusion of binding sites of FBN^{e2cb2} by CM5 matrix, may result in a heterogeneous ligand surface. **C)** The presence of two independent interaction sites between LTBP1 and fibrillin1 may also allow competition between LTBP1 molecules at high concentration and formation of additional 2:1 complexes.

If precise determination of binding affinities were required, alternative chip coupling strategies, or other techniques such as isothermal titration calorimetry could be employed. However for localising the key binding sites between LTBP1 and fibrillin1, more precise affinities would not affect the interpretation of the SPR data.

The SPR data strongly suggest that the LTBP1-fibrillin1 interaction is not confined to a single domain of either protein. Both of the non-overlapping LTBP1^{cb14TB3} and LTBP1^{E3cb15} fragments interact with FBN^{e2cb1}, and so each must contain a distinct fibrillin binding site. As LTBP1^{E3cb15} only binds FBN^{e2cb1}, and not FBN^{NE3} (Figure 5.3), the binding of this region of LTBP1 must be dependent on the presence of the fibrillin hyb1-cbEGF1 domain pair, which is likely to contain the interaction site for LTBP1^{E3cb15}.

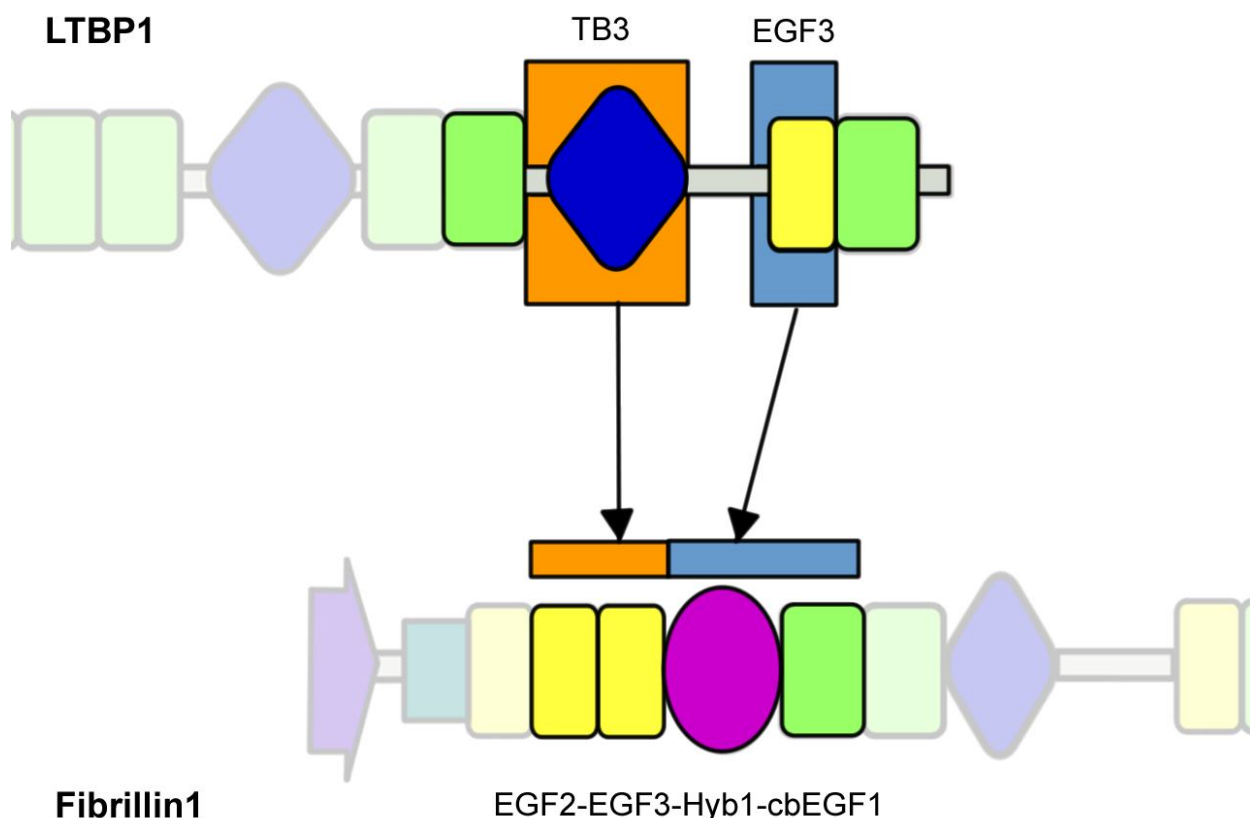
The FBN^{NE3} fragment only interacts with two LTBP1 constructs containing the TB3 domain (Figure 5.3), and shows markedly weaker binding of LTBP1^{TB3E3} construct than

FBN^{e2cb1}, but similar binding to LTBP1^{cb14TB3}. This suggests that the LTBP1^{cb14TB3} binding site might be in the fibrillin1 EGF2-EGF3 region, and could involve the TB3 domain of LTBP1.

The simplest explanation for the observed levels of binding, and for the specificity between the different constructs, is that two binding interactions occur between LTBP1 and fibrillin1 (Figure 5.6). One interaction occurs between the TB3 region of LTBP1 and the EGF2-3 region of fibrillin, and another between the EGF3 region of LTBP1 and the hyb1-cbEGF1 region of fibrillin. Both of these interactions occurring simultaneously could then explain the high level of binding seen between LTBP1^{TB3E3} and FBN^{e2cb1} (Figure 5.3).

Figure 5.6 – LTBP1 – Fibrillin1 interaction model suggested by SPR experiments.

Data suggest two separate binding sites, with LTBP1^{E3cb15} binding dependent on fibrillin hyb1-cbEGF1, and LTBP1^{cb14TB3} binding independent of FBN hyb1-cbEGF1. This figure summarises the simplest double binding site model, but more complex explanations cannot be ruled out, such as redundant binding sites for LTBP1^{cb14TB3} in the fibrillin hyb1-cbEGF1 and N-terminus-EGF1 regions.



However several other explanations cannot be ruled out by these results, such as a significant binding contribution by the poorly conserved TB3-EGF3 linker region¹, or the presence of additional binding sites introducing redundancy.

5.5 NMR titrations to define binding sites

In order to define more precisely which domains of LTBP1 interact with which domains of fibrillin1, and also to determine which residues are involved in the interaction, a number of HSQC titrations were carried out. The principle of these titrations is that when an ¹⁵N-labelled protein is bound in solution by an unlabelled protein, the amide nuclei at or near the binding site experience a change in their environment. This will then affect the chemical shift of the HSQC peaks from these amides. Depending on the rate of exchange between the bound or unbound state this will cause the peaks in an HSQC to shift (if the interchange between bound and unbound states is rapid), or to broaden / decrease in intensity, with new peaks, corresponding to the bound state, appearing elsewhere (if the interchange between states is slow). Intermediate exchange is also possible where both shifting and broadening is observed. The details of the methods and conditions used in these titrations are described in Section 8.34 and in the Figure legends. Particular care was taken to exclude any effects on the spectra due to changes of pH, both by checking and adjusting the pH before each titration point, and comparison of protein titration data with data from separate pH titrations.

For these experiments the same ¹⁵N-labelled and unlabelled LTBP1 and FBN^{e2cb1} fragments were used as produced and characterised earlier for HSQC peak assignments (chapters 3 and 4). Four titrations were carried out in total; two were carried out with ¹⁵N-labelled FBN^{e2cb1}, in order to define its binding sites for the non-

¹ Attempted purification of the isolated LTBP1 linker region for SPR studies was not successful with the FPLC and HPLC methods available.

overlapping LTBP1^{cb14TB3} and LTBP1^{E3cb15} proteins. The other two titrations were carried out with ¹⁵N labelled LTBP1 fragments and unlabelled FBN^{e2cb1}, one with ¹⁵N-labelled LTBP1^{cb14TB3} and the other with ¹⁵N labelled LTBP1^{E3cb15} to define the fibrillin binding sites on these non-overlapping LTBP1 fragments.

The HSQC titration of ¹⁵N-labelled FBN^{e2cb1} with LTBP1^{E3cb15} produced specific peak intensity changes (Figure 5.7), particularly for peaks assigned to the C-terminal β -hairpin of the hyb1 domain. When the peaks showing these intensity changes are mapped onto a preliminary structure of FBN^{e2cb1} (Figure 5.8), they are localised almost entirely on or around the C-terminal β -hairpin of the hyb1 domain, suggesting the binding site for LTBP1^{E3cb15} is in this region. The fact that peak intensity changes were seen with this titration rather than shifts is likely due to the protein-protein interaction being in a slow to intermediate exchange regime. This is often seen with interactions with K_d values below ~ 100 μ M [4], which is consistent with the estimates of affinity from the SPR data².

When ¹⁵N FBN^{e2cb1} is titrated with LTBP1^{cb14TB3} instead of LTBP1^{E3cb15} both peak intensity changes and shifts in peak position are seen (Figure 5.9). This combined peak shifting and intensity loss demonstrates an intermediate rate of exchange between bound and unbound states of the FBN^{e2cb1} and LTBP1^{cb14TB3} proteins. The specific intensity changes in the C-terminus of the hybrid domain seen in LTBP1^{E3cb15} titration are not seen in this titration, and instead intensity changes and shifting peaks mostly correspond to residues in the region of fibrillin-EGF3, and at the very start of the hyb1 domain. When mapped onto a preliminary structure of FBN^{e2cb1} (Figure 5.10), these residues appear to define a binding site that is focused around the main β

² K_d is not a clear predictor of fast or slow exchange however, as k_{on} and k_{off} may be variable.

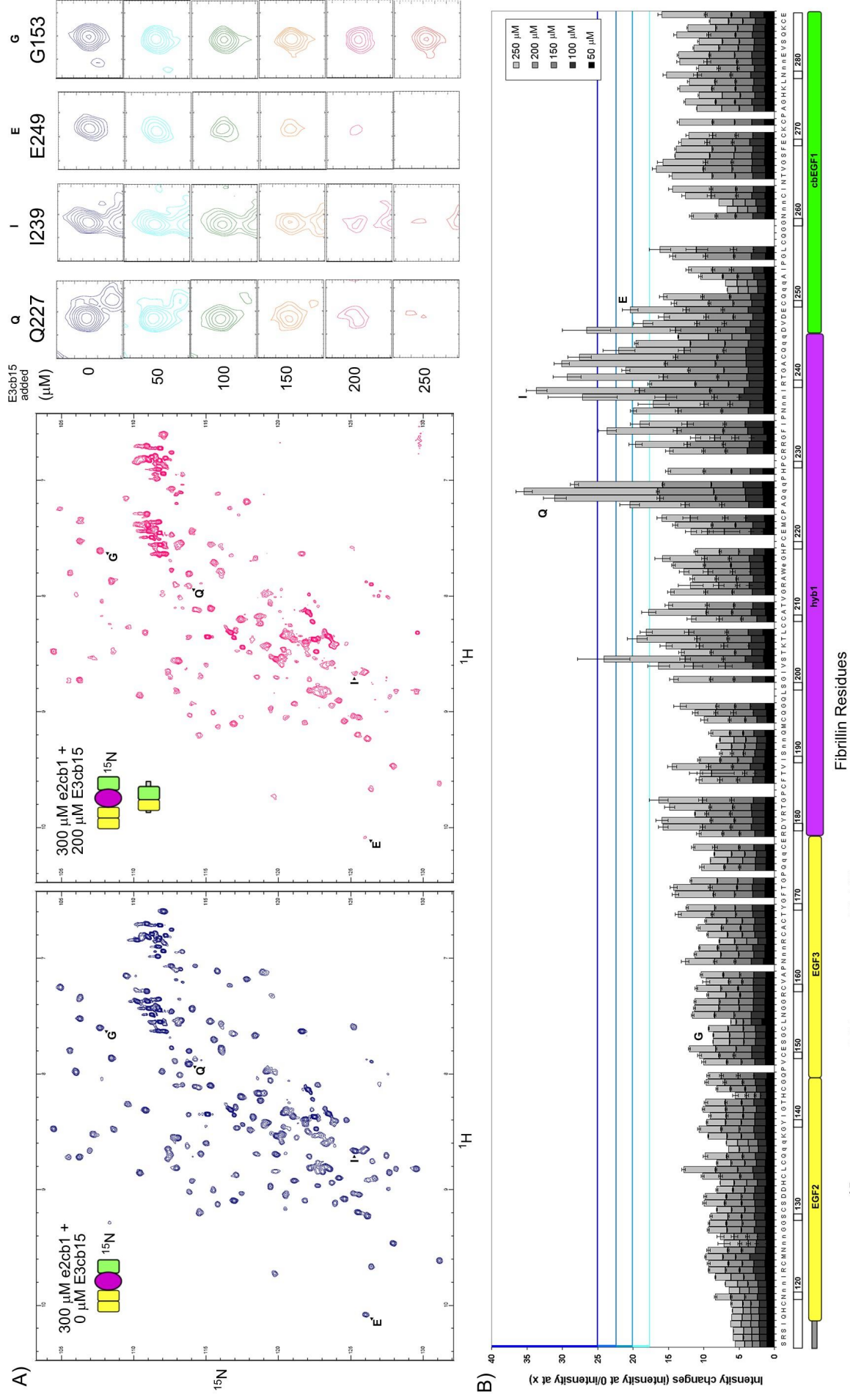


Figure 5.7- HSQC titration of ¹⁵N-labelled FBN^{e2cb1} with LTP1^{E3cb15}

300 μl of a 300 μM ¹⁵N FBN^{e2cb1} sample was used at pH 5.5 with 100 mM NaCl and 10 mM calcium, pH was checked and adjusted before each titration point, HSQC spectra were collected with 256 scans (see Section 8.34). **A)** Two example HSQC spectra from the titration shown side by side, and with identical contouring, displaying specific losses in peak intensity. Four letters (E, I, Q, and G) highlight some example peaks, and to the right of the HSQC spectra the intensity changes of these peaks are followed more closely. **B)** Stacked bar chart of peak intensity change; intensity change measured as "intensity at 0 μM ligand" / "intensity at x μM ligand". Error bars represent standard error determined by intensities of measured peaks relative to background noise (see Section 8.34 for methodology). Letters G, Q, I and E highlight the example peaks in panel A. Blue lines in the bar chart also represent arbitrary cut off's used to generate Figure 5.8. On the x axis lower case letters represent side chain amides (q= glutamine side chain NH, n= asparagine side chain NH, e= tryptophan side chain NH1 indole)

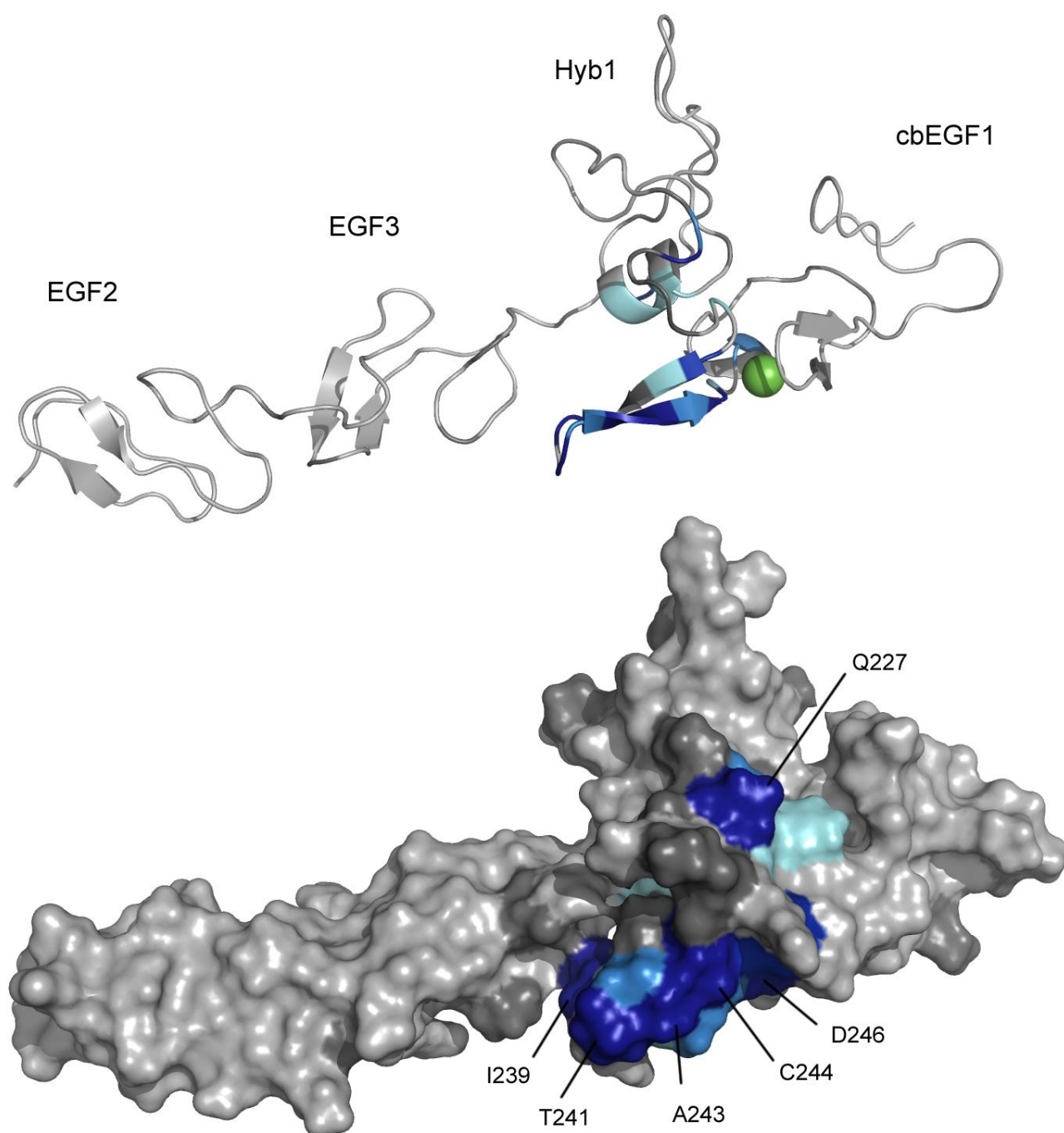


Figure 5.8 - LTBP1^{E3cb15} - ¹⁵N-FBN^{e2cb1} HSQC titration data mapped onto a preliminary FBN^{e2cb1} structure

Residues with significant intensity changes highlighted on a preliminary structure of FBN^{e2cb1} (straight structure 40-3-3 determined in Chapter 3). At the top is a cartoon representation of protein topology, and a surface representation of FBN^{e2cb1} at the bottom, both are coloured according to the colour scale to the right of Figure 5.7 B. Residues with changes below the threshold are coloured white, unassigned residues and prolines are coloured dark grey, and calcium is shown as a green sphere.

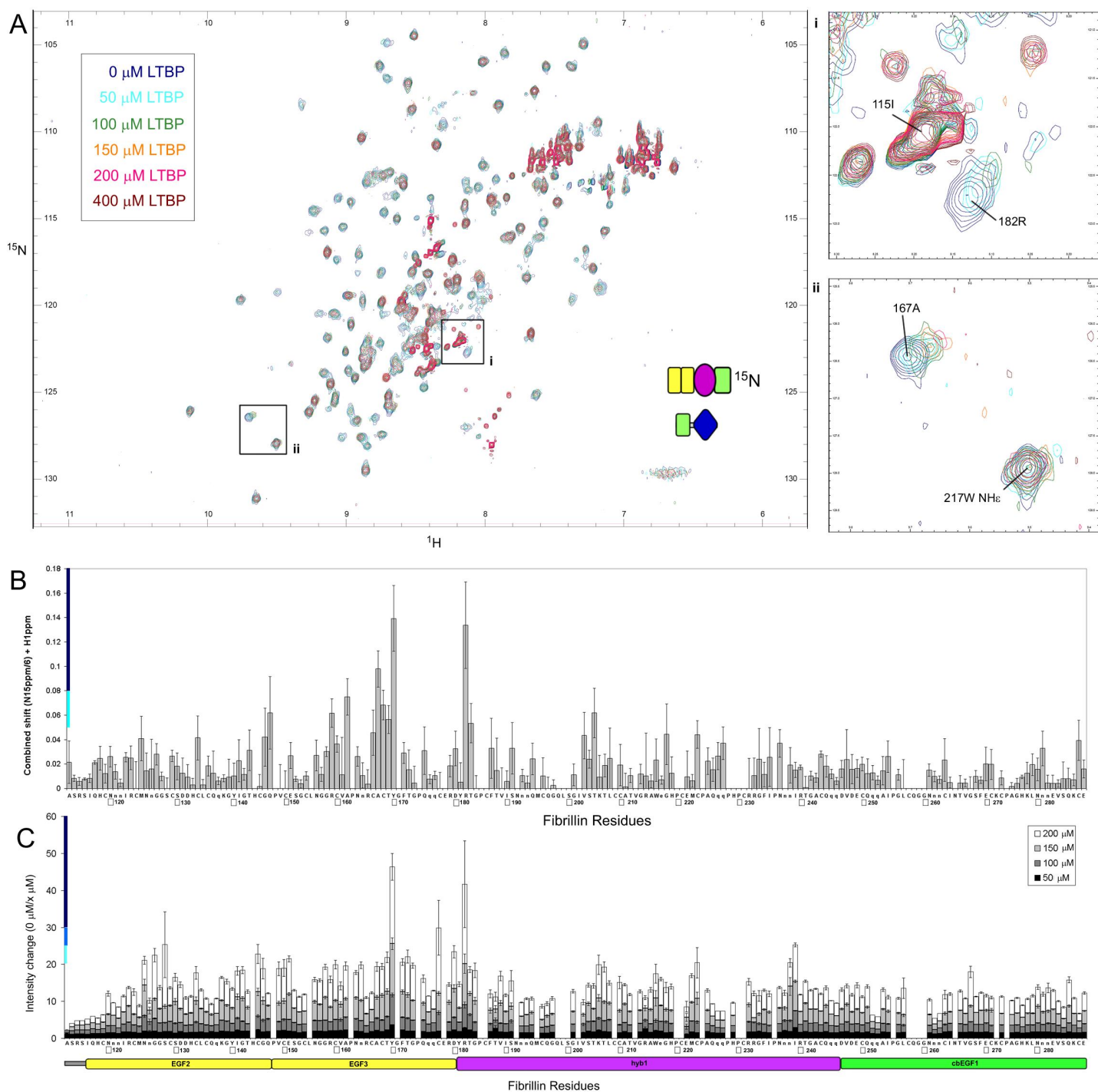
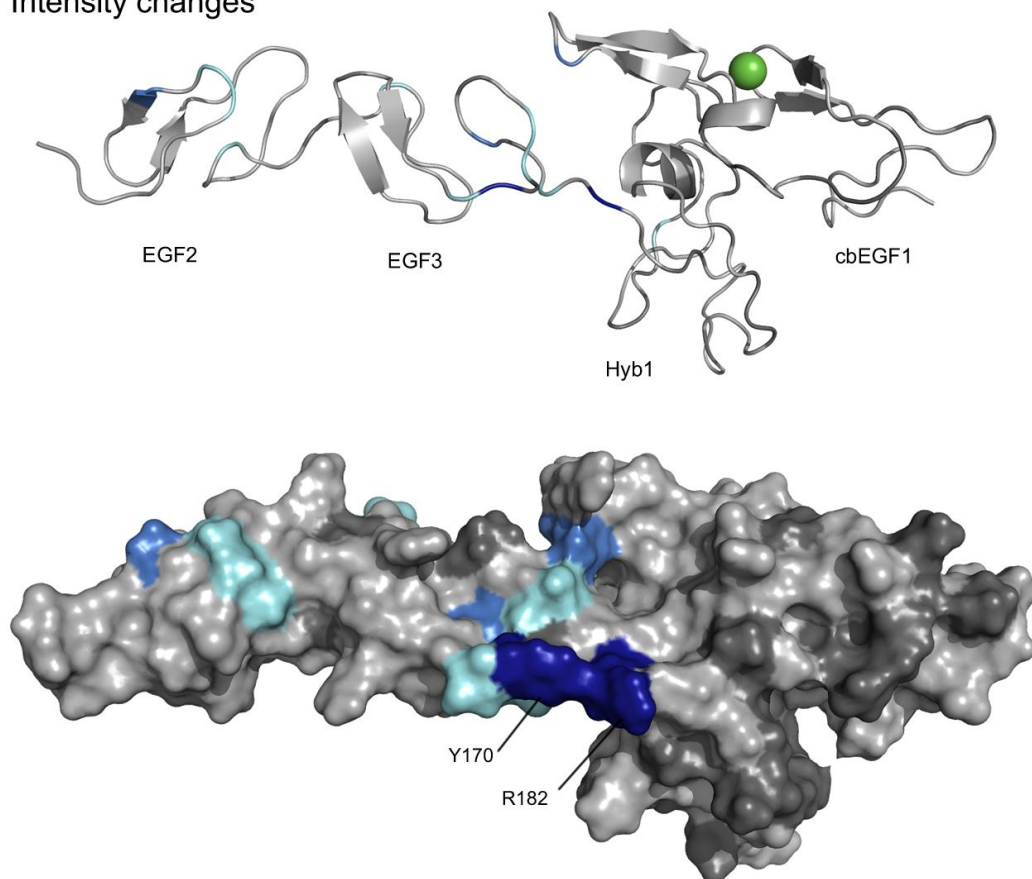


Figure 5.9 - HSQC titration of ^{15}N -labelled $\text{FBN}^{\text{e2cb1}}$ with $\text{LTBP1}^{\text{cb14TB3}}$

290 μ l of a 300 μ M ^{15}N $\text{FBN}^{\text{e2cb1}}$ sample was used at pH 5.3 with 15 mM calcium, (see Section 8.26 and 8.34). **A**) Overlaid HSQC spectra coloured according to amount of $\text{LTBP1}^{\text{cb14TB3}}$ added. Two regions containing shifting and static peaks are highlighted as i and ii, and shown in detail on the right. **B**) Bar chart of the combined shift of each peak as a function of sequence, up to the 200 μ M titration point, (after which some of the peaks become indistinguishable from noise)(Section 8.34). **C**) Stacked bar chart of peak intensity change measured as "intensity at 0 μ M ligand"/ "intensity at x μ M ligand". Blue lines represent cut offs used to generate Figure 5.10.

A Intensity changes



B Peak shifts

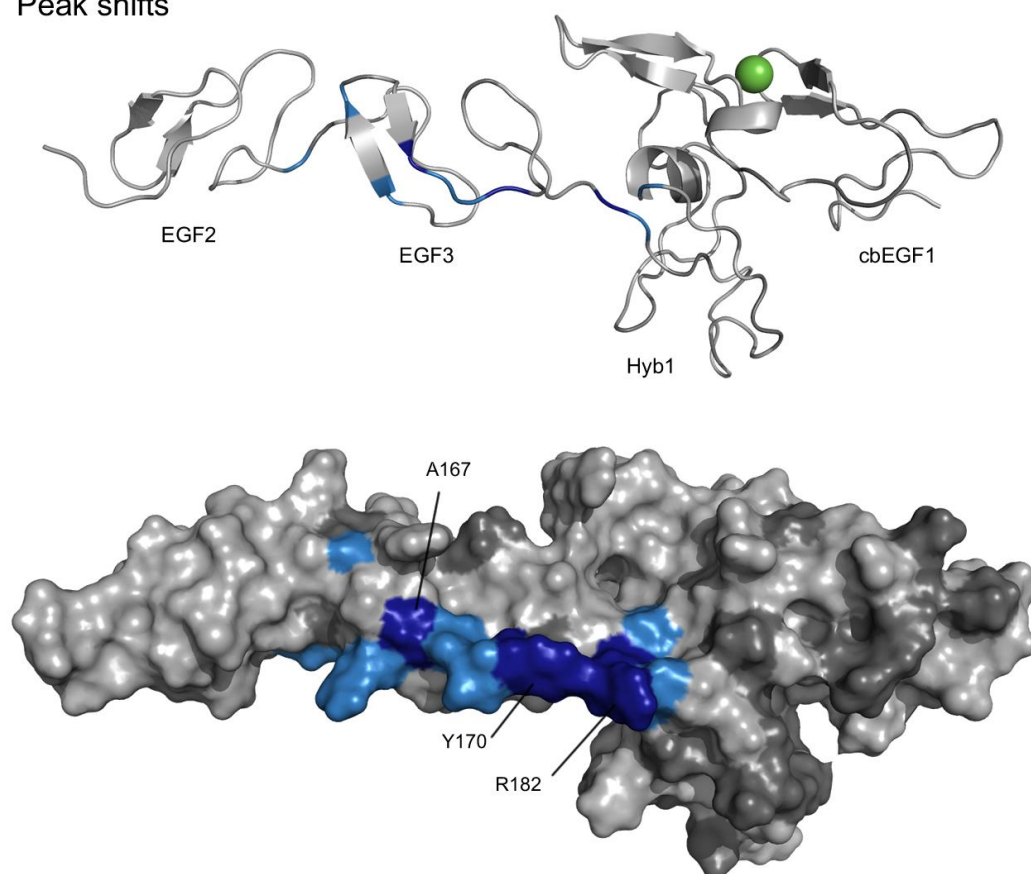


Figure 5.10 - LTBP1^{cb14TB3}-¹⁵N-FBN^{e2cb1} HSQC titration data mapped onto FBN^{e2cb1} preliminary structure

Residues with significant intensity changes and shifts highlighted on a preliminary structure of FBN^{e2cb1} (straight structure 40-3-3 determined in Chapter 3). **A)** Residues with peaks exhibiting significant intensity change, are coloured according to the colour scale to the right of Figure 5.9 C. **B)** Residues with peaks exhibiting significant shifts, are coloured according to the colour scale to the right of Figure 5.9 B. In both panels unassigned residues and prolines are coloured dark grey.

hairpin of fibrillin-EGF3, but which also includes residues at the start of the hyb1 domain and C-terminal region of EGF2.

When the reverse HSQC-titration is carried out with unlabelled FBN^{e2cb1} and ¹⁵N-labelled LTBP1^{cb14TB3}, again peaks were seen to both broaden and shift as the titration progressed (Figure 5.11), confirming that these proteins have an intermediate rate of exchange between bound and unbound states in solution. From this ¹⁵N-LTBP1^{cb14TB3} titration it can be seen that the most affected peaks all correspond to residues in the LTBP1-TB3 domain. However not all peaks that exhibit the highest intensity change also exhibit significant shifts. This can occur with intermediate exchange, where those peaks that experience the greatest change in environment and chemical shift may be seen just to broaden rather than shift.

When mapped onto the homology model of the TB3 domain (Figure 5.12A), the most significant intensity changes are focused around the two α helices, especially around the core helix and tyrosine 1548. The pattern of shifting residues also reflects a similar binding patch on the same face of the TB3 domain (Figure 5.12B).

There are also two residues on the opposite face of the TB domain that exhibit a strong intensity change (glutamate 1529 and arginine 1539). However these appear isolated, and not part of a large patch of affected residues. It is possible that these residues, and some of the other affected residues in the core of the protein, may represent long-range structural changes in the TB3 domain on binding to fibrillin.

Finally titration of ¹⁵N-labelled LTBP1^{E3cb15} with unlabelled FBN^{e2cb1} revealed a number of peaks with specific losses in intensity with increasing levels of added fibrillin (Figure 5.13). As with the reverse titration, no apparent shifts were seen, due to slow or intermediate exchange between bound and unbound states in this experiment. Those peaks with the greatest loss in intensity all correspond to residues localised to the

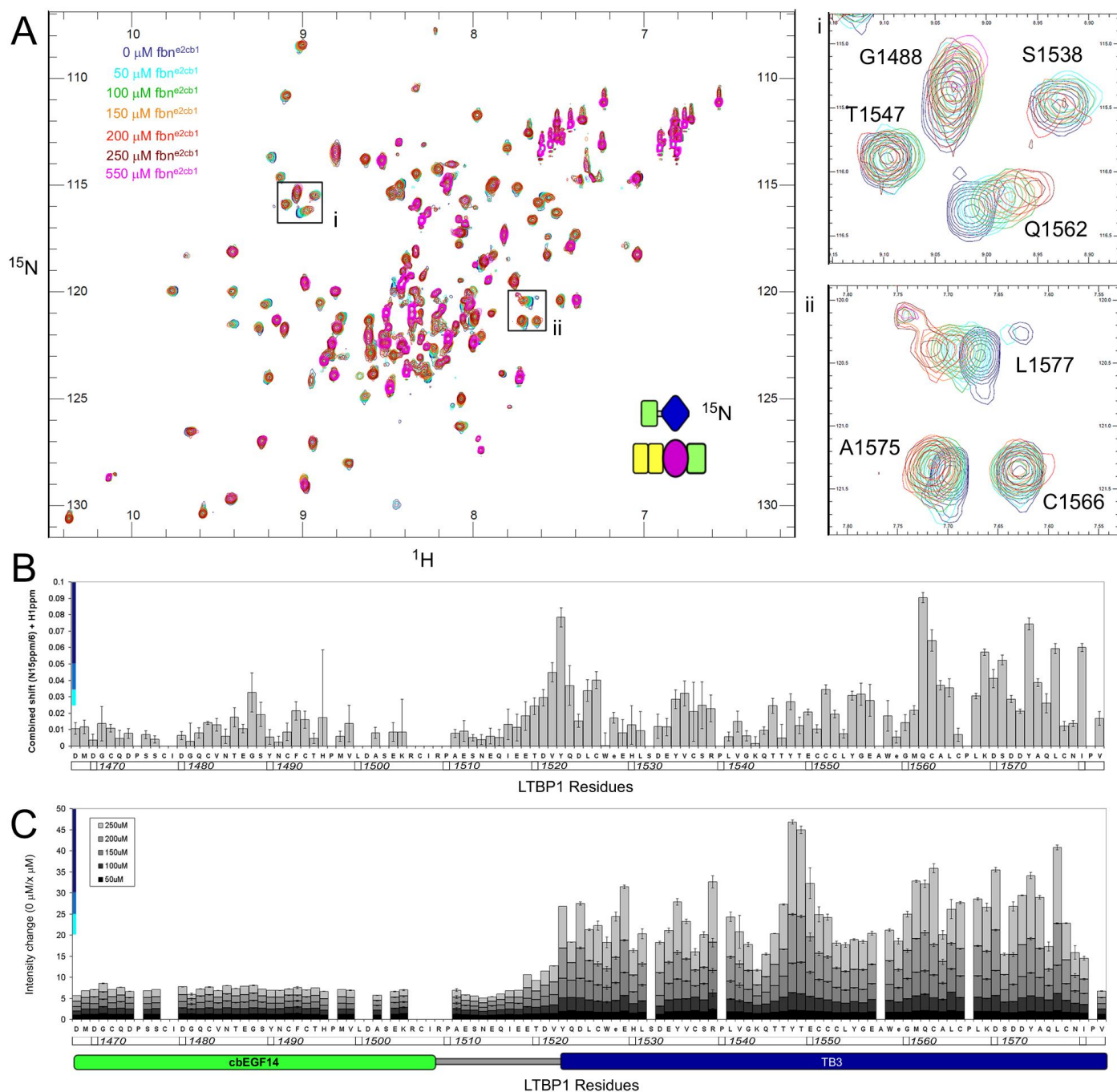
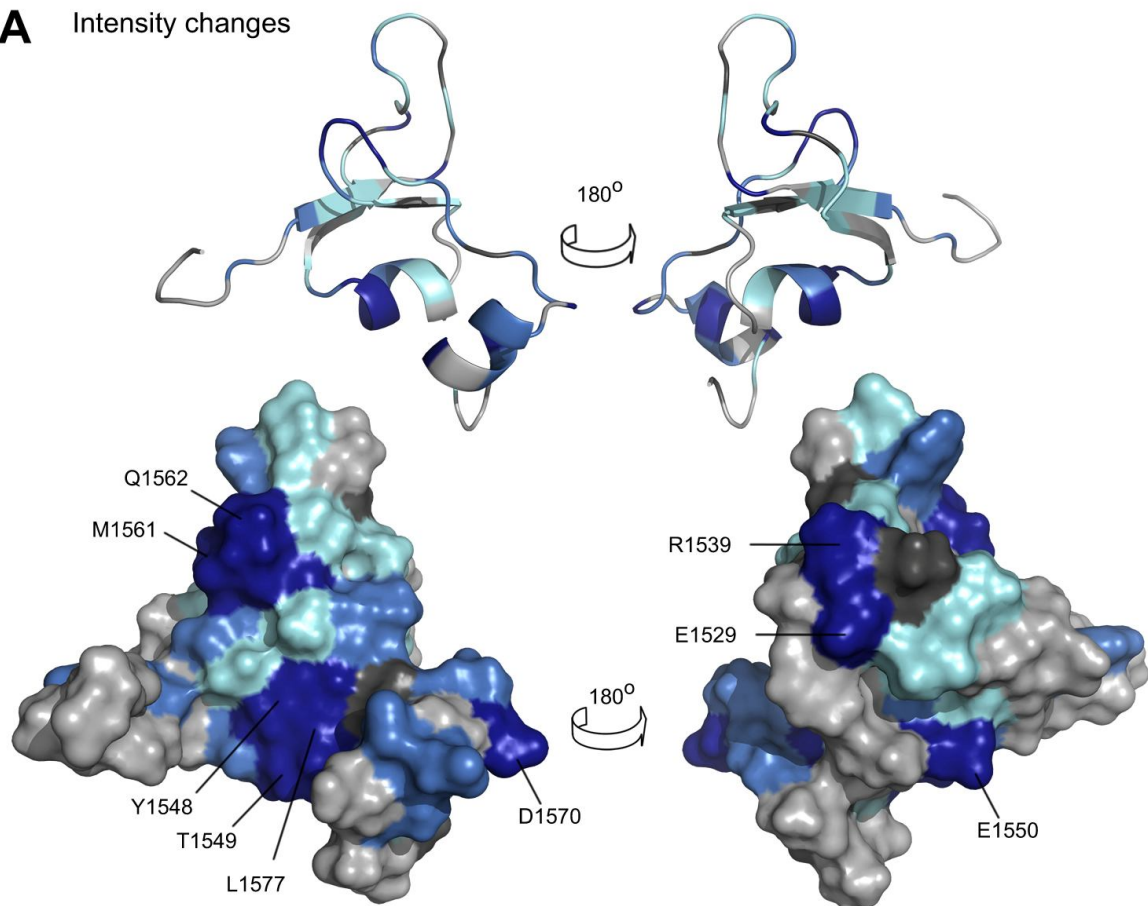


Figure 5.11 - HSQC titration of ^{15}N LTBP1^{cb14TB3} with FBN^{e2cb1}

300 μl of a 300 μM ^{15}N -LTBP1^{cb14TB3} sample was used at pH 5.9 with 20 mM calcium, (see Section 8.27 and 8.34) **A**) Overlaid HSQC spectra coloured according to amount of FBN^{e2cb1} added, two regions containing shifting and static peaks are highlighted. **B**) Bar chart of the combined shift of each peak up to the 250 μM titration point, (after which some of the peaks become indistinguishable from noise). (Section 8.34). **C**) Stacked bar chart of peak intensity change as a function of sequence. Intensity change measured as "intensity at 0 μM ligand"/ "intensity at $x \mu\text{M}$ ligand". Error bars represent standard error determined by intensities of peaks measured relative to background noise (see Section 8.34 for methodology). Blue lines represent arbitrary cut offs used to generate Figure 5.12.

A Intensity changes



B Peak shifts

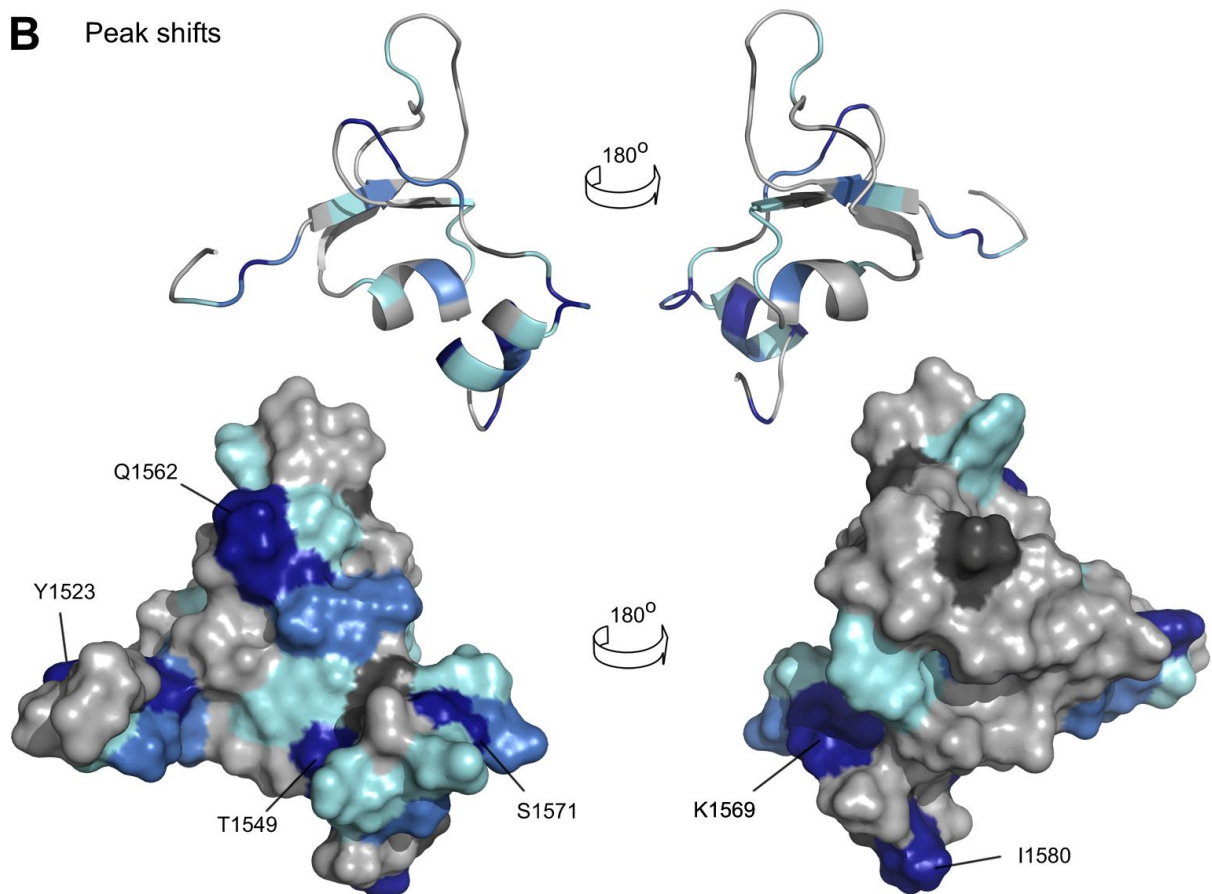


Figure 5.12 - FBNe^{2cb1} titration data mapped onto LTBP1 TB3 homology model

A) Residues with peaks demonstrating a significant loss in intensity, when titrated with FBNe^{2cb1}, mapped onto the homology model of LTBP1 TB3 (from Chapter 4). The cartoon representations at the top of A) demonstrate the topology of the domain, and are oriented identically to the surface representations below. Residues are coloured according to the scale displayed beside Figure 5.11C). **B)** Residues with peaks demonstrating a significant combined chemical shift, mapped onto the homology model of LTBP1 TB3. Colour scale displayed beside Figure 5.11B). In both panels residues with changes below the thresholds are coloured white, and residues coloured dark grey were not assignable in titration spectra.

LTBP1-EGF3 domain (Figure 5.13B), and also include several of the conserved yet flexible residues near the N-terminus of the construct (Section 4.8 fig 4.15). The peaks assigned to flexible residues at the C-terminus of LTBP on the other hand display markedly little change in intensity, while peaks from cbEGF15 residues show some intensity loss, but significantly less than seen with peaks assigned to EGF3 residues.

In Figure 5.14 the residues with peak intensity loss are mapped onto the homology model of LTBP1^{E3cb15}. The small size of the EGF3 domain makes it difficult to discern an obvious binding surface, as residues are often accessible from both sides of the domain. However the clear involvement of the flexible N-terminus, the main β -hairpin, and also threonine 1656 in the C-terminal loop of EGF3, suggest the interaction affects most elements of the domain. It could be speculated that this would be best accommodated by the interaction of fibrillin with the central cavity of EGF3, where a cbEGF would bind calcium.

These HSQC titrations have defined the four binding sites involved in two LTBP1 – fibrillin1 interactions. One interaction seems to involve the LTBP1-TB3 domain and the fibrillin1-EGF3 region, with potential contribution from adjacent EGF3 and hyb1 domains. The other interaction appears to be between the LTBP1-EGF3 domain and the C-terminal region of the fibrillin1-hyb1 domain. These binding sites confirm the model suggested by the SPR data in which LTBP1 TB3 was proposed to interact with fibrillin EGF2-3, and LTBP EGF3 interact with fibrillin hyb1-cbEGF1. However, the titrations provide more detailed information on the domains involved, and implicate specific residues that may define the protein binding sites.

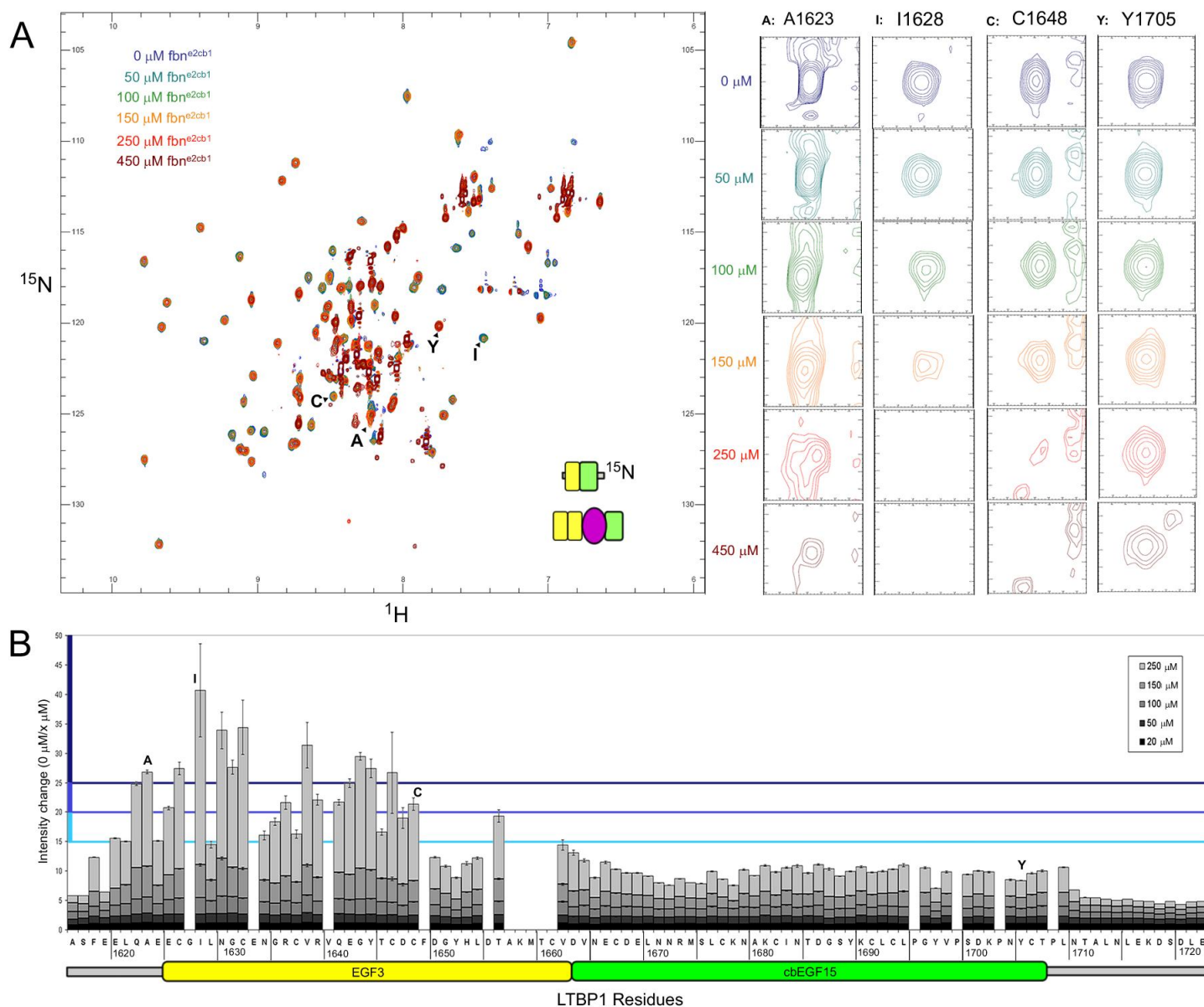


Figure 5.13- HSQC titration of ^{15}N LTBP1^{E3cb15} with FBN^{e2cb1}

300 μl of a 300 μM ^{15}N LTBP1^{E3cb15} sample was used at pH 5.3, with 100 mM NaCl and 10 mM calcium, (Section 8.34). **A**) Overlaid titration HSQC spectra. Letters A, I, C and Y highlight some example peaks, and on the right the intensity changes of these peaks are shown in detail. **B**) Stacked bar chart of peak intensity change measured as "intensity at 0 μM ligand"/ "intensity at x μM ligand". Error bars represent standard error based on the intensities of measured peaks relative to background noise (see Section 8.34 for methodology). Letters A, I, C and Y highlight the example peaks from A. Blue lines represent arbitrary cut offs used to generate Figure 5.14.

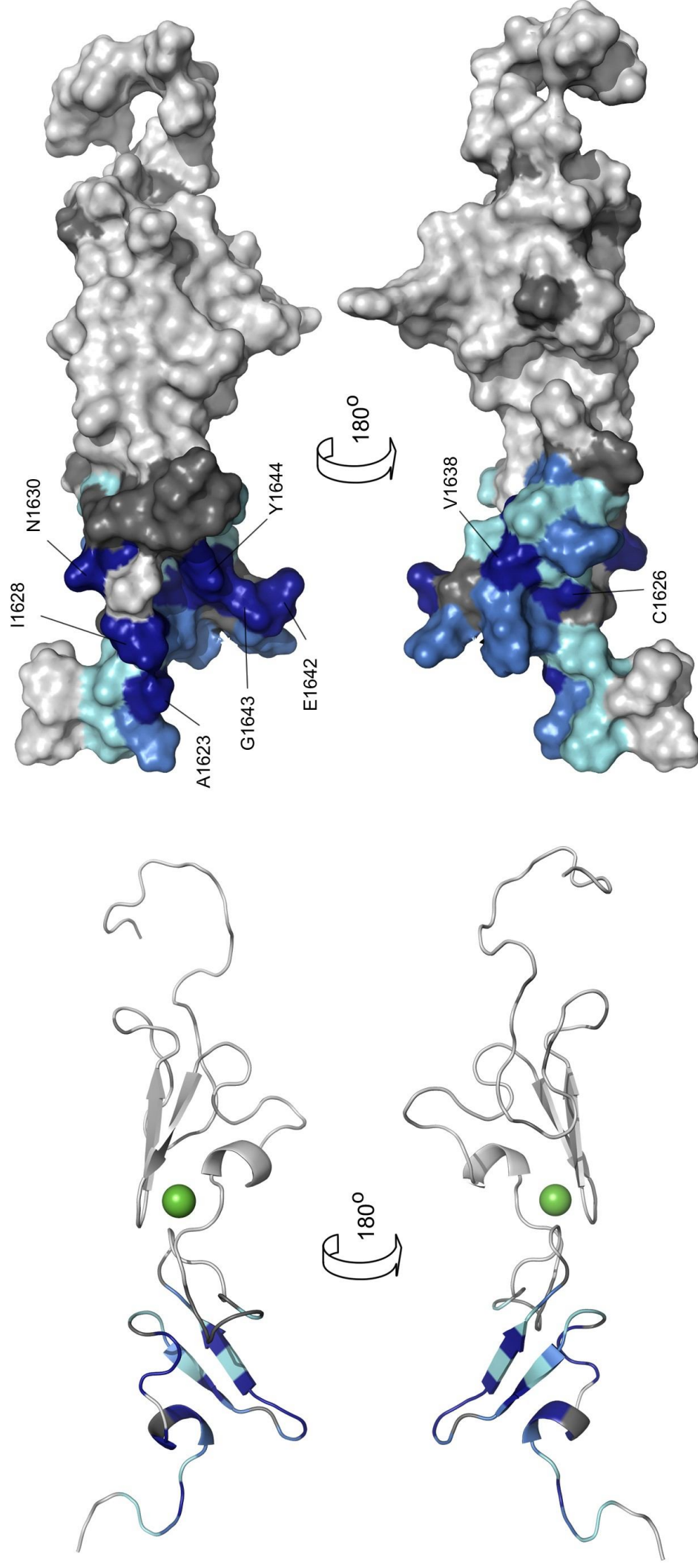


Figure 5.14 - FBN^{E2cb1} titration data mapped onto LTBP1^{E3cb15} homology model

Residues with peaks exhibiting significant intensity changes in the titration with fibrillin are highlighted on a homology model of LTBP1^{E3cb15} (from Chapter 4). A cartoon topology representation is shown on the left, and a surface representation on the right, with an identical orientation to the cartoon representation. Residues and are coloured according to the colour scale to the right of Figure 5.13 B. Residues with changes below the thresholds are coloured white, and unassigned residues and prolines are coloured dark grey.

5.6- Docking simulations between Fibrillin1 and LTBP1 fragments

Using the data obtained from the above HSQC titrations, and homology models of the relevant LTBP1 and fibrillin1 domains³, it is possible to model potential structures of the LTBP1 - fibrillin1 complex. To do this, docking simulations were carried out using the HADDOCK programme [5] (Section 8.36), with which the binding of the LTBP1-TB3 domain to FBN^{e2cb1}, and LTBP1-EGF3 domain to FBN hyb1-cbEGF1 were separately investigated. In both docking simulations multiple runs were performed using different cut offs to select the input restraint residues that define the binding sites. Various flexible regions were also defined in the simulations based on the heteronuclear-NOE data available. These inputs and the resulting HADDOCK scores for the best structures of each run are summarised in table 5.2 and 5.3. The resulting HADDOCK structures were also classified as general “models” (eg model A, model B etc), based on the relative orientation of the interacting proteins, and presence of similar putative salt bridges and other features.

Docking of the LTBP1 TB3 domain to fibrillin was initially carried out with the FBN^{e2cb1} homology model, and to account for the uncertainty in the fibrillin EGF3-hyb1 interface in this model, the residues connecting the two domains were included as flexible, so that alternative conformations could be explored. The highest scoring HADDOCK results with LTBP1 TB3 and FBN^{e2cb1} consistently fitted a single model of binding (Figure 5.15), in which the core region of LTBP1 TB3, including the core tryptophan (W1559) and conserved TYT motif at the start of the core α -helix appears to interact with the main β -hairpin region of fibrillin1 EGF3 domain. In this binding model some interaction is also seen between LTBP1 TB3 and the start of the hyb1 domain, and a putative salt bridge can be seen between aspartates 1572 and

³ The fibrillin1 structures investigated in chapter 3 were not available when this docking analysis was carried out.

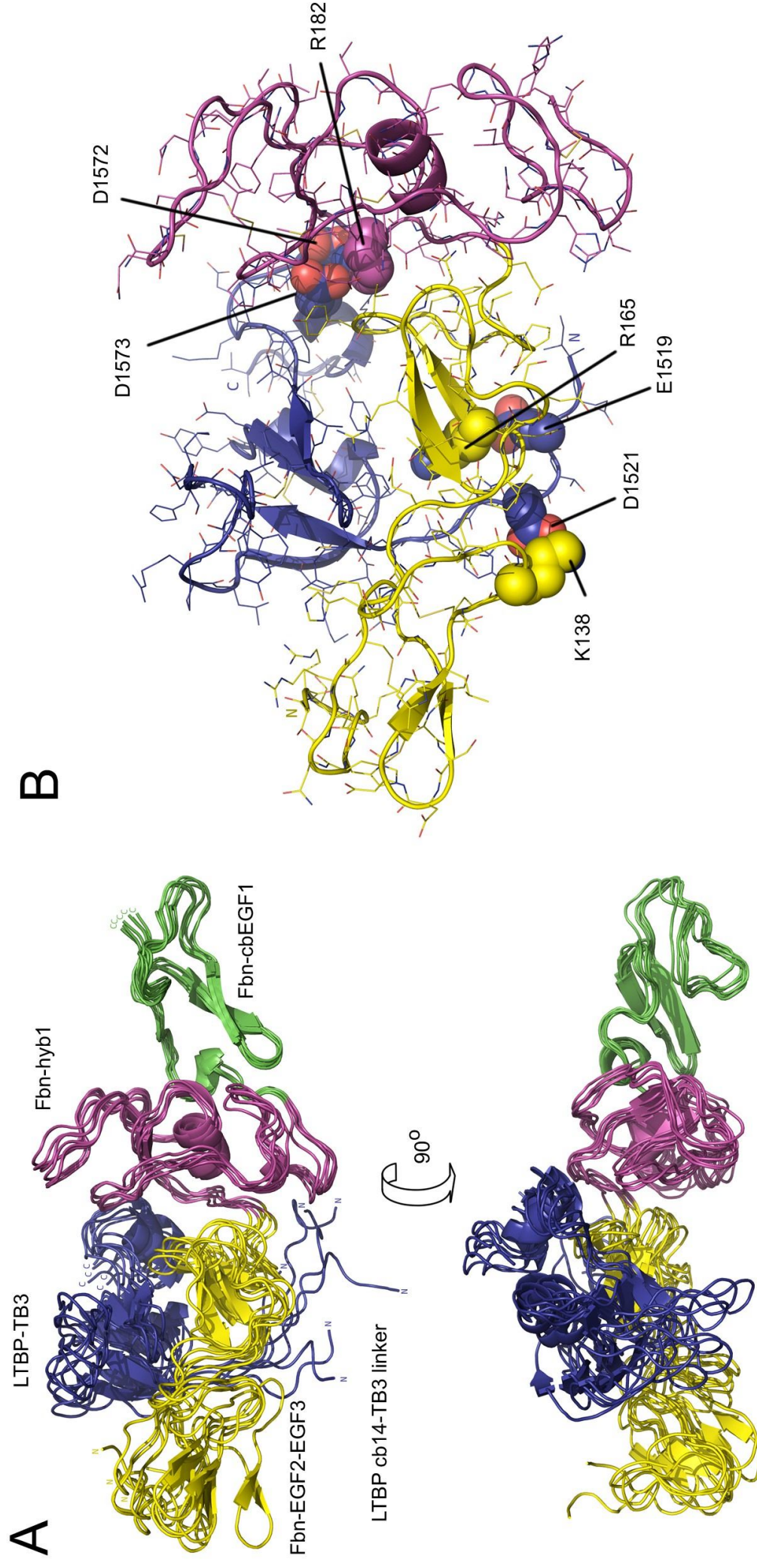


Figure 5.15 - *In silico* docking of FBN1^{ezcb1} and the LTBP1 TB3 domain

A) Overlay of top scoring results from several HADDOCK simulations (Table 5.2). The top scoring results from the following groups of structures were used: run1-cluster1, run2-cluster1, run2-cluster9, run4-cluster3, run5-cluster1, and run5-cluster3. Fibrillin EGF2-3 domains are coloured yellow, hyb1 magenta, and cbEGF1 is coloured green. LTBP1 TB3 is coloured dark blue. Structures superimposed by aligning them to a single FBN^{ezcb1} structure with the pymol alignment tool **B)** Detailed look at a single structure (run2 cluster1). All residues of residues involved in putative salt bridges highlighted as spheres.

1573, on the second α -helix of LTBP1 TB3, and arginine 182 at the very start of hyb1. Putative salt bridges can also be seen between acidic residues in the short flexible linker at the N-terminus of LTBP1 TB3, and basic residues in fibrillin EGF2 and 3. For example, aspartate 1521 in LTBP1 appears to form a salt bridge with lysine 138 in fibrillin EGF2 (Figure 5.15B). However the position of this linker region is the most variable element of these structures (Figure 5.15 A) and the same salt bridges involving the linker region are not seen consistently in all calculations.

Table 5.2 - HADDOCK run summary and top scoring results for LTBP TB3 and FBN^{e2cb1} interaction

Residues chosen from LTBP1 and fibrillin1 as ambiguous interaction restraints selected based on cut offs determined from 0-150 μM ^{15}N FBN^{e2cb1} and 0-250 μM ^{15}N LTBP1^{cb14TB3} titration data (see Section 8.35). Flexible residues were selected on the basis of heteronuclear-NOE data and in regions of structural uncertainty.

Run	Fibrillin restraints	LTBP restraints	Fibrillin fully flexible residues	LTBP fully flexible residues	Other notes	Top scores (model supported) (C=cluster)
1	Top 11 shifts	Top 34 shifts	113-119 177-186 238-241	62-77 92-96 130-133	Histidines modelled as N ϵ protonated	C1- -161.0 +/- 7.8 (A) C6 -123.5 +/- 32.9 (-A)
2	Top 11 shifts and 4 intensity changes (12 restraints)	Top 32 intensity changes	113-119 177-186 238-241	62-77 92-96 130-133	Histidines modelled as N ϵ protonated	C1 -158.5 +/- 10.2 (A) C9 -114.3 +/- 21.3 (A)
3	Top 5 intensity changes	Top 11 intensity changes	113-119 177-186 238-241	68-77 92-96 130-133	Histidines modelled as N ϵ protonated	C4 -103.9 +/- 7.8 (B) C2 -97.6 +/- 8.1 (C)
4	Top 11 shifts and 5 intensity restraints (13 total)	Top 32 intensity changes	113-119 177-186 238-241	68-77 92-96 130-133	Histidines modelled as N ϵ protonated	C1 -147.2 +/- 14.4 (A) C3 -122.7 +/- 13.3 (A)
5	Top 6 shifts	Top 8 shifts	113-119 177-186 238-241	68-77 92-96 130-133	Histidines modelled as N ϵ protonated	C1 -117.0 +/- 5.4 (-A) C3 -107.4 +/- 5.8 (A)

For the interaction of LTBP1-EGF3 and fibrillin, HADDOCK docking was first carried out with a homology model of the fibrillin hyb1 and cbEGF1 domain pair alone, to prevent any interference from the positioning of the fibrillin EGF2-EGF3 pair. The highest scoring results from the FBN^{e2cb1} and LTBP1-EGF3 docking are shown overlaid in Figure 5.16. With these calculations there were two different high scoring

interaction models produced by the HADDOCK calculations (referred to here as model A and model B). The model that scored highest depended and the thresholds for selecting the interaction restraints, and the level of flexibility introduced into the protein structures (table 5.3).

Table 5.3 - HADDOCK run summary and top scoring results for LTBP1 EGF3 and fibrillin Hyb1-cbEGF1 interaction

Residues chosen from LTBP1 and fibrillin1 as ambiguous interaction restraints selected based on cut offs determined from 0-250 μM ^{15}N FBN^{e2cb1} and 0-250 μM ^{15}N LTBP1^{E3cb15} titration data (Section 8.35) Flexible residues were selected on the basis of heteronuclear-NOE data and in regions of structural uncertainty

Run	Fibrillin restraints	LTBP restraints	Fibrillin fully flexible residues	LTBP fully flexible residues	Other notes	Top scores (model supported) (C=cluster)
2	Top 17 intensity changes	Top 29 intensity changes	none	none	Histidines modelled as +ively charged	C2 -132.3 +/- 5.7 (A) C3 -93.7 +/- 2.9 (B)
3	Top 17 intensity changes	Top 29 intensity changes	73-100	1-11 23-31 40-46	Histidines modelled as +ively charged	C3 -131.1 +/- 12.8 (B) C1 -128.4 +/- 10.4 (A)
6	Top 6 intensity changes	Top 21 intensity changes	92-95	1-12	Histidines modelled as +ively charged	C3 -113.2 +/- 7.0 (A) C1 -101.2 +/- 2.8 (B)
7	Top 15 intensity changes	Top 21 intensity changes	92-95	1-12 24-31	Histidine modelled as N ϵ protonated	C2 -113.0 +/- 18.3 (B) C3 -105.8 +/- 15.1 (A)
8	Top 15 intensity changes	Top 28 intensity changes	92-95	1-12 13-17 24-31	Histidine modelled as N ϵ protonated	C1 -114.9 +/- 11.6 (B) C8 -108.3 +/- 13.7 (A)
9	Top 15 intensity changes	Top 28 intensity changes	92-95	Entire domain modelled as fully flexible	Histidine modelled as N ϵ protonated	C1 -123.4 +/- 10.9 (B) C4 -92.9 +/- 29.4 (B)
10	Top 15 intensity changes	Top 28 intensity changes	none	None	Histidine modelled as N ϵ protonated	C6 -123.4 +/- 4.8 (A) C2 -104.0 +/- 4.4 (B)

Interaction model A and model B have several similarities, and in both models fibrillin hyb1 binds within the core of the LTBP1 EGF3 domain, where it is sandwiched between the main β -hairpin, C-terminal loop, and N-terminus of the LTBP1-EGF3 domain. Similar binding residues are also used in both models, but in model A only the C-terminal loop of fibrillin1 hyb1 is inserted into the core of LTBP1 EGF3, and the main β -hairpin of the LTBP1 EGF3 domain is pointed in towards the core α -helix of

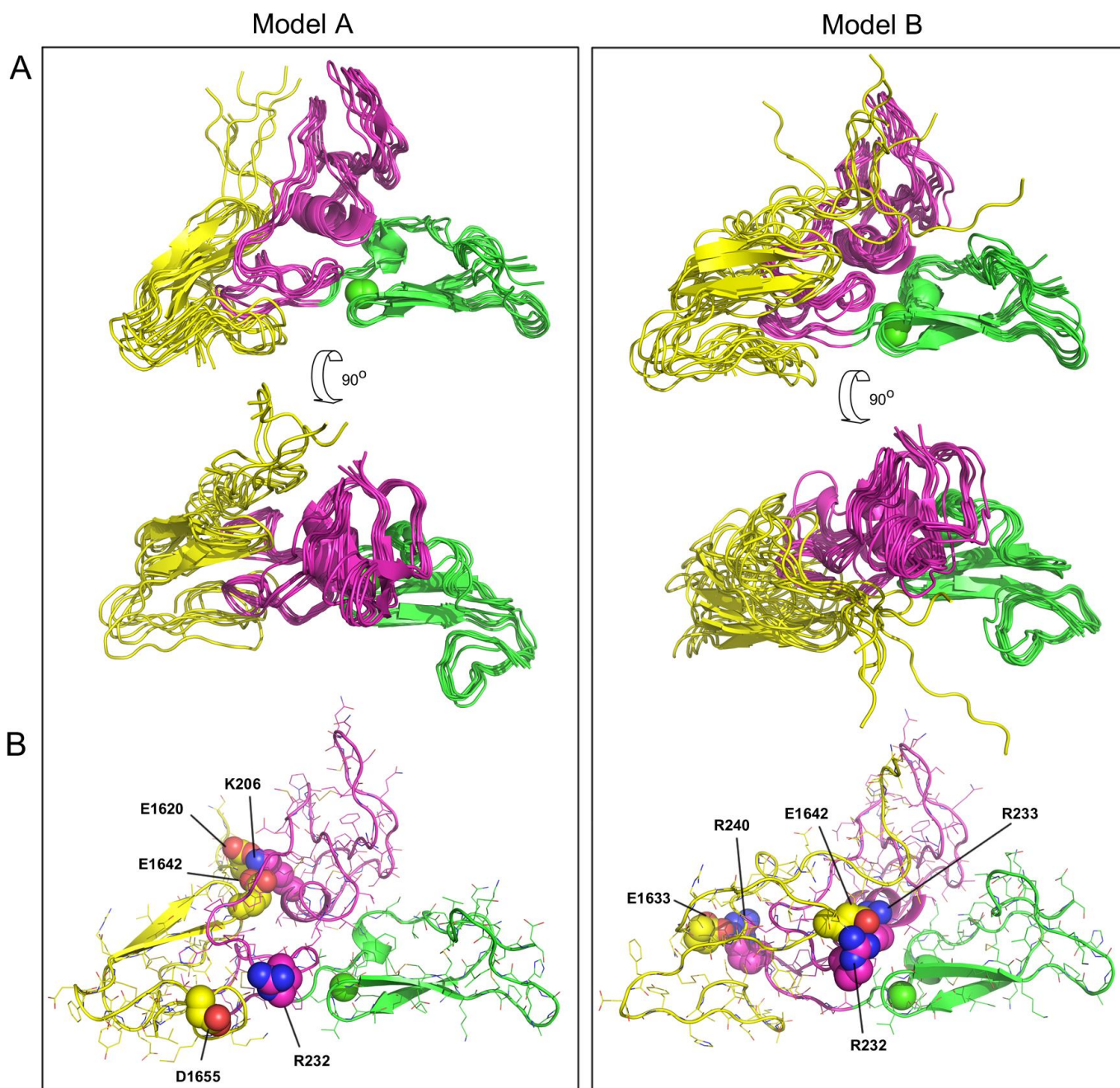


Figure 5.16 - *In silico* docking of FBN1 hyb1-cbEGF1 and the LTBP1 EGF3

Haddock modelling of the fibrillin and LTBP1 EGF3 interaction produced two dominant structural models (Table 5.3). **A**) Overlays of the top-scoring results grouped by model. For model A the following structures were used: run2- cluster2, run3-cluster1, run6-cluster3, run7-cluster3, run8-cluster8, run10-cluster6. For model B the best results are shown from: run2-cluster3, run3-cluster3, run6-cluster1, run7-cluster2, run8-cluster1, run9-cluster 4 and 1, run10-cluster2. LTBP1 EGF3 is coloured yellow, fibrillin hyb1 coloured magenta, and fibrillin cbEGF1 coloured green. Structures superimposed by aligning them to a single FBN^{e2cb1} structure using the pymol alignment tool **B**) Putative salt bridges in structures representative of the two models, with side chains of potential salt bridge residues highlighted as spheres. For model A the top structure from run10-cluster6 is used here. For model B the top structure from run3-cluster3 is shown.

fibrillin1 hyb1. In model B, the LTBP1 EGF3 sandwich contains both the C-terminal loop of fibrillin hyb1 and the preceding loop defined by a “PHP” motif.

The two models also contain different patterns of putative salt bridges; arginine 232 in the fibrillin1 hyb1 domain is potentially involved in electrostatic interactions in both models, but in model A it is positioned near to aspartate 1655 in the C-terminal loop of LTBP1 EGF3, while in model B it appears to interact with glutamate 1642 on the end of the main β -hairpin of LTBP1 EGF3. In model A there are also potential salt bridges between lysine 206, on the core α -helix of fibrillin hyb1, and glutamates 1620 and 1642 in the N-terminus and main β -hairpin of LTBP1 EGF3. In model B however lysine 206 does not appear to be involved in the interaction. In some model B structures, including the one shown in Figure 5.16 B, an additional putative salt bridge can be seen between arginine 240, in the C-terminal loop of the fibrillin hyb1 domain, and glutamate 1633, in the turn region prior to the main β -hairpin in LTBP1 EGF3. However this putative salt bridge is only seen in a subset of model B structures (run 3-cluster3, run 8-cluster1, and run 6-cluster1), while the one between arginine 232 and glutamate 1642 is always seen in this model.

While these docked fibrillin - LTBP complex structures are only models, and based on limited binding site and structural data, they still provide valuable information on the potential mechanisms of the LTBP1 - fibrillin1 interaction⁴. They also give predictions that can be tested experimentally with site-directed mutagenesis.

5.7- Conclusions

These studies of limited fragments have provided a variety of novel insights into the interaction between LTBP1 and fibrillin1. The SPR results show no interaction of the

⁴ Attempts at docking modelled LTBP1^{TB3E3} and FBN^{e2cb1} structures did not give consistent results due to the significant ambiguity introduced into the restraints, and by the large flexible linker between TB3 and EGF3 generating a very large number of potential domain orientations.

LTBP fragments with the negative control protein, while all three LTBP1 fragments do interact with FBN^{e2cb1}. This verifies the interaction seen by other groups using eukaryotically expressed proteins, and also demonstrates that the interaction is not critically dependent on any eukaryotic post-translational modifications. Although the affinities measured by other groups give lower K_d values than the estimates of affinity determined here, this may be explained by use of different buffers, and their use of kinetics based measurements of K_d rather than the equilibrium measurements used here⁵. These previous studies also used lower concentrations of analyte, and did not report whether plots of their data were consistent with 1:1 binding, or whether higher protein concentrations were tested.

The SPR experiments also revealed an interesting pattern of specificity between certain fibrillin1 and LTBP1 fragments, suggesting that multiple binding sites on both fibrillin1 and LTBP1 molecules may be involved in the interaction. HSQC titrations confirmed this, demonstrating binding between LTBP1 - TB3 and fibrillin1 EGF3, with involvement of adjacent domains, and between LTBP1 – EGF3 and the C-terminal region of the fibrillin hyb1 domain. This binding site information has also allowed models to be generated of the LTBP1 TB3 – fibrillin1 and LTBP1 EGF3 - fibrillin complexes. These complexes all contain putative salt bridges, which may suggest suitable target residues for site-directed mutagenesis.

A question still remains over the precise role of the poorly conserved interdomain linker region between LTBP1 TB3 and EGF3. Dynamics studies have shown this linker to be flexible (Section 4.8, figure 4.15), and it is possible the length and flexibility of this linker is important to allow simultaneous interaction of LTBP1 TB3 and EGF3 with their separate binding sites. This would explain why the presence of a long linker is conserved but not its specific sequence (Chapter 2, Figure 2.7). It cannot be ruled

⁵ Most sensorgrams had insufficient curvature to attempt kinetic based estimates of K_d

out that this linker region also interacts with fibrillin, as attempts to purify the linker peptide by itself were unsuccessful⁶. Attempted HSQC titrations with the ¹⁵N-labelled LTBP1^{TB3E3} fragment were also un-interpretable due to excessive protein degradation. However the poor sequence conservation in the linker region would suggest it does not have a direct and significant role in the LTBP1 - fibrillin1 interaction. The use of a homology model of FBN^{e2cb1} in these docking interactions also introduces some uncertainty into the models, particularly in relation to the fibrillin EGF3-hyb1 interface. Future docking studies, using the structural insights into FBN^{e2cb1} gained in chapter 3, will help to address this issue.

From the data in this chapter and HADDOCK docking simulations, two structural models for the complete LTBP1 - fibrillin1 interaction are proposed (summarised in Figure 5.17). Further studies are required to validate the accuracy of these interaction models. At the end of NMR titrations, HSQC and HMQC spectra collected with an excess of unlabelled protein rarely demonstrated any new peaks that could correspond to residues in the bound state (Figures 5.7, 5.9, 5.11, and 5.13). This is potentially due to the high molecular weight of the bound complex, or due to failure to reach complete saturation of the labelled protein. This difficulty obtaining spectra for proteins in the bound state makes it unrealistic to use other available NMR techniques, such as RDC's, or NOESY experiments, to probe the complex⁷. Instead attempting to disrupt the LTBP1 – fibrillin1 interaction via site-directed mutagenesis could provide the best way to validate the interaction models proposed here.

⁶ SPR experiments with an impure preparation of the interdomain linker showed no binding response with a concentration of 50 µM total protein.

⁷ TROSY experiments may provide a viable alternative to HSQC titrations, but require costly perdeuterated protein samples that were not produced in the course of this project.

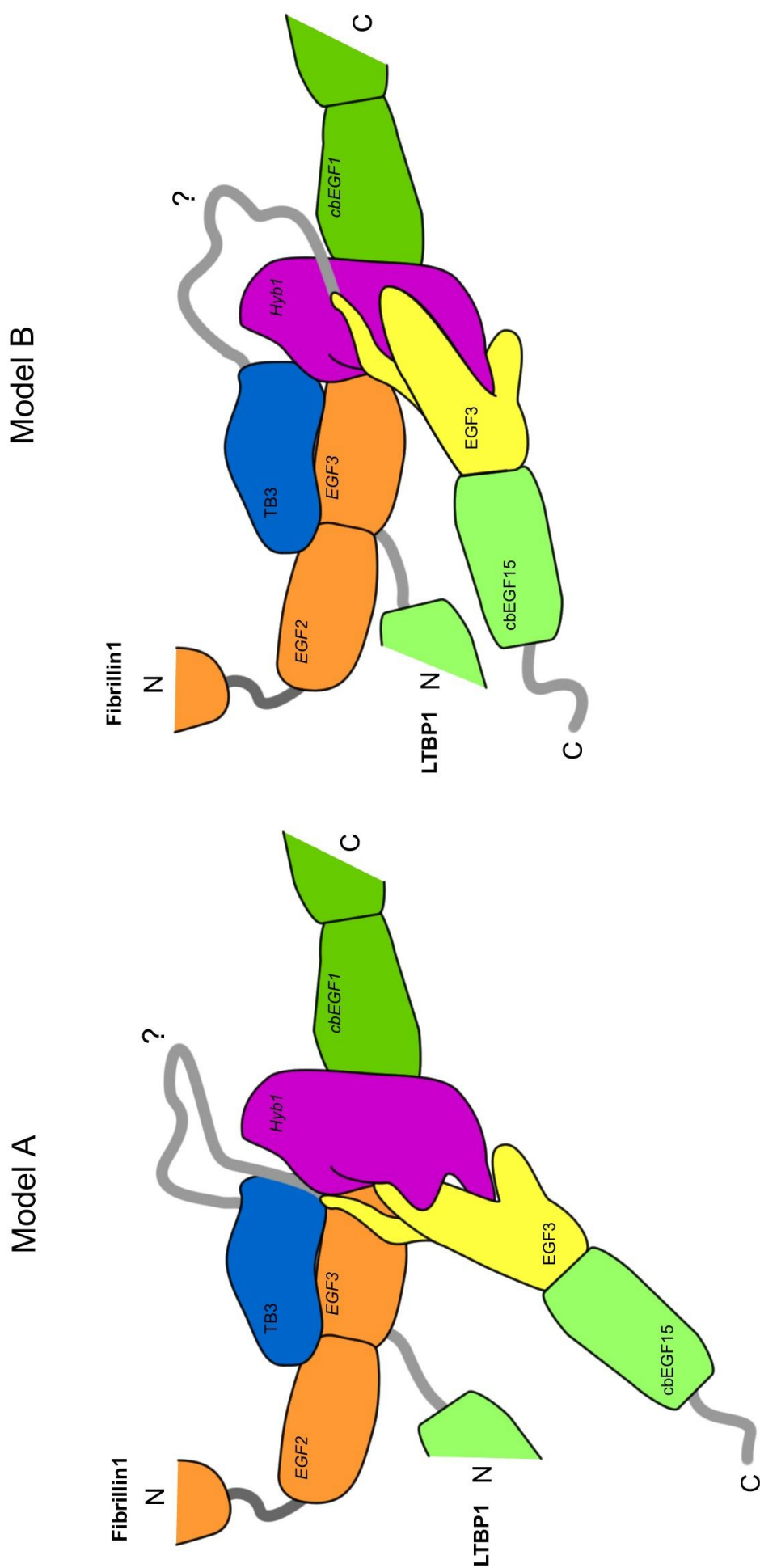


Figure 5.17 - Summary of the models for the LTBP1 - fibrillin1 interaction based on HADDOCK docking of homology models. Simplified versions of the structures in Figures 5.15 and 5.16 combined into an overall model for the LTBP1 fibrillin1 interaction. Fibrillin domain labels are italicised while LTBP1 domains are not. Domains are coloured according to the same scheme as previous figures except that fibrillin EGFs are coloured orange, and fibrillin cbEGFs are coloured a darker green. Question marks indicate that the position of the LTBP1 TB3-EGF3 interdomain linker is unclear. The relative positioning of the fibrillin EGF3 and hybrid domains in this figure is also speculative, and certain orientations may block the binding of LTBP1 EGF3 in model A (see chapter 3 and 7).

References

1. Isogai, Z., et al., *Latent transforming growth factor beta-binding protein 1 interacts with fibrillin and is a microfibril-associated protein*. J Biol Chem, 2003. **278**(4): p. 2750-7.
2. Massam-Wu, T., et al., *Assembly of fibrillin microfibrils governs extracellular deposition of latent TGF beta*. J Cell Sci. **123**(Pt 17): p. 3006-18.
3. Ono, R.N., et al., *Latent transforming growth factor beta-binding proteins and fibulins compete for fibrillin-1 and exhibit exquisite specificities in binding sites*. J Biol Chem, 2009. **284**(25): p. 16872-81.
4. Zuiderweg, E.R., *Mapping protein-protein interactions in solution by NMR spectroscopy*. Biochemistry, 2002. **41**(1): p. 1-7.
5. de Vries, S.J., M. van Dijk, and A.M. Bonvin, *The HADDOCK web server for data-driven biomolecular docking*. Nat Protoc, 2010. **5**(5): p. 883-97.
6. Jensen, S.A., et al., *Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins*. Structure, 2009. **17**(5): p. 759-68.
7. Lee, S.S., et al., *Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization*. Structure, 2004. **12**(4): p. 717-29.

Chapter 6- Disruption of the LTBP1 - Fibrillin1 interaction by site directed mutagenesis

6.1 Introduction

In order to test the interaction models generated by HADDOCK using HSQC-titration data, site-directed mutagenesis was carried out in an attempt to disrupt the LTBP1 – fibrillin1 interaction. Putative salt bridges could be identified in all of the high scoring interaction models produced by HADDOCK (Figures 5.15 and 5.16). These have the potential to contribute significant enthalpy to the binding interactions, and so provide good targets for site-directed mutagenesis. The residues involved in these salt bridges were specifically targeted, by mutating their DNA codons to code for residues of opposite charge. The rationalisation for targeting and reversing the charge of salt bridge residues was that, as well potentially contributing significant enthalpy to any interaction, it may be possible confirm specific residue-residue interactions if charge reversal of both residues restores or partially restores the interaction. This restoration of binding could theoretically occur if a new salt bridge were to form with the opposite orientation to that present between WT proteins, although this is not guaranteed to happen, as in folded proteins steric and local charge effects may interfere [1].

A total of three charge substituted FBN^{e2cb1} mutants, and five LTBP1^{TB3E3} mutants were generated to test the models proposed in Chapter 5 (Figure 6.1). A third model (model C) for the LTBP1 EGF3 – fibrillin hyb1 interaction is included in Figure 6.1, which was not discussed in Chapter 5, as it was only present in some lower scoring HADDOCK clusters¹. This model was also tested by replacement of glutamate 1625 in LTBP1 EGF3 with arginine.

¹ This third model was selected for testing after the first round of mutagenesis highlighted the importance of fibrillin arginine 232, and so other low scoring HADDOCK models involving this residue were investigated.

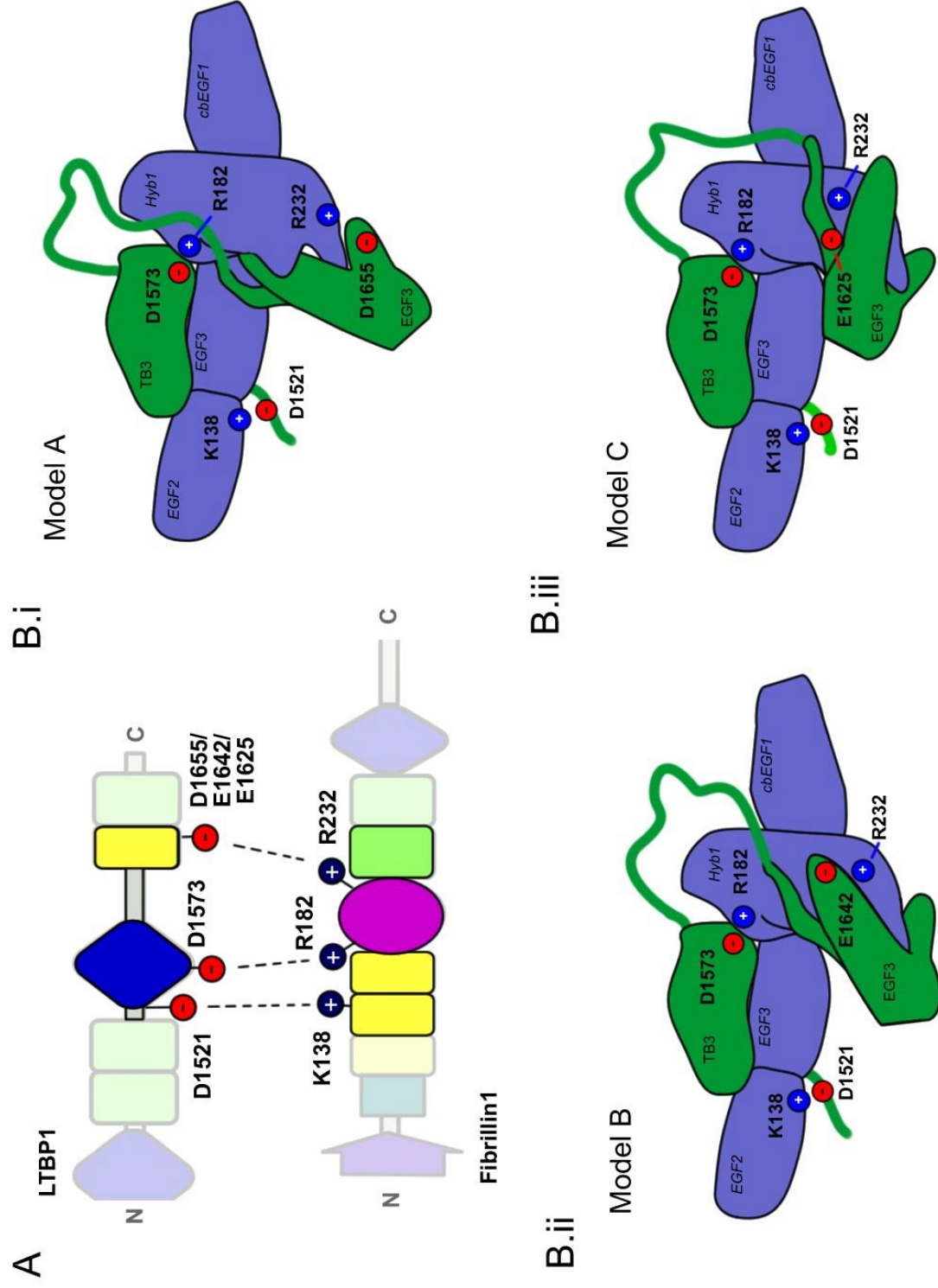


Figure 6.1 - Models for the interaction of LTBP1 and Fibrillin1 and selected salt bridges

A) Simple domain summary of the LTBP1 C-terminus and fibrillin 1 N-terminus (domain legend as in Figure 5.1). Residues suggested by HADDOCK models as forming putative salt bridges are highlighted. **B)** Simplified structural models for the interaction of LTBP1^{TB3E3} and FBN^{e2cb1} inferred from the experiments in chapter 5. Fibrillin domains are coloured marine blue and LTBP1 domains are coloured green. Putative salt bridges that were targeted by mutagenesis are highlighted in each model. The following salt bridges are model specific; R232-D1655 : model A, R232-E1642 : model B, R232-E1625 : model C.

As well as being selected for their potential to distinguish between models, where possible residues were also selected that were predicted to be structurally unimportant. For example, fibrillin1 lysine 206 was not targeted as it is predicted to pack with the core tryptophan of the hyb1 domain, and is also highly conserved (Figure 2.8).

Substituted proteins were produced and characterised in the same way as WT proteins, via prokaryotic expression and *in vitro* refolding (see materials and methods, Section 8.15). Their interactions were then analysed by SPR to examine the effects of the substitutions on binding.

6.2 Production and characterisation of Fibrillin1^{e2cb1} mutants

FBN^{e2cb1} cloned into the pQE30 expression vector (with existing C204S substitution) was specifically altered using the “quik-change” mutagenesis protocol (Agilent) (Section 8.8). Three substitutions were introduced into separate constructs; a K138>D substitution was introduced to test whether the putative salt bridge between fibrillin lysine 138 in EGF2 and LTBP1 aspartate 1521 in the flexible region before TB3 contributed to the interaction, or was just imposed by the docking calculations. A R182>E substitution was introduced as this residue appears to form a salt bridge between fibrillin hyb1 and the LTBP1 TB3 domain, specifically aspartate 1573 in the second α -helix of the TB3 domain, (and possibly also aspartate 1572 depending on the parameters used in docking). Finally an R232>E substitution was introduced, as in both HADDOCK models A and B of the fibrillin1 hyb1 - LTBP1 EGF3 interaction this residue is implicated as forming a salt bridge, either with aspartate 1655 in the C-terminal loop of LTBP1 EGF3 in model A, or with glutamate 1642 on the main β -hairpin of LTBP1 EGF3 in model B.

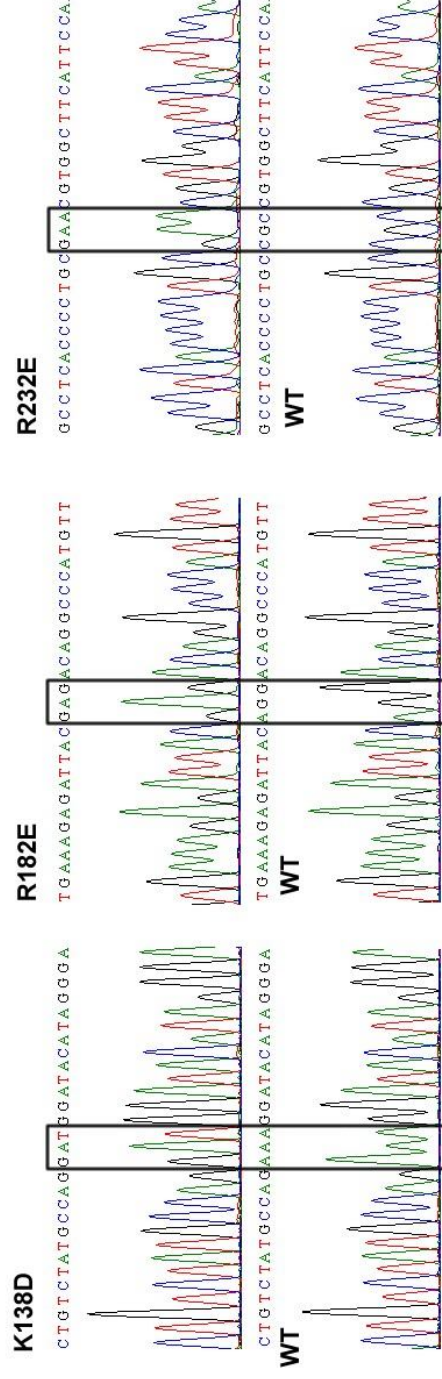
Sequencing confirmed the success of the mutagenesis (Figure 6.2 A), and the substituted proteins were then produced by expression in *E. coli* BL21 cells, inclusion body extraction, *in vitro* refolding, and HPLC purification as described for WT proteins (Section 3.2 and Section 8.9 - 8.12). The His-tag was also removed by Xa cleavage followed by FPLC and HPLC purification.

Figure 6.3 shows the final HPLC traces for the mutant fragments, as well as SDS PAGE and mass spectrometry data. From these data it appears that all mutants behave similarly during HPLC, in reducing and non-reducing SDS-PAGE, and mass spectrometry results were consistent with their expected masses². After interaction experiments, the FBN^{e2cb1} protein with an R232E substitution was singled out for further structural analysis by 2D NOESY NMR spectra (Section 8.27). 2D NOESY spectra from this protein were very similar to those of WT FBN^{e2cb1}. The aromatic regions of the spectra from both proteins are overlaid in Figure 6.3 D, demonstrating similar peaks for all aromatic residues³, including tryptophan 217 at the core of the hybrid domain. Phenylalanine 235, a residue just downstream of arginine 232, and at the interface of hyb1 and cbEGF1, seems to exhibit a very slight shift, which could be just due to the nearby side chain substitution. The similarity of these peak positions demonstrated that the R232E substitution does not disrupt the fold of FBN^{e2cb1}. Also arginine 232 is not conserved in organisms lacking LTBP, and is in fact replaced with a glutamate residue in *T. castaneum* (beetle) fibrillin, and an aspartate residue can be found in the corresponding position in many other early fibrillins (Chapter 2 fig 2.8), which further supports this residue being structurally unimportant.

² There were some minor mass increases (< 6 Da) in some cases, but these were consistent between those samples that were measured at the same time, and were likely due to altered calibration of the instrument.

³ Only the aromatic regions are shown in Figure 6.3 and 6.4 as in this case these NOESY peaks provide information on all domains, and also maintain clarity.

A Fibrillin mutants



B LTBP mutants

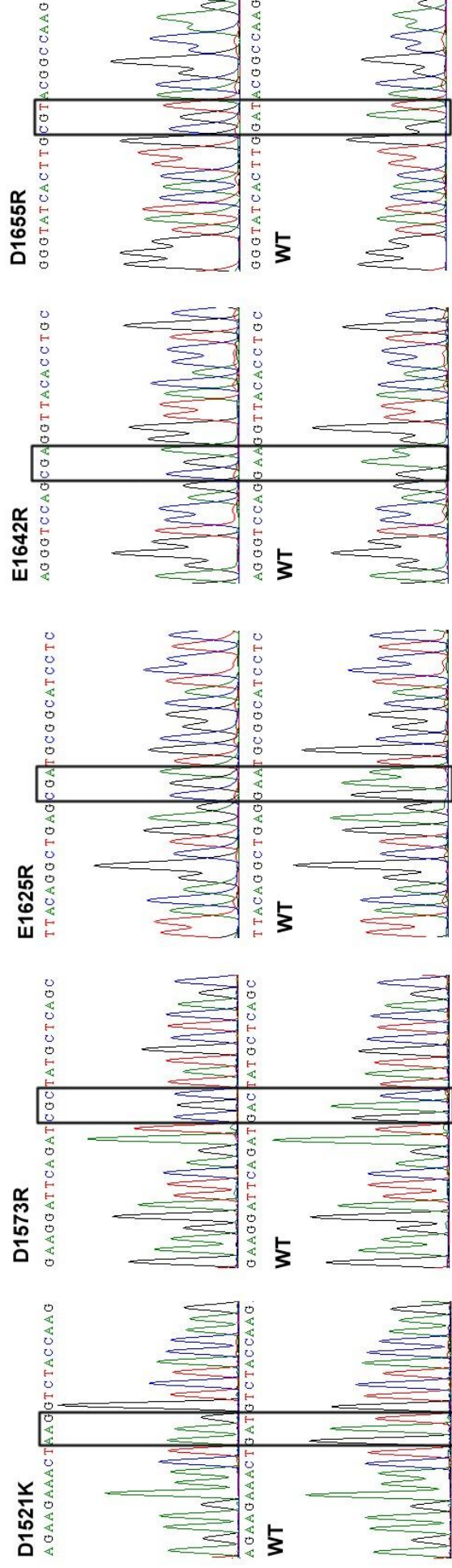


Figure 6.2 - Sequence confirmation of fibrillin1 and LTBP1 mutagenesis

Sanger sequencing chromatograms from relevant regions of **A)** FBN1^{TB3E3} expression plasmid sequences. Targeted codons are highlighted by black boxes, with the relevant DNA sequences shown above the chromatograms. All results give the correct sequence with no signs of WT DNA contamination.

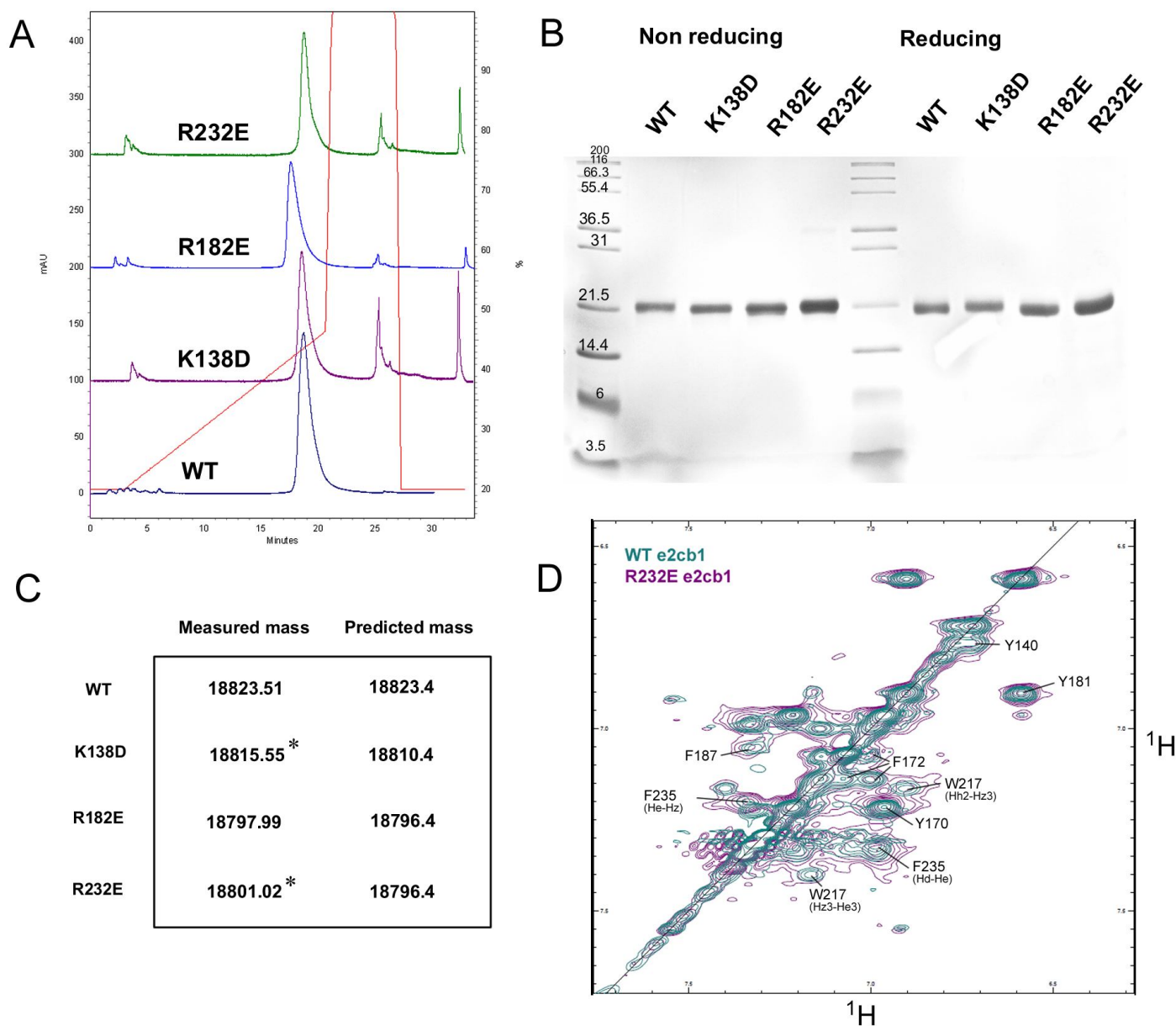


Figure 6.3 - Characterisation of FBN^{e2cb1} substituted proteins

A) C8 reverse phase HPLC chromatograms of refolded FBN^{e2cb1} fragments after factor Xa cleavage and ion exchange FPLC (as in Chapter 3). Red line represents the acetonitrile gradient used (Section 8.13) **B)** SDS-PAGE of final purified fragments under both reducing and non-reducing conditions, stained with Coomassie (Section 8.19). **C)** ESI mass spectroscopy results of substituted proteins. There were minor mass increases (~5 Da) in two of the samples (marked with an asterisk), but were consistent between samples measured at the same time, and were likely due to calibration of the instrument. **D)** Overlaid aromatic region of 2D NOESY spectra of WT and R232E FBN^{e2cb1} showing similar peak positions in both fragments (Section 8.27). Both spectra were collected without calcium at pH 5.4, and labelled with available aromatic assignments.

6.3 Production and characterisation of LTBP1^{TB3E3} mutants

LTBP1^{TB3E3} cloned into the pQE30 expression vector was also specifically altered using the “quik-change” mutagenesis protocol (Agilent) (Section 8.8). The following substitutions were introduced individually: D1521>K, D1573>R, E1625>R, E1642>R, and D1655>R, in order to test the putative salt bridges previously discussed. Sequencing confirmed the success of mutagenesis (Figure 6.2 B), and proteins were then expressed and purified as described for WT LTBP1^{TB3E3} (Section 4.2 and Section 8.9-8.11).

Some slight differences in C8 reverse phase HPLC chromatograms were observed between refolded proteins with various substitutions (Figure 6.4 A), particularly with the E1625R substitution. However the presence of single peaks in similar positions is still suggestive of protein folding. When SDS-PAGE was carried out on refolded LTBP1^{TB3E3} proteins, marked differences in migration could be seen between different substituted proteins under both reducing and non-reducing conditions (Figure 6.4 B). One explanation for these SDS-PAGE results might be a significant change in protein mass, but the mass spectrometry results were all consistent with those expected for each substitution (Figure 6.4 C). An alternative explanation would be that some of the substitutions of negative with positively charged side chains are suitably placed to affect the overall mobility of the proteins in an electric field, potentially through altered interactions with SDS. After interaction experiments the D1655R substitution was singled out for further structural analysis to check its fold, and a sample of this protein was analysed with a 2D NOESY NMR experiment (Figure 6.4D). Generally the spectrum of D1655R appeared consistent with those from wild type LTBP1^{TB3E3} protein, although a few peaks that are present in the WT spectra appear absent, or slightly shifted, in the D1655R spectra. This may be due to local changes brought

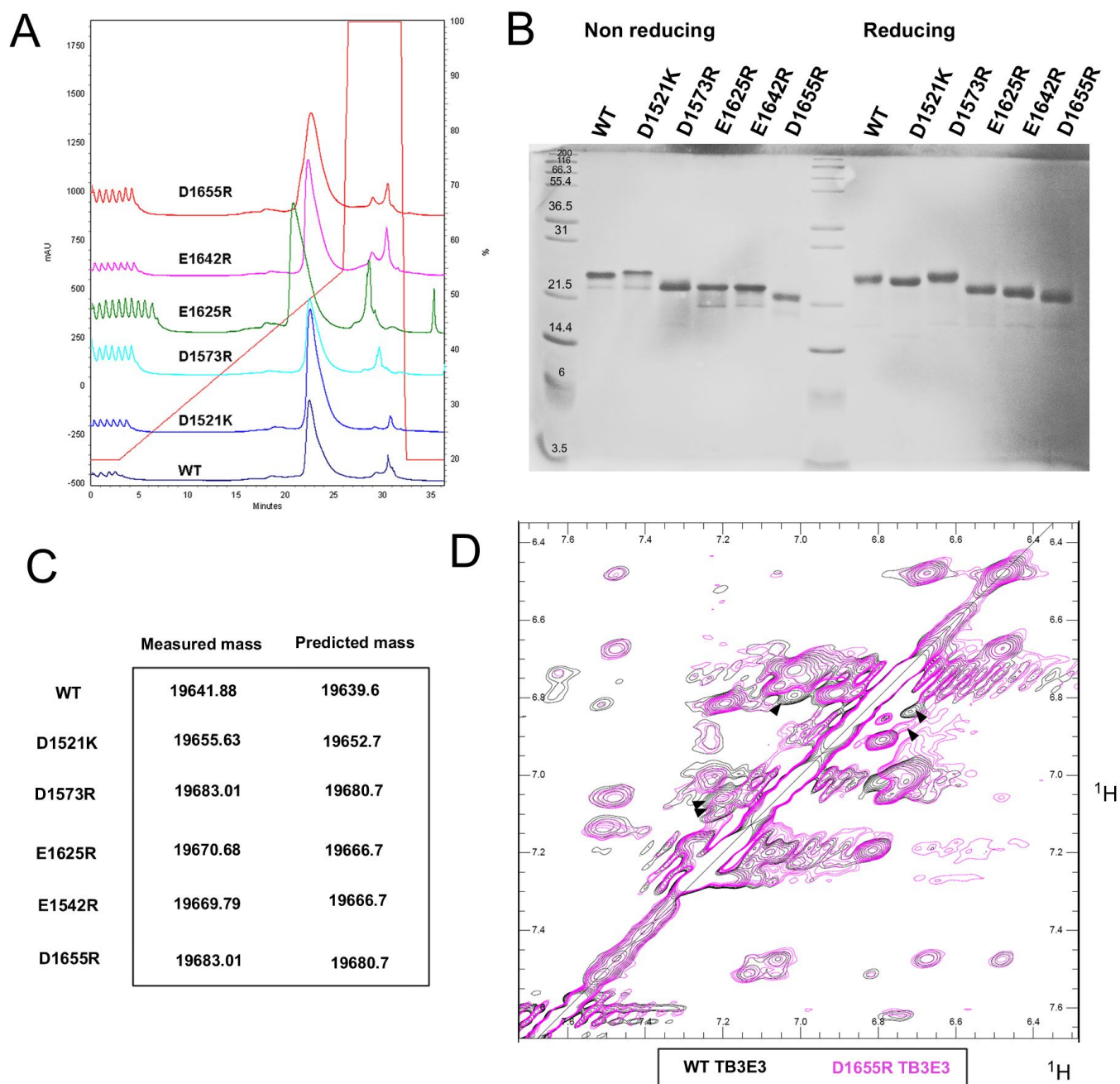


Figure 6.4 - Characterisation of LTBP1^{TB3E3} substituted proteins

A) HPLC chromatograms of LTBP1^{TB3E3} fragments directly after refolding. Red line represents the acetonitrile gradient used (Section 8.13) **B)** Refolded samples analysed by SDS-PAGE with Coomassie staining, under both non-reducing and reducing conditions. **C)** ESI mass spectrometry results for the five LTBP1^{TB3E3} fragments, confirming purification of the correct protein. Measured masses were consistently ~2-3 Da greater than predicted masses, due to the calibration of the instrument. **D)** Overlaid aromatic region of 2D NOESY spectra collected from WT and D1655R LTBP1^{TB3E3} fragments, the pH of both samples was adjusted to 7.7, and spectra were collected on a 600 MHz spectrometer, although the samples were at different protein concentrations. WT and D1655R spectra have a very similar pattern of peaks, and the D1655R sample did not display any new sharp peaks indicative of misfolding. Some peaks can be seen to be absent or slightly shifted in the D1655R sample which may be due to localised changes brought about by the aspartate to arginine substitution. (Comprehensive aromatic assignments were not available for this protein)

about by the presence of a different side chain in the EGF3 domain. The absence of any new sharp peaks in the D1655R spectrum suggests there is no significant disruption of folding. This is consistent with the fact that D1655R LTBP1^{TB3E3} still showed a significant level of interaction with FBN^{e2cb1}, and the fact that aspartate 1655 is predicted to be a surface residue, makes it unlikely that the D1655R substitution would cause any significant disruption to the folding of LTBP1^{TB3E3}.

6.4 - Characterisation of mutant interactions by SPR

In order to test whether any of the amino acid substitutions diminished the fibrillin1 - LTBP1 interaction, SPR experiments were performed in a similar manner to those described previously (Section 5.4 and Section 8.26). The FBN^{e2cb1} fragments were immobilised on CM5 sensor chips as ligands, and WT and substituted LTBP1^{TB3E3} proteins flowed over as analytes. Four flow cells were available on each CM5 chip, one of which was used as the blank reference and the other three were coupled with different FBN^{e2cb1} fragments. Two chips were used in total, one coupled with WT, K138D, and R232E FBN^{e2cb1} fragments, and the other with WT, R182E, and R232E fragments. Amine coupling was achieved with all proteins to 1000 RU and was stable when regenerated with 50 mM HCl. All fibrillin proteins appeared soluble in coupling buffer, and LTBP^{TB3E3} fragments were soluble in SPR buffer, and no fragments showed any sign of aggregation, that could disrupt the SPR measurements, and also no non-specific binding to negative control proteins was seen.

Substituted proteins were produced in two rounds and the single cycle SPR data from the proteins produced in the first round is shown in Figure 6.5. Only the R232E substitution in this round seemed to disrupt binding, and this was taken forward for further analysis along with the second round of substituted proteins. The results of the

second round of single cycle SPR experiments are shown in Figure 6.6. From these initial experiments substituted proteins of particular interest were further analysed using multi-cycle SPR experiments (Figure 6.7) (Section 8.26), with similar results obtained to those in the single cycle experiments. Plots of SPR data are shown in Figure 6.8, and estimates of K_d are shown in Table 6.1 and 6.2. As discussed in Chapter 5, the non-linear nature of the Scatchard plots suggests using 1:1 Langmuir model binding, to estimate affinity is not accurate. However these estimates may still be useful for qualitative comparison between substituted proteins.

From the substitutions tested, several did not appear to have any significant affect on the interaction. For example the replacement of lysine 138 in fibrillin EGF2 with aspartate did not have a measurable impact on binding response, and neither did the replacement of aspartate 1521 with lysine in the flexible region at the start of LTBP1 TB3. In several high scoring HADDOCK structures of the LTBP1 TB3 – FBN^{e2cb1} interaction these two residues were predicted to form a salt bridge, but these results suggest either this salt bridge does not form, or is redundant in the overall interaction of LTBP1 and fibrillin1.

The substitutions of glutamate 1625 and 1642 with arginine in LTBP1 EGF3 also do not appear to significantly disrupt the LTBP1 - fibrillin1 interaction, giving binding responses similar to, or greater than, WT LTBP1^{TB3E3}. This suggests they do not play a significant role in driving the fibrillin1 – LTBP1 interaction, and that the models where they form putative salt bridges may be inaccurate.

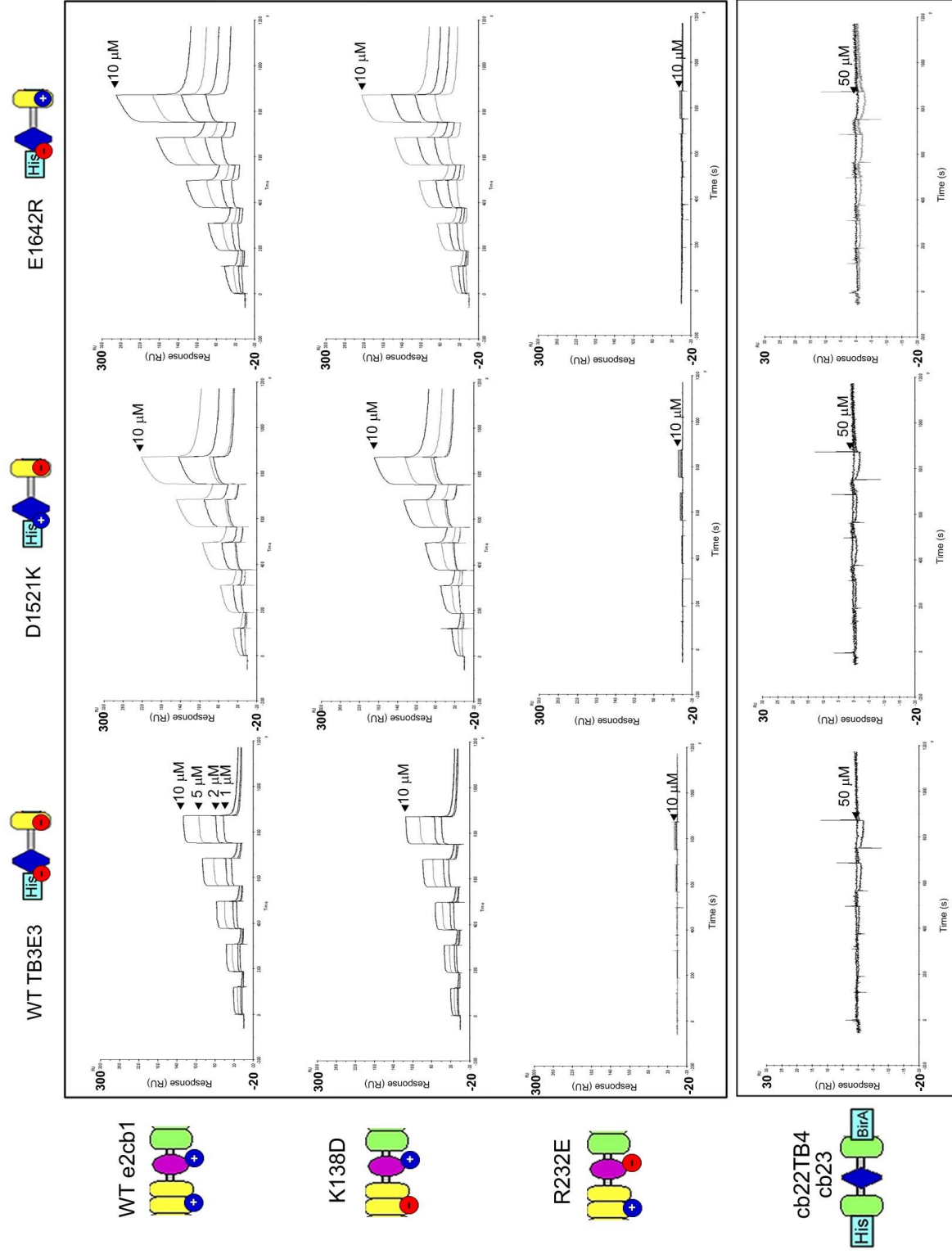


Figure 6.5 - Single cycle SPR results for the first round of LTPB1 and fibrillin1 substitutions
Blank subtracted sensorgrams for WT, D1521K, and E1642R LTPB1^{TB3E3} fragments flowed over a CM5 chip coupled with WT, K138D, and R232E FBN1^{e2cb1} fragments. The single cycle method was used (Section 8.26), with four runs of five 2x serial dilutions, starting from stocks of 10, 5, 2 and 1 μ M, for both mutant and WT LTPB1^{TB3E3} fragments. LTPB1 proteins were also flowed over another negative control chip coupled with the fibrillin1 cb22TB4cb23 fragment (bottom), and showed no interaction with this protein. SPR conditions were the same as used previously (Section 8.26).

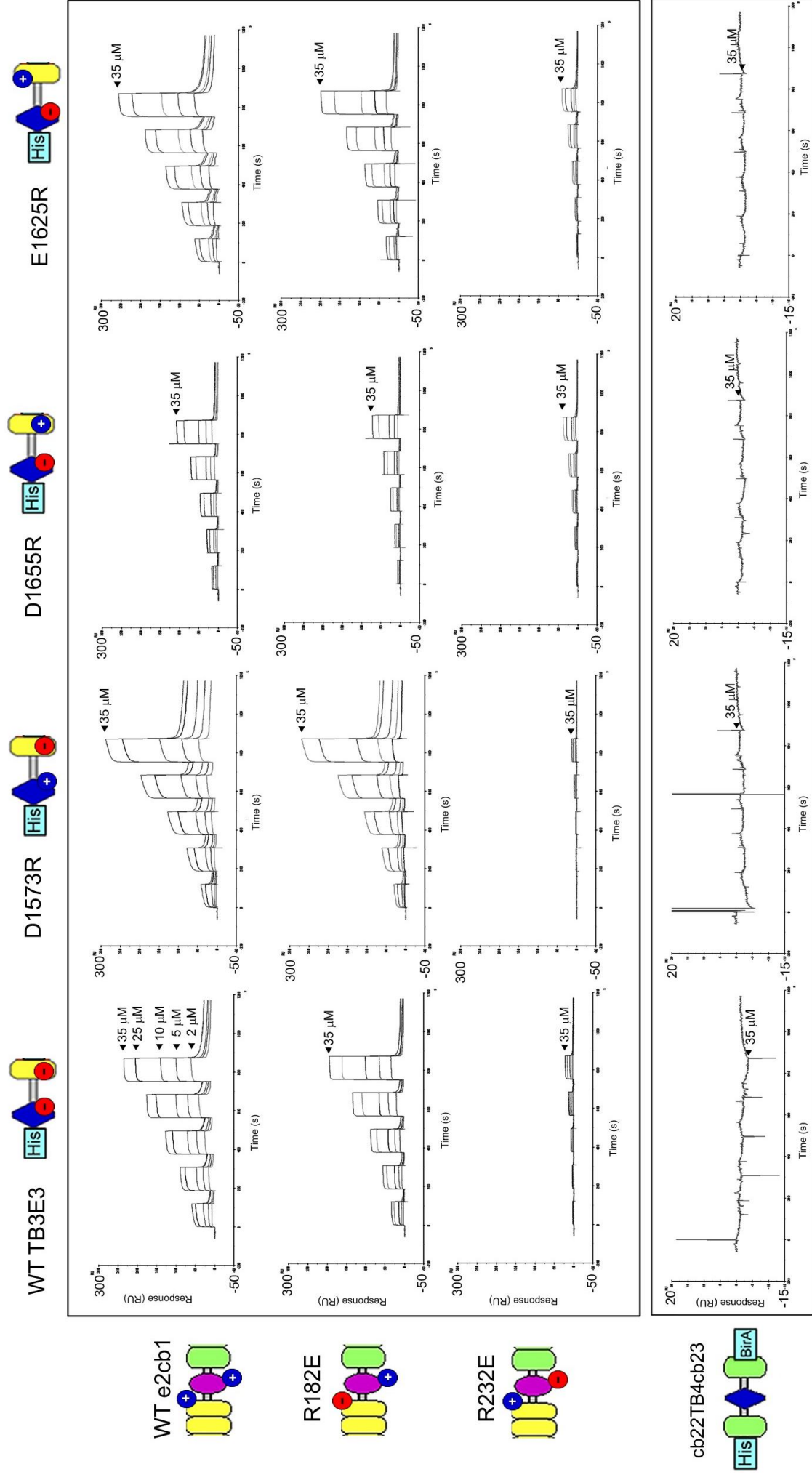


Figure 6.6 – Single cycle SPR results for the second round of LTP1 and fibrillin1 substitutions

Blank subtracted sensorgrams for WT, D1573R, E1625R, and D1655R LTPB1^{TB3E3} fragments flowed over a CM5 chip coupled with WT, R182E, and R232E FBN1^{e2cb1} fragments. The single cycle method was used (Section 8.26), with five runs of five 2x serial dilutions, starting from stocks of 35, 25, 10, 5, and 2 μ M, for all mutant and WT LTPB1^{TB3E3} fragments. LTPB1 proteins were also flowed over another negative control chip coupled with the fibrillin1 cb22TB4cb23 fragment (bottom), and showed no interaction. SPR conditions were the same as used previously (Section 8.26).

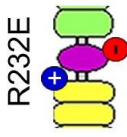
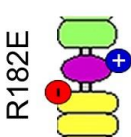
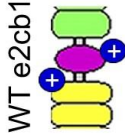
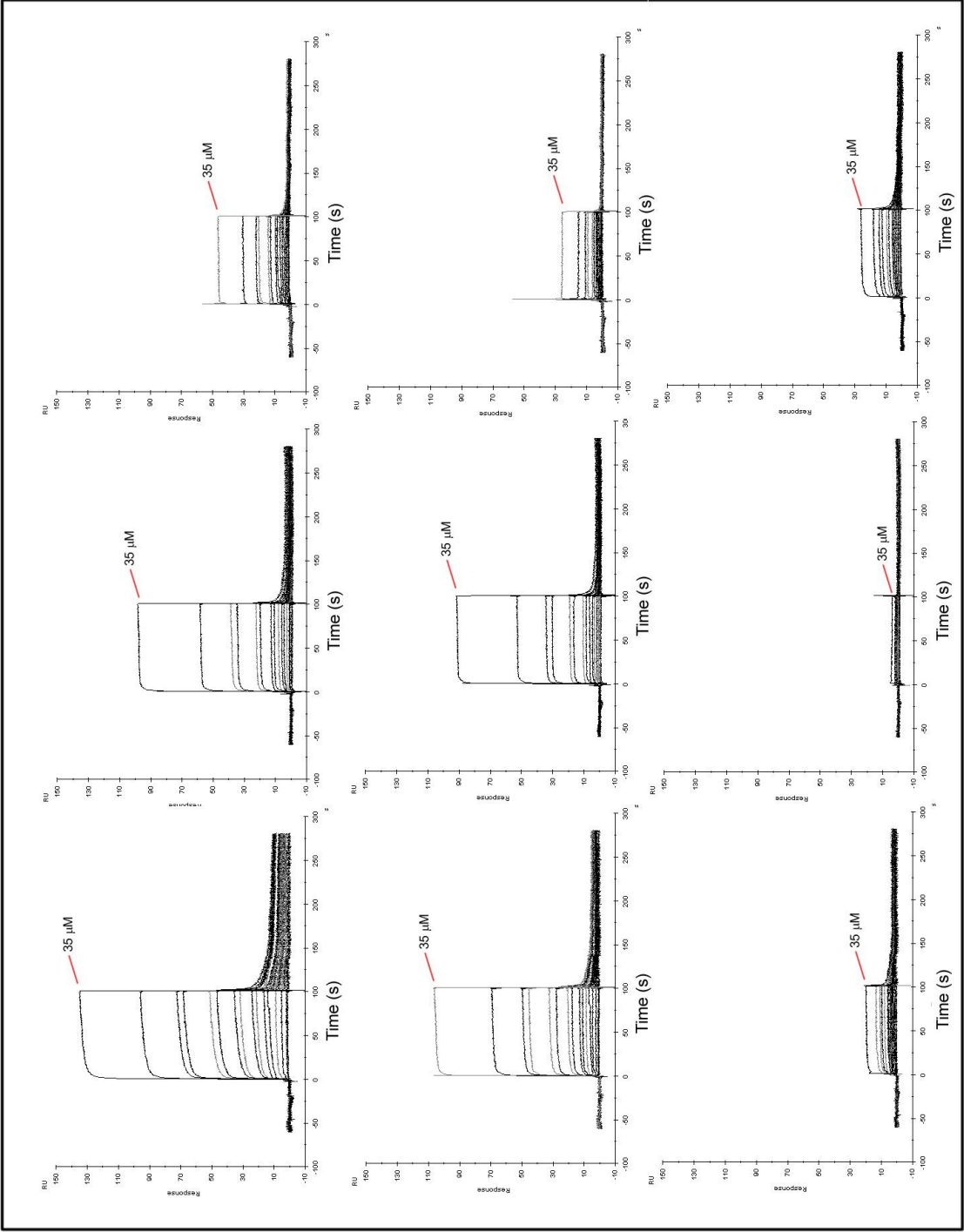
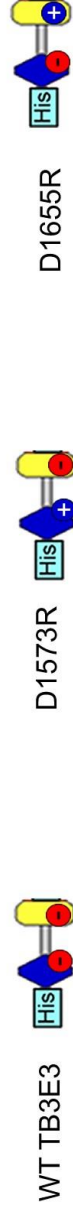


Figure 6.7 - Multi-cycle SPR experiments with LTP1 and fibrillin1 substituted fragments of partiuclar interest
Blank subtracted sensorgrams for the interaction of WT, D1573R, and D1655R LTP1^{TB3E3} fragments in solution with WT, R182E, and R232E FBN^{e2cb1} fragments on the chip. The following series of concentrations were used for all LTP1^{TB3E3} proteins: 35, 17.5, 10, 8.75, 5, 4.38, 2.5, 1.25, 1.09, 0.625, 0.547, 0.273, 0.137, 0.068, and 0.034 μ M. SPR conditions were the same as used previously (Section 8.26).

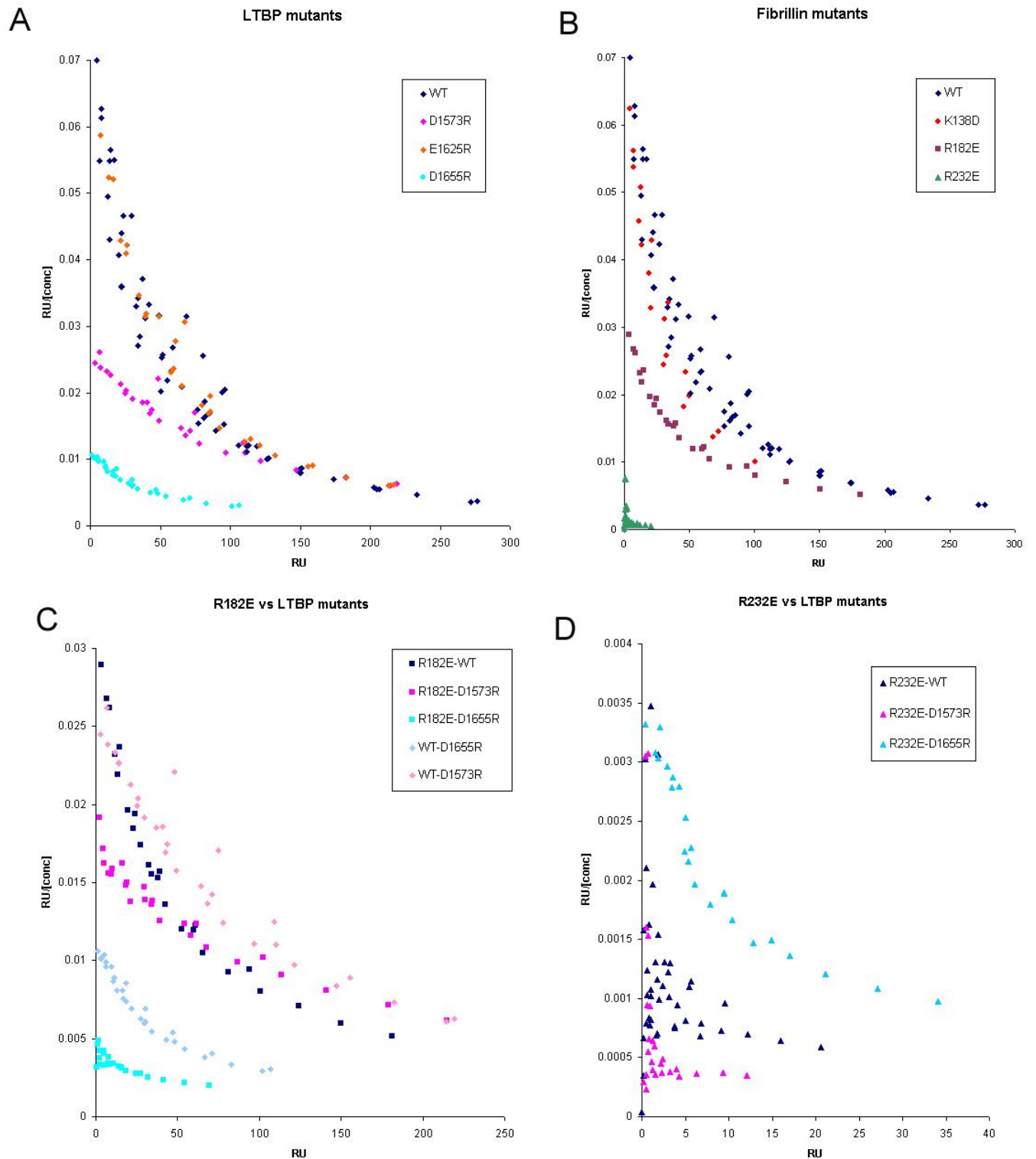


Figure 6.8 - Scatchard plots of single cycle SPR data with selected LTBP1 and fibrillin1 substitutions
 Plots of (response/concentration) vs (response) for; **A)** Selected substituted LTBP1^{TB3E3} fragments binding to WT FBN^{e2cb1}, **B)** WT and substituted fibrillin fragments binding to WT LTBP1^{TB3E3}, **C)** Different LTBP1^{TB3E3} fragments binding to R182E FBN^{e2cb1}, with selected responses with WT FBN^{e2cb1} included for comparison (shown as faded diamond shaped data points), **D)** Selected LTBP1^{TB3E3} fragments binding to R232E FBN^{e2cb1} (responses below 1-2 RU become difficult to distinguish from background noise and other artefacts, which may account for the range of values for each protein in this region). To represent data points specific shapes were used for interactions with specific fibrillin ligands as follows; diamonds = WT FBN^{e2cb1}, circles = K138D, squares = R182E, and triangles = R232E. Where possible different colours were used consistently for each LTBP1^{TB3E3} protein, although this was varied in B) and C) so that different sets of data could be easily distinguished.

Table 6.1 – Estimates of affinity from single-cycle data.

Non-linear regression based on 1:1 binding was used to generate these values. As Scatchard plots of the data are non linear, these results are likely to be inaccurate, but are shown here for qualitative comparison. The maximum concentration of analyte used for fitting is given before each set of values, and due to the non-Langmuir binding seen, comparisons should only be drawn between values using the same maximum concentration. Significant variation in predicted $R_{\max}$ is also seen, which in theory should be about the same for all constructs, as their masses are very similar. This underlines the potential inaccuracy of these values. Values given to one decimal place. Affinity values in table **A)** were obtained from single cycle data, and those in table **B)** were obtained from multi-cycle data. In B the maximum concentration of analyte used for fitting was 35 μM for all proteins. Results in red are likely to be especially inaccurate due to the significantly smaller estimate of $R_{\max}$.

6.1 A)

Analyte Ligand	WT ^{TB3E3}	D1573R	D1655R	E1625R	D1521K	E1642R
WT ^{E2cb1}	(35 μM max) K_d 12.5 μM $R_{\max}$ 290.3 χ^2 47.9 (10 μM max) K_d 4.5 μM $R_{\max}$ 162.0 χ^2 49.2	(35 μM max) K_d 23.4 μM $R_{\max}$ 463.8 χ^2 16.6	(35 μM max) K_d 31.1 μM $R_{\max}$ 202.6 χ^2 7.24	(35 μM max) K_d 14 μM $R_{\max}$ 329.4 χ^2 42.6	(10 μM max) K_d 9.4 μM $R_{\max}$ 417.1 χ^2 6.76	(10 μM max) K_d 5.2 μM $R_{\max}$ 383.8 χ^2 76.2
R182E	(35 μM max) K_d 21 μM $R_{\max}$ 290.4 χ^2 11.2	(35 μM max) K_d 33.7 μM $R_{\max}$ 518.3 χ^2 8.44	(35 μM max) K_d 60.5 μM $R_{\max}$ 193.8 χ^2 0.289	(35 μM max) K_d 26.4 μM $R_{\max}$ 330.2 χ^2 7.10	nd	nd
K183D	(10 μM max) K_d 4.9 μM $R_{\max}$ 157.8 χ^2 13.9	nd	nd	nd	(10 μM max) K_d 8.5 μM $R_{\max}$ 337.4 χ^2 5.79	(10 μM max) K_d 5 μM $R_{\max}$ 311.5 χ^2 57.7

6.1 B)

Analyte Ligand	WT ^{TB3E3}	D1573R	D1655R	E1625R
WT ^{E2cb1}	K_d 15.3 μM $R_{\max}$ 176.2 χ^2 13.9	K_d 56.5 μM $R_{\max}$ 252.9 χ^2 1.59	K_d 28.5 μM $R_{\max}$ 79.42 χ^2 0.911	K_d 18.4 μM $R_{\max}$ 162.5 χ^2 9.10
R182E	K_d 28.2 μM $R_{\max}$ 181.4 χ^2 4.07	K_d 70.8 μM $R_{\max}$ 275.0 χ^2 0.916	K_d 58 μM $R_{\max}$ 65.63 χ^2 0.221	K_d 35 μM $R_{\max}$ 158.5 χ^2 1.68

6.4.1 Substitutions disrupting the LTBP1 EGF3 - fibrillin1 hyb1 interaction

A clear reduction in binding is observed with the fibrillin R232E substitution in hyb1, giving SPR responses of less than 25 RU with 35 μ M WT LTBP1^{TB3E3} as analyte. This residue is implicated as forming a salt bridge in several HADDOCK interaction models, and protein with this substitution has been shown to fold normally (Figure 6.3). This result therefore supports an important role for this fibrillin residue in binding LTBP1.

In HADDOCK interaction models B and C (Figure 6.1) fibrillin arginine 232 forms a salt bridge with E1625 or E1642, but substitution of these residues in LTBP1^{TB3E3} does not affect binding. However in HADDOCK model A, aspartate 1655, in the C-terminal loop of LTBP1-EGF3, is positioned close to fibrillin arginine 232. The LTBP1^{TB3E3} fragment with a D1655R substitution in LTBP1 EGF3, also demonstrates a clear reduction in binding to fibrillin1, although not as significant as that seen with the R232 substitution in fibrillin. Aspartate 1655 is predicted to be a surface residue in the homology models of LTBP EGF3, and it is also highly conserved through evolution (Figure 2.7B). From the substitutions tested, this residue provides the best candidate for forming a salt bridge, or driving a polar interaction, with arginine 232 in fibrillin, and so these mutagenesis results support HADDOCK model A.

When D1655R LTBP protein is flowed over R232E fibrillin (simultaneously with the WT experiment), only a small response is seen when compared to the WT - WT interaction. This means charge reversal of both residues does not seem to have restored a normal interaction via a swapped salt bridge. However close inspection of the response levels (figure 6.8 D) does show a clear increase in response relative to the interaction of WT LTBP1^{TB3E3} with R232E FBN^{e2cb1}. In contrast when the D1655R

LTBP protein is flowed over R182E fibrillin a significant reduction in binding is seen, compared with the interaction of WT LTBP1^{TB3E3} with R182E FBN^{e2cb1} (Figure 6.8 C). This exacerbated loss of binding demonstrates that both fibrillin arginine 182, and LTBP1 aspartate 1655, appear to contribute to different elements of the overall LTBP-fibrillin interaction.

The small increase in binding that is seen with the R232E FBN^{e2cb1} – D1655R LTBP1^{TB3E3} combination supports both residues providing a mutually dependent element of the LTBP1 – fibrillin1 interaction. The failure to restore the normal interaction by charge reversal does not disprove the interaction of aspartate 1655 and arginine 232, as the surrounding residues may dissipate changes in surface charge, or alter side chain arrangement. In fact, attempts at re-engineering salt bridges through charge reversal are often frustrated by these factors in folded proteins [1]. Close inspection of selected HADDOCK model A structures also reveals that in LTBP1 EGF3 several backbone carbonyls, and the side chain OH of threonine 1656, can also be positioned close to fibrillin arginine 232, and these features may also contribute to the interaction with this residue. These features may also oppose the effects of charge reversal in the substituted proteins.

6.4.2 Substitutions disrupting the LTBP1 TB3 – fibrillin1 interaction

As well as the R232E and D1655R substitutions targeting the FBN^{hyb1} – LTBP1^{EGF3} interaction, the R182E and D1573R substitutions were introduced to disrupt a putative salt bridge in the fibrillin EGF3/hyb1 – LTBP1 TB3 interaction. These substitutions also seem to reduce the binding of LTBP1^{TB3E3} and FBN^{e2cb1}, although much less markedly than the R232E and D1655R substitutions. The difference in binding with these substituted proteins is most noticeable at low analyte concentrations, and most clearly

observed in Scatchard plots (fig 6.8 C). When the interaction of R182E and D1573R was tested, to see if a new reversed salt bridge could be introduced, no improvement in binding was seen in the Scatchard plots (Figure 6.8 C). This again may be due to the compensating effects of the surrounding residues, and this lack of compensation does not suggest the model is incorrect. If the model was inaccurate, and both R182 and D1573 residues were involved in different elements of the LTBP1-fibrillin1 interaction, a greater loss of overall binding would be expected with the two mutants together, as with that seen in the interaction of R182E FBN^{e2cb1} with D1655R LTBP1^{TB3E3}, which appears considerably weaker than the interaction of either protein with WT LTBP1/fibrillin1 (Figure 6.6, 6.7, and 6.8 C). Also the interaction of LTBP1 D1573R with fibrillin R232E appears weaker than the interaction of R232E with WT LTBP, barely giving any measurable response at all in SPR experiments. This supports LTBP1 aspartate 1573 and fibrillin1 arginine 182 being involved in a separate element of the interaction to that involving the fibrillin arginine 232 and LTBP1 aspartate 1655.

These results are consistent with the multi-domain interaction model inferred from SPR and NMR studies. In this model the interaction of the LTBP1-TB3 domain and the fibrillin EGF3/hybrid1 domains, (involving LTBP1 D1573 and fibrillin R182), contributes one binding element to the overall interaction, and the interaction of LTBP1-EGF3 and fibrillin-hybrid1 domains, (involving LTBP1 D1655 and fibrillin R232), contributes another binding element (Figure 6.11). The relative binding levels of the various substituted proteins, with fibrillin R232E and LTBP1 D1655R substitutions having the greatest impact, suggests that the interaction of the LTBP1-EGF3 and fibrillin hyb1 domains makes the largest contribution to the overall affinity. This is consistent with

the levels of interaction seen between overlapping LTBP1 and fibrillin fragments observed in Chapter 5 (Figure 5.3 and 5.4).

6.5 - Interaction of fibrillin mutants with overlapping LTBP1 fragments

To further confirm that the effects of fibrillin residue substitutions are specific to interactions with particular LTBP1 domains, the binding of LTBP1^{cb14TB3} and LTBP1^{E3cb15} fragments to FBN^{e2cb1} with R182E or R232E substitutions, was also investigated (Figure 6.9). In these experiments the interaction of LTBP1^{E3cb15} with the R232E FBN^{e2cb1} is almost entirely eliminated, while LTBP1^{E3cb15} binds the R182E substituted protein at similar levels to WT. In contrast LTBP1^{cb14TB3} gives a similar binding response with both WT and R232E fibrillin (~8 RU with 50 μ M LTBP1^{cb14TB3}), but gives a more limited response in binding fibrillin R182E (~3 RU at 50 μ M). This further supports the multi-domain interaction model (Figure 6.11), with fibrillin R182 residue contributing to the interaction with LTBP1-TB3, and fibrillin R232 involved in the interaction with LTBP1-EGF3.

6.6 – Effect of R232E substitution in NMR titration experiments.

To further examine the interactions of the FBN^{e2cb1} R232E substituted fragment, and confirm the significant reduction in binding was not specific to the SPR assay, an HSQC titration was carried out with ¹⁵N-labelled LTBP1^{E3cb15} and R232E FBN^{e2cb1}, using the same conditions described previously with the WT FBN^{e2cb1} titration (Figure 5.13).

Only very small peak intensity changes were seen with addition of the R232E fibrillin compared with WT fibrillin (Figure 6.10), suggesting the LTBP - fibrillin interaction is disrupted in solution, as well as on the surface of an SPR chip. With the R232E

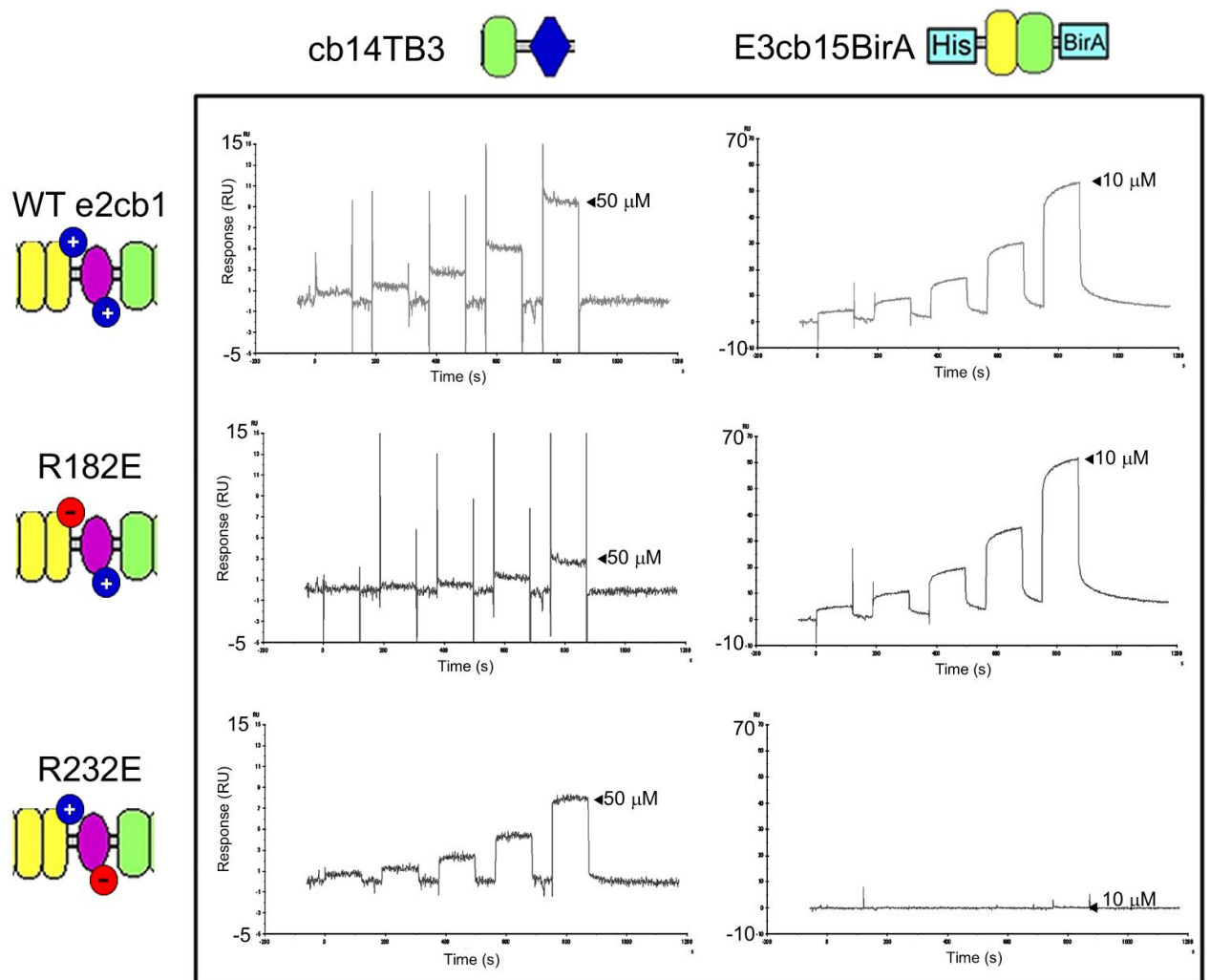


Figure 6.9 - Single cycle SPR experiments with limited LTBP1 fragments

Blank subtracted sensorgrams for LTBP1^{cb14TB3} and LTBP1^{E3cb15} analytes injected across an SPR chip with WT, R182E, and R232E fibrillin proteins bound (same chip as used in figure 6.6 and 6.7). Five concentrations of LTBP1^{cb14TB3} were injected from a 2x serial dilution of 50 μ M, and five concentrations of LTBP1^{E3cb15}-BirA were injected as a 2x serial dilution of 10 μ M.

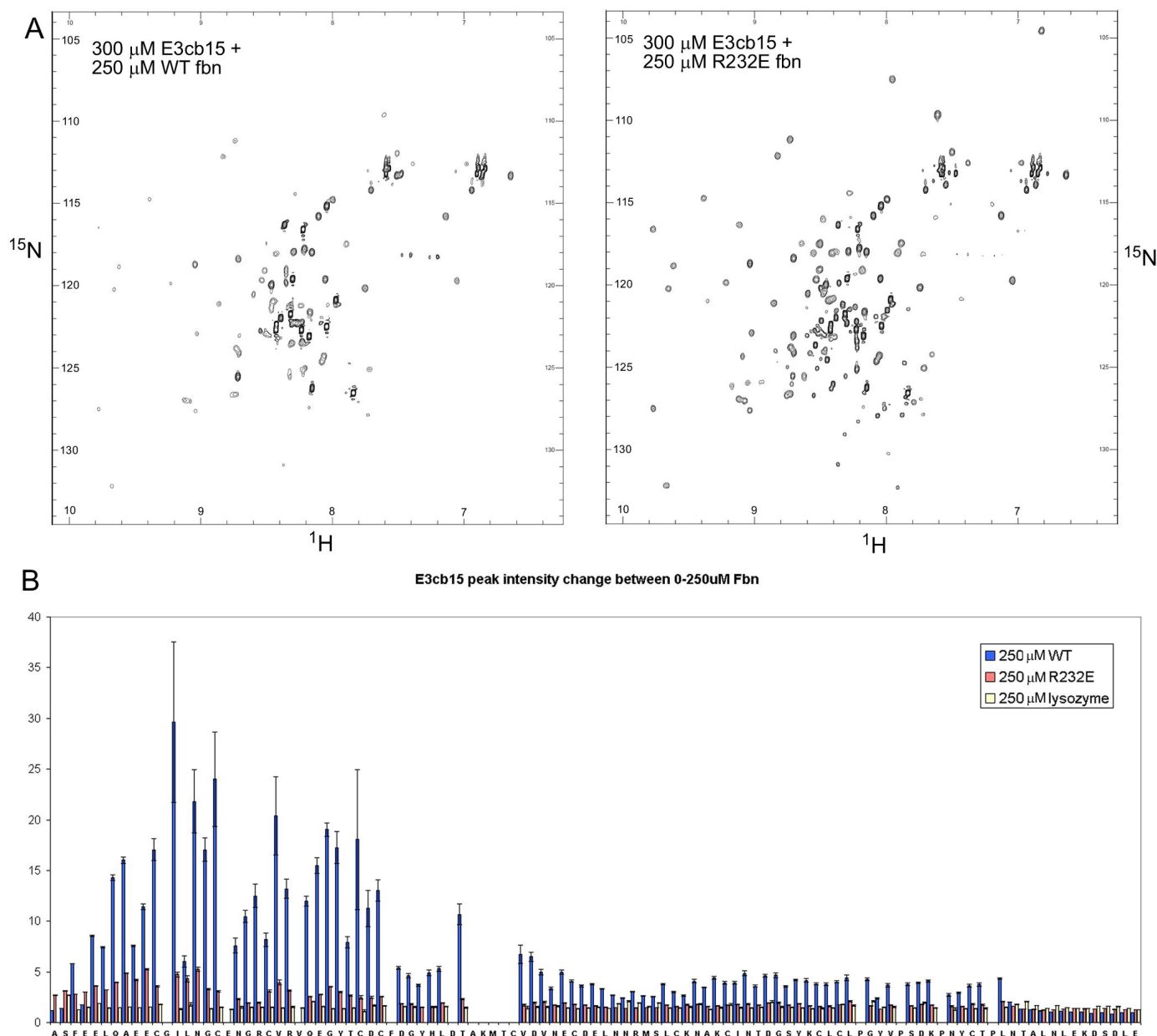


Figure 6.10 - HSQC titration of ^{15}N LTBP1^{E3cb15} with R232E FBN^{e2cb1}

A) HSQC spectra of LTBP1^{E3cb15} with WT FBN^{e2cb1} (left) and R232E FBN^{e2cb1} (right). Both spectra were contoured to identical intensity levels, and both were collected on a 500 MHz spectrometer, under the same conditions (see chapter 5 figure 5.13), initial HSQC spectra of the separate samples used for each titration, without any added protein, looked identical. **B)** Peak intensity change between 0 and 250 μM added protein. Experiments were carried out with 250 μM WT FBN^{e2cb1} (blue bars), 250 μM R232E FBN^{e2cb1} (red), and 250 μM lysozyme (cream). Lysozyme was used as a negative control for increased sample viscosity and non-specific protein-protein interactions. Intensity change measured as "intensity at 0 μM ligand"/ "intensity at 250 μM ligand", with separate 0 μM blank spectra measured for each experiment. Error bars represent standard error determined by intensities of measured peaks relative to background noise (see Section 8.34 for methodology).

titration some minor specific intensity losses were still seen with peaks from the LTBP1 EGF3 domain, which were not seen when 250 μ M lysozyme was added as a control protein (Figure 6.10 B). This suggests some interaction still occurs in solution, where both proteins are present at very high concentration, and so the interaction of LTBP1 EGF3 and fibrillin hyb1 may not be entirely dependent on arginine 232. However the pattern of peaks with the greatest intensity changes appears to be different in the R232E titration, suggesting the binding site of the FBN^{e2cb1} mutant on LTBP1 EGF3 may be subtly altered.

It was hoped that the reduction in binding affinity with R232E may have resulted in peak shifts rather than broadening in the HSQC titration, but this was not seen. It appears then that, despite the significant reduction in binding observed by SPR, in solution state there is still slow to intermediate exchange between the bound and unbound states with LTBP1^{E3cb15} and R232E FBN^{e2cb1}.

6.7 – Conclusions

The SPR data presented here, showing a specific pattern of interactions between WT and substituted LTBP1^{TB3E3} and FBN^{e2cb1} fragments, are consistent with the multi-domain interaction model inferred from the SPR and NMR titration data in Chapter 5. The data also support the importance of specific salt bridges in the LTBP1 - fibrillin1 interaction, which were predicted by HADDOCK simulations. The results of the mutagenesis specifically highlight one preferred structural model for the LTBP1 EGF3 – fibrillin hyb1 interaction (model A), with substitution of LTBP1 aspartate 1655, and fibrillin arginine 232, significantly disrupting the LTBP1 – fibrillin1 interaction. Substitution of LTBP1 glutamate 1625 or 1642 did not have any significant impact however, suggesting interaction models B and C are inaccurate (Figure 6.1). The

moderate reduction in binding seen with the LTBP1 D1573R and fibrillin1 R182E substitutions also supports their contribution to the interaction of the LTBP1 TB3 domain and fibrillin^{e2cb1} as proposed by HADDOCK simulations.

The LTBP1 – fibrillin1 interaction model supported by mutagenesis is summarised in Figure 6.11 along with a simplified table of the key SPR results. Further mutants to test the importance of other potentially interacting residues may also be of future interest, and double mutants in LTBP1 to try and obtain a more complete loss of interaction might be interesting to direct further biological studies.

References

1. Hwang, J.K. and A. Warshel, *Why ion pair reversal by protein engineering is unlikely to succeed*. Nature, 1988. **334**(6179): p. 270-2.

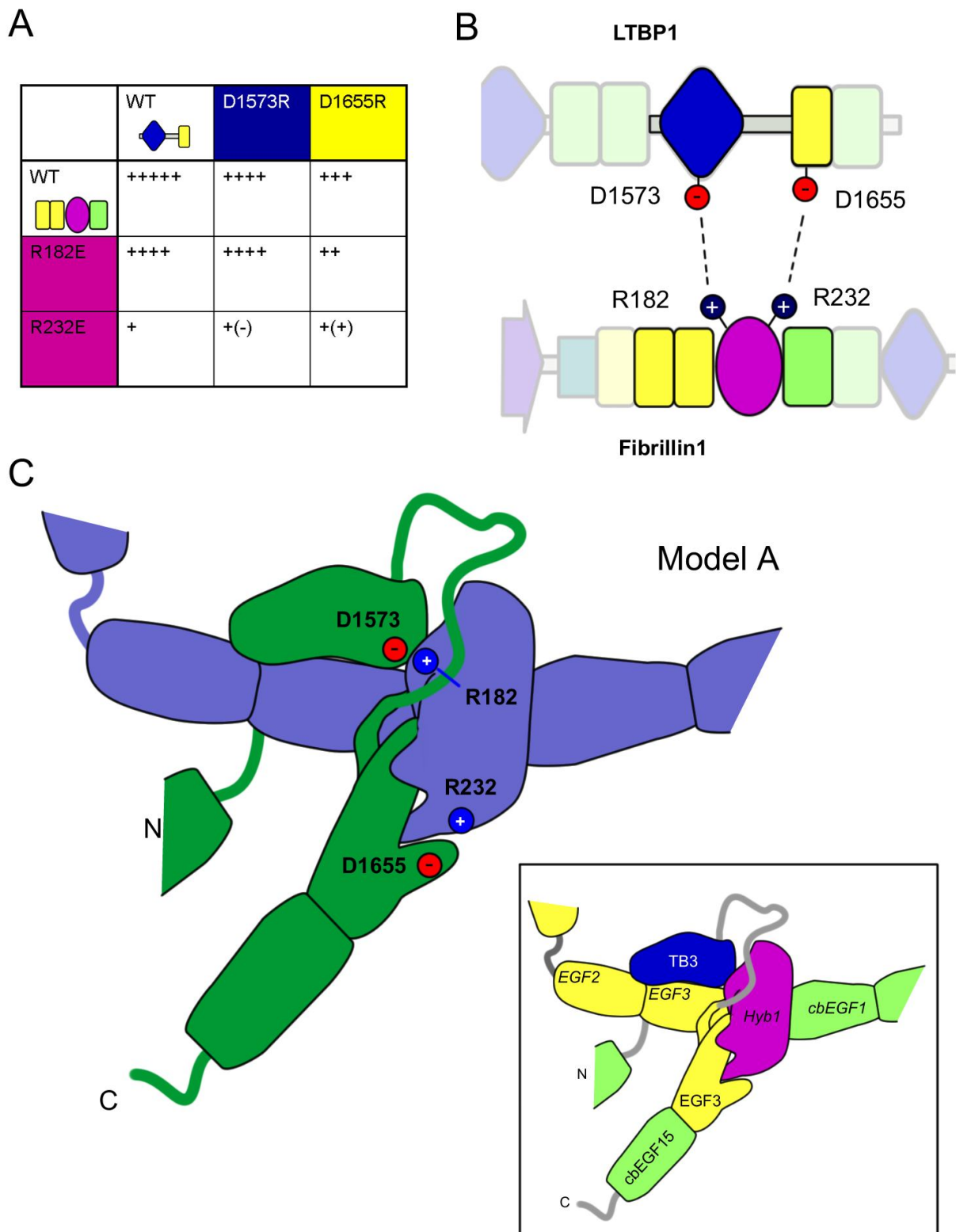


Figure 6.11 - Summary of LTBP1-fibrillin1 interaction model supported by site directed mutagenesis

A) Summary of the main interaction results with key substituted proteins. LTBP1^{TB3E3} substitutions are arranged across the top and FBN^{e2cb1} mutations are arranged down the left hand side. Their level of interaction is given qualitatively as a number of + signs, with the WT- WT interaction having the strongest interaction and the most + signs. The bracketed (+) and (-) signs in the bottom left two cells indicate a slight increase or decrease in measurable binding relative to the R232E-WT interaction. **B)** Linear domain representation of the LTBP1 C-terminus and fibrillin 1 N-terminus, including the two potential salt bridges that are supported by mutagenesis. **C)** Simplified structural model of the LTBP1-fibrillin1 interaction as supported by mutagenesis. In the main image LTBP1 is coloured green and fibrillin1 marine blue, charge interactions supported by mutagenesis are also shown. The smaller image in the bottom right identifies the specific domains shown in the larger image.

Chapter 7 - Discussion

7.1 Summary

This thesis has investigated the structure and evolution of the fibrillin N-terminus and LTBP1 C-terminus, and how these two proteins interact. Initially bio-informatics studies were used to highlight conserved areas of LTBP1 and fibrillin1 of particular interest. Then expression, refolding, and purification conditions were established for a number of novel LTBP1 and fibrillin1 constructs, and, once protein fragments were generated, various biochemical and biophysical experiments were carried out. These revealed a number of novel insights into the structure and interactions of these proteins, which may in turn have further biological implications.

7.2 Structure of the fibrillin N-terminus

In Chapter 3, preliminary NMR structures of the FBN^{e2cb1} multi-domain fragment were generated and other structural studies carried out. Due to limited NOE connections between the N-terminal residues of the hyb1 domain and the rest of this domain, both straight and bent protein structures were generated, hinging near the EGF3 – hyb1 interface region (Figure 3.22). Both conformations were consistent with the available NMR restraints, including RDCs. It is hoped that further experiments will be able to distinguish which is the accurate conformation of FBN^{e2cb1}. However for the purposes of this discussion these preliminary structures, and the NMR data used to generate them, still provide significant insight into how this region of fibrillin is structured.

7.2.1 Domain interfaces

The four domains in the FBN^{e2cb1} fragment all demonstrate folds consistent with their domain types [1-6]. The interfaces between domains are of particular interest as they

can be difficult to predict by homology modelling, but may significantly affect the overall protein structure. Heteronuclear NOE data, looking at fast time scale dynamics, does not demonstrate any flexible interdomain linkers in the FBN^{e2cb1} fragment (Figure 3.9), and specific interdomain NOEs (Figure 3.16) also support all four domains forming interfaces, and the protein forming a single rigid structure. In preliminary NMR structures, the interface between EGF2 and EGF3 is consistent with that observed previously in the structure of the FBN^{NE3} fragment¹ (D. Yadin, personal communication). The interface between hyb1 and cbEGF1 is also closely similar to the interfaces seen between fibrillin1 hyb2-cbEGF10, and TB4-cbEGF23 (Figure 3.26). The small difference in the hyb1-cbEGF1 interface is that isoleucine 202, near the end of the second hyb1 β -strand, appears to pack with the α -helical region of cbEGF1, instead of leucine 199 in the middle of this β -strand, as would have been predicted on the basis of the hyb2-cbEGF10 and TB4-cbEGF23 structures. Alignments of fibrillins from various species show that bulky hydrophobic residues are consistently found at the same position as isoleucine 202 in the hyb1 domain (Figure 2.8), while in other fibrillin TB domains various polar and charged residues can be found in this position (Figure 7.1, green boxed), consistent with a conserved role for Isoleucine 202 in hyb1.

The interface between EGF3 and hyb1 represents a novel packing interaction (Figure 3.24 and 7.1 B), which is quite different to that seen at the interface of cbEGF22 and TB4 (Figure 3.24 C and 7.1 C), and also to the limited packing interaction between cbEGF9 and hyb2 (Figure 1.4). Alignments were generated with other regions of fibrillin1 to examine whether the EGF3-hyb1 interface is unique (Figure 7.1 A), or whether a similar interface might also be found in other cbEGF-TB domain pairs.

¹ This would be expected, as some of the restraints for this region from FBN^{NE3} were included in the FBN^{e2cb1} structure calculations to improve the agreement of this region. Their inclusion did not introduce any significant violations or energy penalties into the calculations.

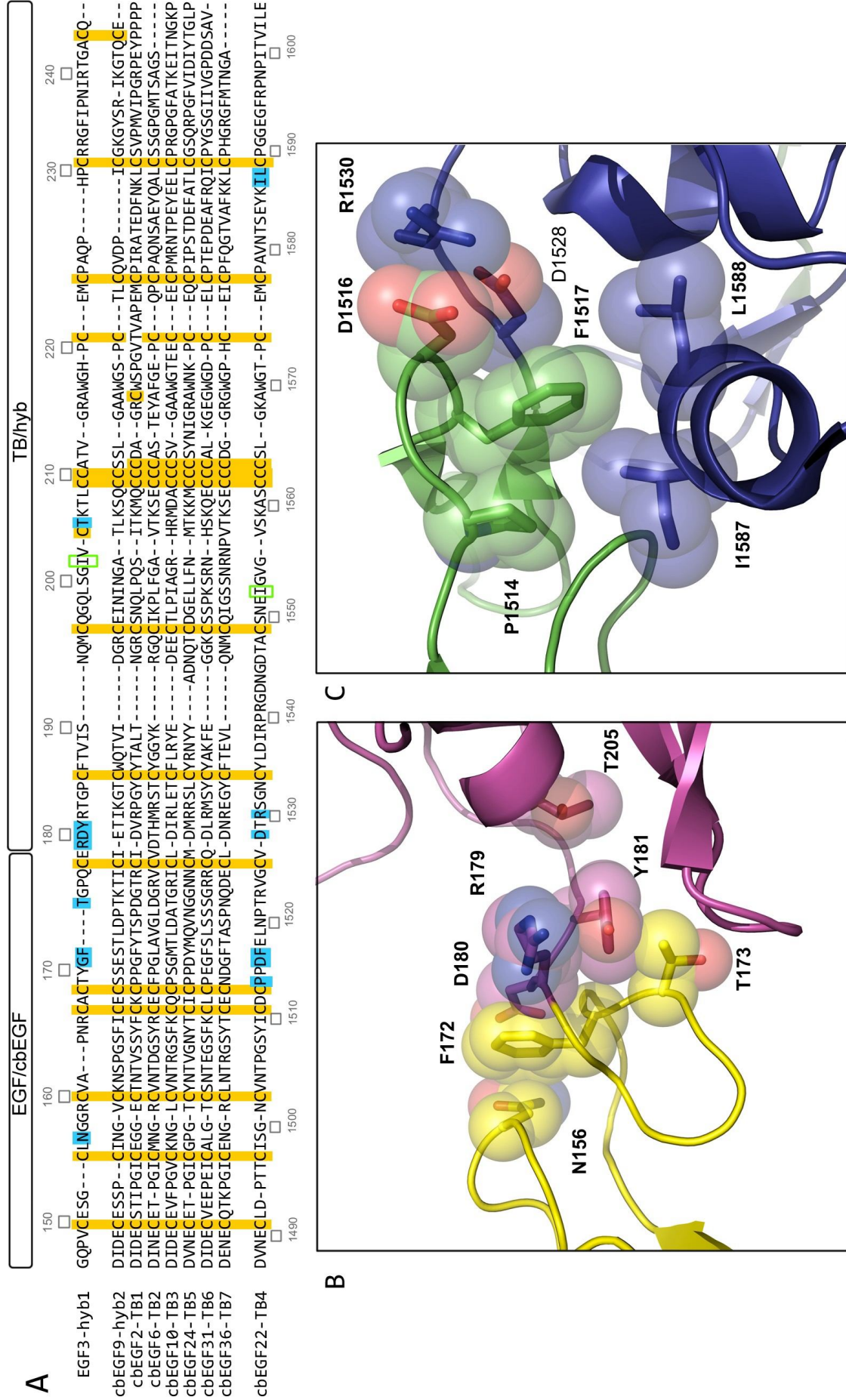


Figure 7.1 - Potentially unique nature of EGF3-hyb1 packing

A) Alignments of human fibrillin1 EGF/cbEGF - TB/hyb domain pairs. Cysteine residues are highlighted with a yellow background, and residues shown in B) and C) as important for domain packing are highlighted with a blue background. The two residues highlighted with green boxes are also discussed in the text as important for packing of this domain with the C-terminal cbEGF. **B)** Interdomain interactions of EGF3-hyb1 (extracted from Figure 3.24) **C)** Interdomain interactions of cbEGF22-TB4 (extracted from Figure 3.24), other interdomain contacts also exist between cbEGF22 and TB4, but for simplicity these are not shown here.

These alignments show that the cbEGF22-TB4 structure may not be representative of all other cbEGF-TB domain pairs (as suggested by Lee et al [2]), as only the cbEGF24-TB5 domain pair also has an acidic residue at an equivalent position to that in the “DF” motif of cbEGF22, which forms a salt bridge with the arginine at the start of the following TB domain. However the EGF3-hyb1 interface also has a number of unique features defined by residues not seen in other domains, such as the inclusion of an extra arginine residue at the start of hyb1, which may affect the coordination of other residues in this region. In addition, no cbEGF-TB domain pairs have an equivalent residue to tyrosine 181, which appears central to EGF3-hyb1 packing. While some cbEGF-TB domain pairs do have hydrophobic residues in a similar position to the tyrosine 181 of hyb1, it is not clear if they would pack in the same way as an aromatic side chain. Hybrid domains also lack the second helix of TB domains that in TB4 provides additional hydrophobic contacts with the N-terminal cbEGF domain, which also appear conserved in other TB domains. Collectively these data suggest the cbEGF22-TB4 structure may still provide the better model for predicting other cbEGF-TB domain interfaces, but the novel elements presented by the EGF3-hyb1 interaction should also be born in mind.

In sequences of EGF3-hyb1 from multiple organisms, the packing residues are highly conserved (Figure 2.8), as are the packing residues involved in the EGF2-EGF3 and hyb1-cbEGF1 interfaces. This suggests that these interdomain interactions are biologically significant enough to be conserved through 600 million years of evolution [7]. This conservation may relate to a role of these interdomain interfaces in the structure or stability of microfibrils.

7.2.2 Position of cysteine 204 in hyb1

One observation from the FBN^{e2cb1} structures, which may have implications for the mechanism of microfibril assembly, is the partially buried position of serine 204 in hyb1. In the wild-type protein there is a cysteine residue in this position, which introduces a 9th cysteine into hyb1 with no disulphide-bonding partner. It has been demonstrated that this cysteine forms a free thiol in fibrillin secreted by eukaryotic cells [8]. Pulse chase labelling of embryonic chick aortae has also demonstrated that fibrillin molecules rapidly form disulphide cross-linked aggregates in the ECM, and it has been proposed that cysteine 204 may play a role in this. This notion is supported by its complete conservation in all hyb1 domain sequences discovered so far (Figure 2.8). Homology modelling of hyb1 based on the structure of hyb2 initially suggested cysteine 204 would be solvent exposed [1], but in the structures determined here, the serine 204 side chain appears to be consistently buried (Figure 3.27), although serine 204 it is still quite close to the protein surface, and may still be accessible to small molecules. Partial protection of cysteine 204 would be consistent with experimental observations where the accessibility of free thiols in N-terminal fragments of fibrillin1 and 2 was measured by labelling by Ellman's reagent [8]. In these experiments it was shown that thiol labelling occurred much more rapidly in 6 M guanidine-HCl than in Tris-buffered saline, suggesting that partial denaturation of the hyb1 structure by guanidine improves reagent accessibility of cysteine 204.

If cysteine 204 occupies the same position as serine 204, some structural rearrangement of hyb1 would be required in order to form disulphide bonds with another protein. This rearrangement may be possible with local movements of the surrounding residues, without any disruption of disulphide bonds. For example, a protein interacting with this region could perhaps encourage the leucine 199 residue in

hyb1 to pack with cbEGF1 rather than isoleucine 202, and this altered packing of hyb1 and cbEGF1 would push the loop containing cysteine 204 outwards, exposing it for disulphide formation.

7.2.3 Context of FBN^{e2cb1} within the fibrillin N-terminus

The structures of FBN^{e2cb1} help to improve our picture of the overall fibrillin microfibril. Together with information from the FBN^{NE3} structure (D. Yadin, personal communication), we can present two potential structural models for the N-terminal region of fibrillin1 (Figure 7.2). As well as being important for predicting the overall architecture of fibrillin1, and giving insights into microfibril organisation, the structure of this region may also provide insight into numerous protein-binding sites (Figure 1.7), which allow fibrillin to perform a regulatory role in the extracellular matrix.

7.3 Structure of the LTBP1 C-terminus

In Chapter 4 the structure of the LTBP1 C-terminus was investigated, by using various NMR methods to study three overlapping LTBP1 constructs. Calcium binding and D₂O exchange experiments support the LTBP1 cbEGF14, TB3, EGF3 and cbEGF15 domains all adopting folds consistent with those predicted by homology modelling and sequence conservation (Figures 4.10, 4.12, 4.13 and 4.14). Unlike the FBN^{e2cb1} fragment however, the C-terminus of LTBP1 does not appear to behave as a single rigid unit, and instead contains numerous flexible regions, specifically between cbEGF14 and TB3, TB3 and EGF3, and the residues at the very C-terminus, (as demonstrated by heteronuclear NOE analysis (Figure 4.15)). These dynamics data, and the lack of significant chemical shifts between overlapping constructs (Figure 4.9), demonstrate a lack of interdomain interfaces between cbEGF14 and TB3, and between TB3 and EGF3. However a canonical aromatic packing interface does

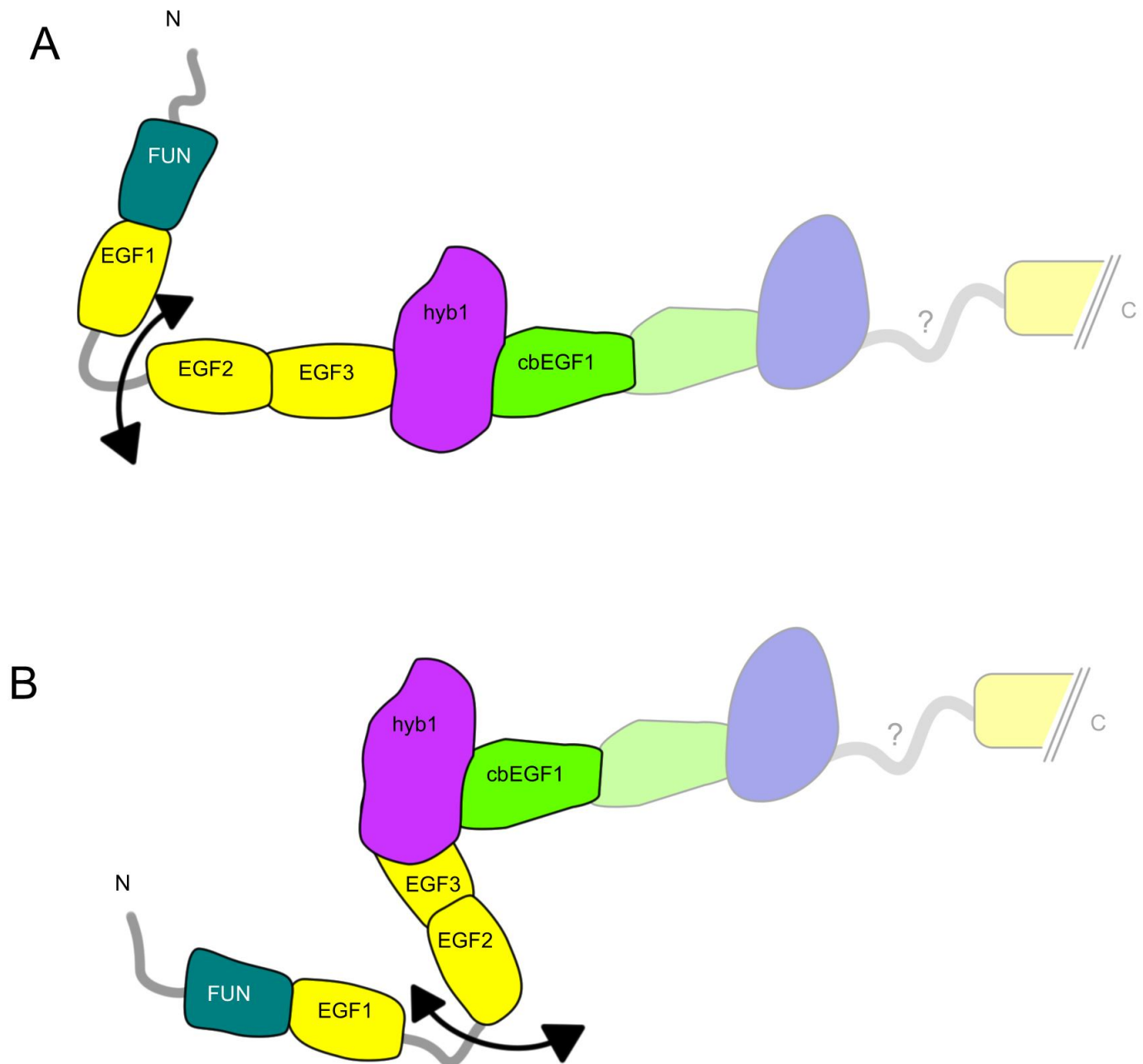


Figure 7.2 - Proposed simplified structures of the fibrillin1 N-terminus

A simplified representation is shown of the N-terminal domains of fibrillin (coloured as in Figure 1.1), cbEGF2, TB1, the proline rich region and EGF4 are also shown semi transparently to provide context, although their structures and interdomain interfaces are unverified. A black double headed arrow signifies the potential for a significant range of motion around the EGF1-EGF2 linker (D Yadin, personal communication). **A)** Summarises the likely structure of the N-terminus of fibrillin if the straight FBN^{e2cb1} NMR structure is used, and **B)** if the bent structure is used.

appear to be present between EGF3 and cbEGF15, as seen in other cbEGF-cbEGF interfaces [5, 6].

Together these experiments imply a “double jointed tail” like behaviour for the structure of the LTBP1 C-terminus, as summarised in Figure 7.3. This flexible structure may have significant implications for the activities of LTBP1 in the extracellular matrix. The long flexible linker between TB3 and EGF3 may explain why this region is sensitive to a number of proteases [9-11], which have in some cases been postulated to affect TGF β signalling [10]. While the precise sequence of the linker between TB3 and EGF3 is poorly conserved (Figure 2.7) a large number of amino acids consistently separate these two domains in all LTBP1-like sequences found, suggesting the flexible interdomain linker separating TB3 and EGF3 may have a biologically significant role, potentially related to its interaction with fibrillin (discussed below).

7.4 Structure and interfaces of the fibrillin1- LTBP1 complex

7.4.1 Multi-domain binding

In Chapter 5, SPR and NMR titration studies both demonstrated interactions between multiple domains of the fibrillin1 N-terminus and LTBP1 C-terminus. Specific binding sites were determined between particular domains, with LTBP1-TB3 binding to fibrillin1-EGF3, (including potential contributions from the flanking fibrillin EGF2 and hyb1 domains), and LTBP1-EGF3 binding to the C-terminal β -hairpin of fibrillin1-hyb1. These binding site data were used to generate potential LTBP1-fibrillin1 complex structures, based on the available homology models, and the HADDOCK simulations carried out suggested a number of potential binding models, of which two were particularly favoured (Figure 5.17).

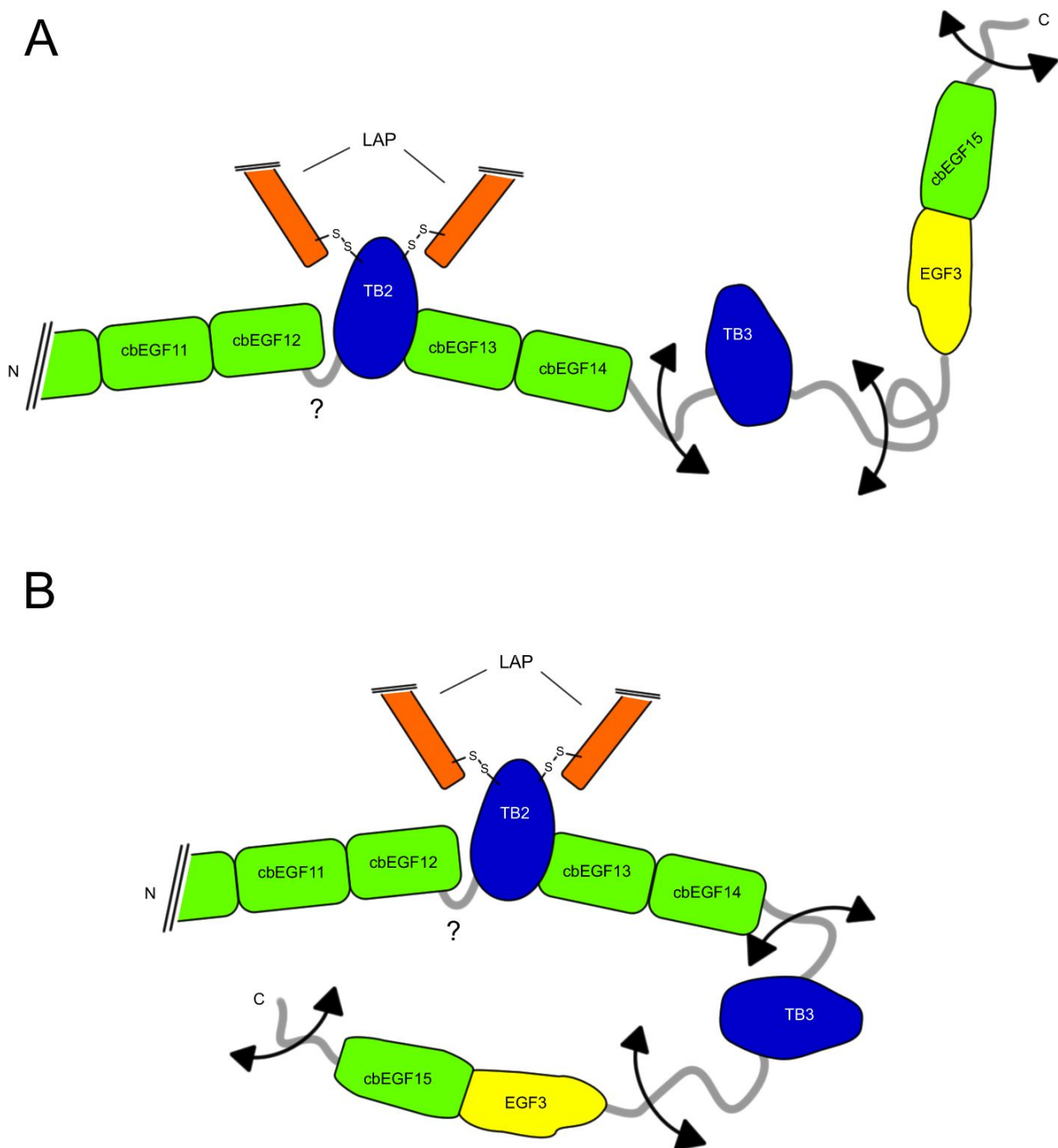


Figure 7.3 - Schematic diagrams of the LTBP1 C-terminus

A) and **B)** show two of many possible domain orientations, with black double-headed arrows denoting linkers where a significant range of motion is possible. The N-termini of the TGF β -propeptide are also included, and shown in orange where they form disulphide bonds with the second TB domain of LTBP1 (for LAP structure see Figure 1.10). Interfaces of cbEGF11-12, TB2-cbEGF13, and cbEGF13-14 are based on conservation of canonical packing residues, but no canonical packing residues are seen at the cbEGF2-TB2 interface (denoted by a question mark).

7.4.2 Confirmation of interfaces by mutagenesis

Site directed mutagenesis was then used in Chapter 6 to test these interaction models, targeting potential salt-bridges by substituting residues involved with residues of opposite charge. Of five LTBP1 mutants and three fibrillin mutants tested in SPR studies, only two LTBP1 substitutions (D1573R and D1655R) and two fibrillin1 substitutions (R182E and R232E) appeared to affect LTBP1-fibrillin1 binding. The greater loss of binding observed with R232E and D1655R substitutions was also consistent with a more significant role of LTBP1-EGF3 binding fibrillin1-hyb1 in the overall interaction. The data showing some substitutions reduce binding while others do not, together with other interaction data presented in Chapter 6, favour one particular model of fibrillin1-LTBP1 complex formation (Figure 6.11), and are not consistent with other potential modes of docking proposed by HADDOCK.

7.4.3 Conserved interface residues

As well as putative salt bridges, the interaction model supported by mutagenesis suggests a number of other polar and hydrophobic contacts between the fibrillin1 and LTBP1 proteins (Figure 7.4 and 7.6). Many of these residues are also well conserved, supporting the hypothesis that they perform a biological role (Figure 7.5 and 7.7). For example tyrosine 1548, in the conserved “TYT” motif in LTBP1 TB3 N-terminal to the core α -helix (Figure 7.4C), along with the core tryptophan 1559, is predicted to interact with a hydrophobic region on the main β -hairpin of fibrillin EGF3, which is partly defined by valine 161 and alanine 167 (Figure 7.4 D). The conserved alanine 1564 - leucine 1565 motif in LTBP1 TB3 also appears to interact with a similar region of fibrillin1 EGF3. The LTBP1 EGF3 domain on the other hand contains many highly conserved residues, several of which have side-chains that pack with the fibrillin1 hyb1 domain.

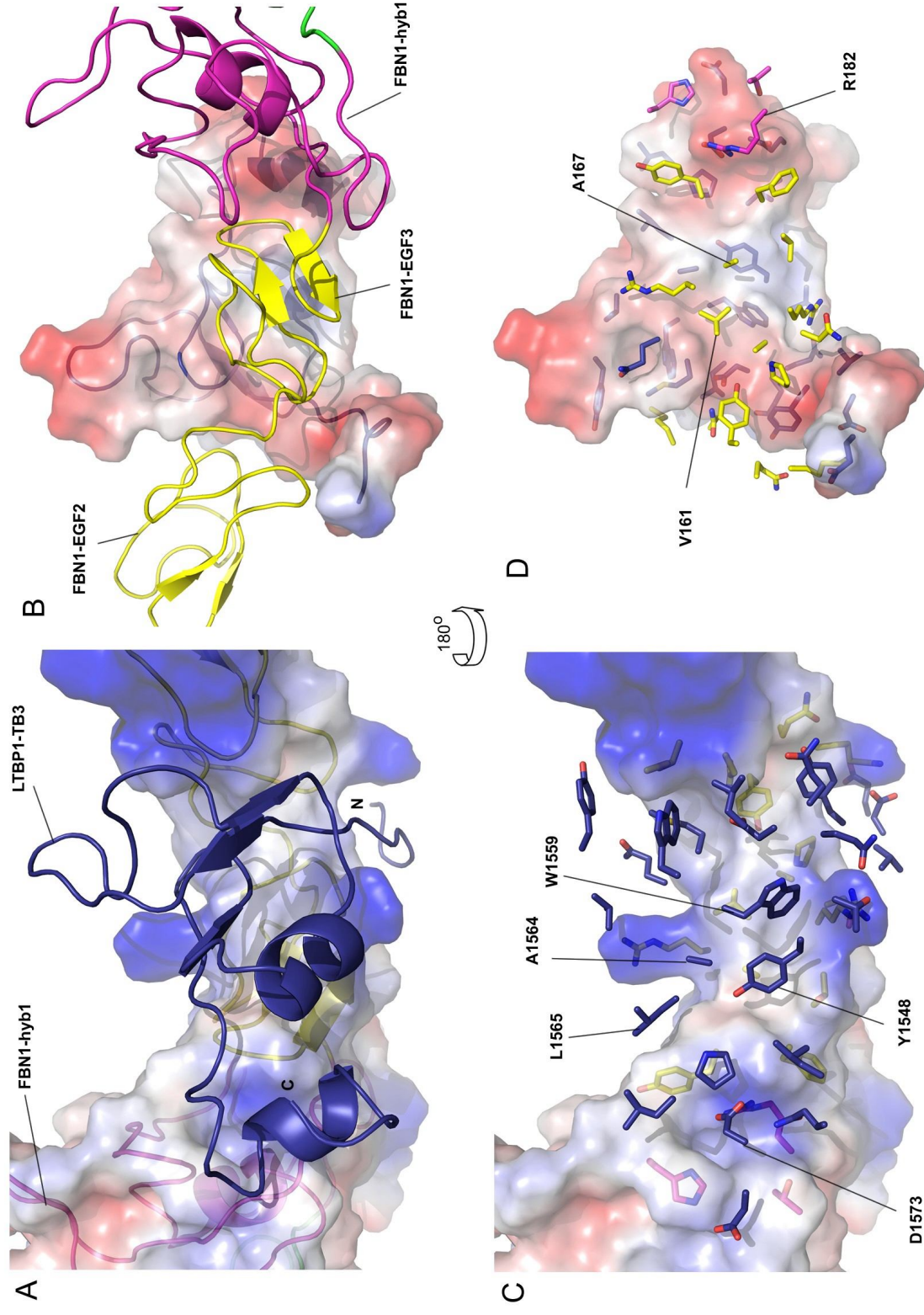


Figure 7.4 - Modelled interaction surfaces between LTPB1 TB3 and fibrillin1 EGF2-3-hyb1

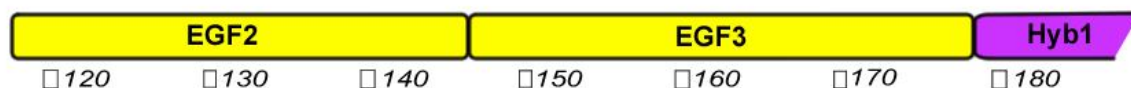
Surfaces of fibrillin (A and C) and LTPB1 TB3 domain (B and D) coloured according to electrostatic contact potential. Red represents negative polarity, blue represents positive polarity, and white represents non-polar regions, (representations of surface charge generating using pymol, which may not be quantitatively accurate). Structure from HADDOCK run 4 cluster1 was used to make this figure (Table 5.2). **A**) LTPB1 TB3 domain represented in cartoon format binding to fibrillin. The surface of the fibrillin molecule is rendered semi-transparently and coloured according to electrostatic potential. **B**) Fibrillin EGF2-3 and hyb1 domains represented in cartoon format and LTPB1 TB3 shown in surface format, same structures used as in A) but rotated 180°. **C**) LTPB1 TB3 residue side chains forming surface contacts with fibrillin. **D**) Fibrillin residue side chains that form surface contacts with LTPB1 TB3. Selected residues in C) and D) are labelled here and all surface residues are highlighted in figure 7.5 above the relevant sequence alignments.

LTBP



Surface residues:	<u>ET</u> <u>DVYQDL</u> W	YV	Q TY	WGMQ AL PL <u>DD</u> QL
Human_LTBP1	- NEQI EETDVYQDL CWEHL SDE- YVCSRPLVG- KQTTYTECCCLYGEA- WGMQCAL CPLKDSDDYAQLCNI PVT-			
Platypus_LTBP1	- NEQTEETDVYQDL CWEHL SNE- YVCSRPI VG- KQTTYTECCCLYGEA- WGMQCAL CPMKDSDDYAQLCNI PVP-			
Chicken_LTBP1	- SEQTEETEVYQDL CWOHL SDD- FVCSRPLVG- KQTTYTECCCLYGEA- WGMQCAL CPMKESDDYAQLCNI PVP-			
Frog_(tropicalis)_LTBP1	- SEPTTEETDVYQDF CWQEL SEE- FVCSQPLVG- RRTTYTECCCLFGEA- WGMQCAL OPTKDSDDYAELCNI - - -			
Zebra_fish_LTBP1	- - - GQSEADLYSGV CWERVTEN- FSCANPI PN- I HTSFTECCCLYGEA- WGNDCAL CPGKNTDSYAQMCNI R- - -			
Stickleback_LTBP1	- EEHEAEQI DYLG I CWQAVTET- MTCTRP L GPNRKT TYTECCCLHGEA- WGMDCAL CPPRNTE DYASMCNV- - -			
Medaka_fish_LTBP1	- NQHESQTD F QGI CWQVTVD T- KMCTRP L GNNRK I TYTECCCLHGEA- WGMDCAL CPPRN TAGFNHLCVQ- - -			
Salmon_LTBP	- AEQHSVQTDYQGI CWQVTVTGTRKTCTKPLAN- RKTTFTECCCLFGEA- WGMDCAL CPLKNTEDYASMCNVPI SE			
Lamprey_LTBP1	- RHEEDAEMLSMCMWRAMGND- FMCSQLI LG- KETTYTECCCQYGEA- WGPDCAL CPERNTDDFAQLCNI PRGG			
Human_LTBP2	- TEDLPDHI HMDI CWKKVT- N- DVCSSEPLRG- HRTTYTECCCQDGEA- WSQQCAL CPPRSSEVYAQLCNVARI E			
Human_LTBP4	- - SLDD- - - NLGV C WQEVGAD- LVCSHPRLD- RQATYTECCCLYGEA- WGMDCAL CPAQDSDDF EALCNVLRPP			
Human_LTBP3	- ESPAERAPERDVCWSQRGED- GMCAGPLAG- PALTFDCCCRQGRG- WGAQCRCP PRGAGSH- - - CPTSQS-			
Lancelet_LTBP	- PPAASYI FEPDRV CWT SVI GPNLF CSNALP- - FNMTRAQCCCRSDVQAWGVDCRLCPKLGTD AFRDL CER I KPG			
Acorn_worm_LTBP	- - - VDTVD TNATDV CWT ELSN- EVVCNSAMDG- PESTYDCCCTAGVG- WGFECRRCP TPNSMEYNEL CSEYK- -			

fibrillin



Surface residues:	MNG	<u>QK</u> Y	H GQ	R VAPNR A YGFT	<u>RT</u>
Human FBN1	HCNI RCMNGGSCSDDHCL CQKGYI GT HCGQPV CESGCLN GGRCA VAPNRCA CTYGF TGPQCERDYRTGPCF				
Mouse FBN1	HCSI RCMNGGSCSDDHCL CQKGYI GT HCGQPV CESGCLN GGRCA VAPNRCA CTYGF TGPQCERDYRTGPCF				
Chicken FBN1	HCNI RCMNGGSCSDDHCL CQKGYT GT HCGQPV CENGCLN GGRCA VAPNRCA CTYGF TGPQCERDYRTGPCF				
Frog FBN1	QCNI RCMNGGT CSDEQCHCQKGYA GAHCGHPV CEQGCLN GGRCA VAPNRCA CTYGF TGPQCERDYRTGPCF				
Tilapia fish FBN1	HCNI RCMNGGT CAEDQCL CQKGYV GNHCGQPV CENGCLN GGRCA VAPNRCA CTYGF TGAQCERDYRTGPCF				
Fugu FBN1	NCNI RCMNGGT CAEDHCL CQKGYI GKHCQGPV CENGCLN GGRCA VAPNRCA CTYGF TGAQCERDYRTGPCF				
Human FBN2	QCSVR CMNGGT CADDHCQ CQKGYI GTYCGQPV CENG CQNGGRCI GPNRCA CVYGF TGPQCERDYRTGPCF				
Human FBN3	GCSVSCMNGGT CRGASCL CQKGYT GTVCGQPI CDRGCHN GGRCI GPNRCA CVYGF MGPQCERDYRTGPCF				
Lamprey FBN	QCNI RCMNGGSCADDHCS CGKGYT GT HCGQPV CDTGCQNGGRCI GPNRCA CVYGF TGPHCERDYRTGPCF				
Sea squirt FBN	RCNMRCMNGGECVDNQRCPY GTY GPHCGQPI CGPGCLN GGRCI AGRCA CVYGF TGMR CERDYRKGPCF				
Sea urchin FBN	PCTTQCLN GGS CQGN SCNC PN GF SGT NCGMPT CNPPCQNGGEC SGGNRCS CPPGF GGSQCQTDYRTGV CY				
Fibrillin pre-LTBP					
Aphid FBN	SCKYPCLN GGT CQGSV CT CRQGYT GENCGDPI CRDGCLN GGRCI GPDRCA CVYGYA GRRCEADYRI GPCY				
Beetle FBN	PCRAF CMNGGT CT DGRCV CAPGYSGEF CTEPI CLEPCKN GGRCI GPDRCA CI YGFTGNHCETDYRTGPCY				
Daphnia FBN	SCP NL CQNGGSCI GK KCV CRPGF SGEA CEDGVCTE PCLN GGRCI GPDRCA CVYGF TGRRCEADYRTGPCF				
Louse FBN	QCRTF CLN GGT CQGSKCI CPPGF SGKHCSEPI CREPCLN GGRCI GHDRCA CPYGYT GRRCEADYRTGPCY				
Sea anemone FBN	RCKI LCFN GGSCTDDT CT CLKGYT GRF CETPVCTE PCLN GGRCT GPDVCA CPYGFAGKRCQSDYRTGPCY				

Figure 7.5 - Alignments of LTBP1 TB3 and fibrillin EGF2-EGF3-hyb1 domains

Alignments adapted from Figures 2.7 and 2.8 and coloured identically. Putative surface residues identified in HADDOCK models are highlighted above the alignments. Residues potentially involved in electrostatic surface contacts or salt bridges are coloured blue. Of these residues those that have been shown by mutagenesis to play a significant role in the interaction are underlined in blue, while those that do not affect binding when substituted are underlined in red.

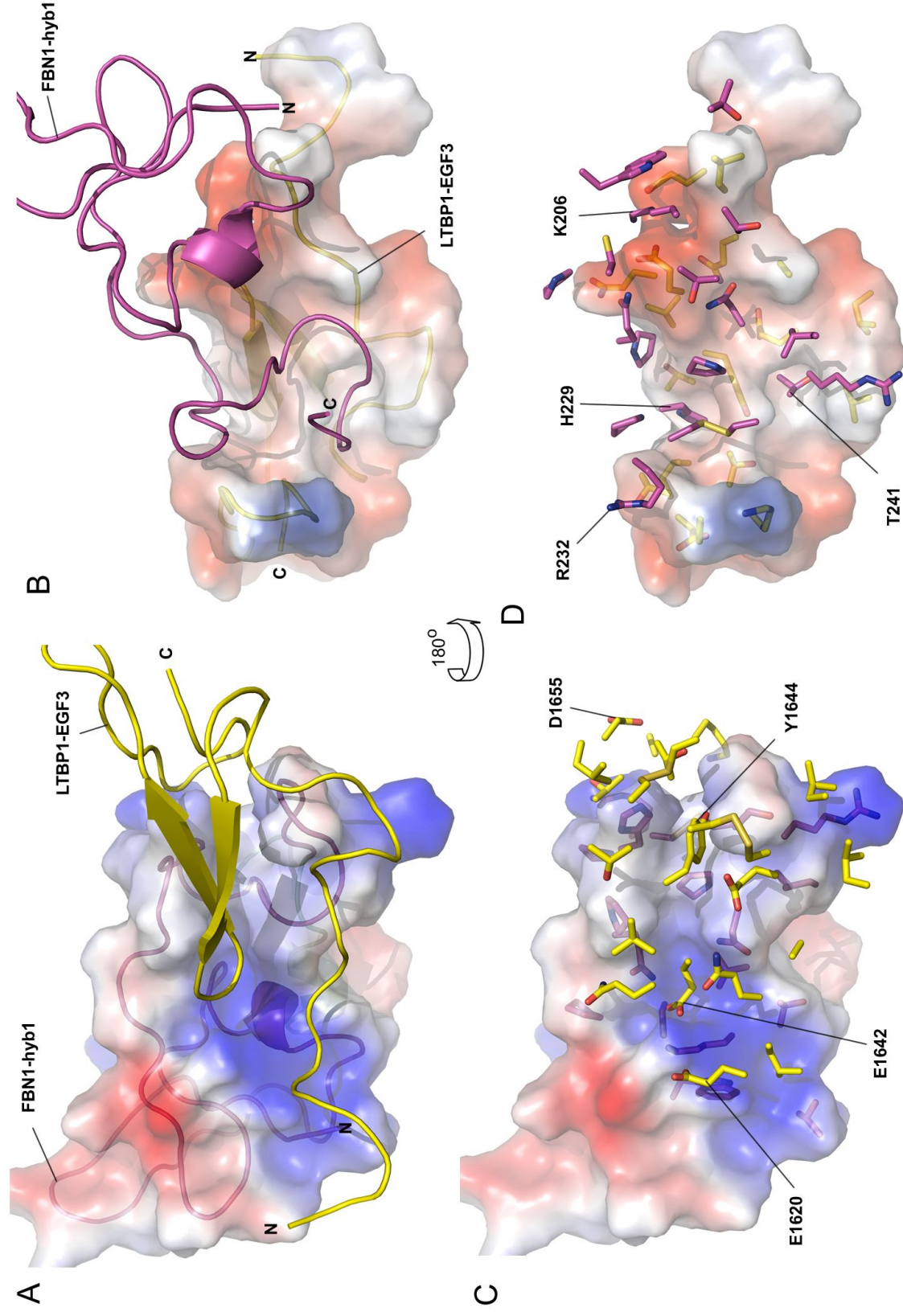


Figure 7.6 - Modelled interaction surfaces between LTBP1 EGF3 and fibrillin1 hyb1

Surfaces of fibrillin (A and C) and LTBP1 EGF3 domain (B and C) coloured according to electrostatic contact potential (as in Figure 7.4). Structure from HADDOCK run 10 cluster6 used to make this figure (Table 5.3). **A)** LTBP1 EGF3 domain represented in cartoon format binding to fibrillin. **B)** Fibrillin hyb1 domain represented in cartoon format binding LTBP1. **C)** LTBP1 EGF3 residue side chains forming surface contacts with fibrillin hyb1. **D)** Fibrillin hyb1 residue side chains that form surface contacts with LTBP1 EGF3. Selected residues in C) and D) are labelled and all surface residues are highlighted in figure 7.7 above the relevant sequence alignments.

LTBP

LTBP

EGF3

Surface residues:

□ 1620

□ 1630

□ 1640

□ 1650

□ 1660

Model A

ELQA

EC

IL

GC

C

VQ

EG

YTC

LD

TAK

T

Human_LTBP1

SFEELQAEECGI LNGCENGRCVRVQEG- YTCDCFDGYHLD TAKMT CV

Platypus_LTBP1

SFEELQAEECGI LNGCENGRCVRVQEG- YTCDCFDGYHLD MT KMT CV

Chicken_LTBP1

SFEELQAEECGI LNGCENGRCVRVQEG- YTCDCFDGYHLD MA KMT CV

Frog_(tropicalis)_LTBP1

SYEELQAEECGI LNGCENGRCVRVQEG- YTCDCFDGYHLD MA KMT CV

Zebra_fish_LTBP1

AFEGLRAEECGV LNGCENGRCVRVQEG- YTCDCFDGYT LDMSRMA CV

Stickleback_LTBP1

AFEGLRAEECGI LNGCENGRCVRVQEG- YTCDCFDGYT LDLSRMA CV

Medaka_fish_LTBP1

AFEGLLAEECGI LNGCENGRCVRVQEG- YTCDCFDGYT LDLSRMA CV

Salmon_LTBP

AFEGLRAEECGI LNGCENGRCVRVQEG- YTCDCFDGYT LDMSRMA CV

Lamprey_LTBP1

SFEGLQAQEECGI LNGCENGRCVRVPEG- YTCDCYDGYQLDL SRMVCT

Human_LTBP2

GFEGQLAEECGI LNGCENGRCVRVREG- YTCDCFEGFQLDAAHMA CV

Human_LTBP4

PYEELEAEECGI LDGCTNGRCVRVPEG- FTCRCFDGYRLDMTRMA CV

Human_LTBP3

DSSEEDSDECR- - - CVSGRCVPRPGG- AVCECPGGFQLDASRARCV

Lancelet_LTBP

YASETLDADCKI EGGCGFGRCLRV DGGGYI CDCYSGYKLDNKG LKCV

Acorn_worm LTBP

DGQEELELDCGFLDGCVNGRCVRVHGG- YACDCYRGFKLDRT GTTCF

fibrillin

fibrillin

Hybrid1

Surface residues:

180

190

200

210

220

230

240

Model A	T	TKT	C	W	P	QHPHCR	PNIRTG	C
Human FBN1	DYRTGPCFTVI -	SNQMCQGGQLSGI	VCTKTL	CCATV	GRAWGHPC	CEMCPAQPHPCRRGFI	PNI RTG- ACQ	
Mouse FBN1	DYRTGPCFTVV-	SNQMCQGGQLSGI	VCTKTL	CCATV	GRAWGHPC	CEMCPAQPHPCRRGFI	PNI RTG- ACQ	
Chicken FBN1	DYRTGPCFTLV-	TNQMCQGGQLSGI	VCTKTL	CCATV	GRAWGHPC	CEMCPAQPHPCRRGFI	PNI RTG- ACQ	
Frog FBN1	DYRTGPCFTVV-	SNQMCQGGQLSGI	VCTKTL	CCATV	GRAWGHPC	CEMCPAQPHPCRRGFI	PNI RTG- ACQ	
Tilapia fish FBN1	DYRTGPCFASV-	NNQMCQGGQLTGI	VCTKTL	CCATV	GRAWGHPC	CEMCPAQPHPCRRGFI	PNHRTG- ACQ	
Fugu FBN1	DYRTGPCFASV-	DNQVCQGGQLTGI	VCTKTL	CCATV	GRAWGHPC	CEMCPAQPNPCRRGFI	PNHRTG- ACQ	
Human FBN2	DYRTGPCFTQV-	NNQMCQGGQLTGI	VCTKTL	CCATI	GRAWGHPC	CEMCPAQPPCRRGFI	PNI RTG- ACQ	
Human FBN3	DYRTGPCFGQV-	GPEGCQHQLTGLV	CTKAL	CCATV	GRAWGLPCEL	CPAQPHPCRRGFI	PNI HTG- ACQ	
Lamprey FBN	DYRTGPCFTQV-	SNQMCQGGQVSGI	VCTKGL	CCATI	GRAWGHPC	CEVCPAQPHPCRRGFI	PNI RTG- ACQ	
Sea squirt FBN	DYRKGPCFHTR-	VNGTCKDEVPGVM	CTKLF	CCATV	GEGWGHPC	CEECPARVGNCTKGRLPPL- - - -	CT	
Sea urchin FBN	DYRTGVCYLQST	SNQYCSAGVEGLHCT	KAL	CCATI	GQGWGDP	CENDTFDNP	CDRGYI	LNGRTG- QCI
Fibrillin pre-LTBP								
Aphid FBN	DYRI GPCYTRV- I	NDMCQGGQLEGVV	VCTKQL	CCATI	GLAWGHPC	CEQCSLS- YDCDI	GFLKNLHSS-	QCI
Beetle FBN	DYRTGPCYTG-	RGALCVNQLEGVV	VCTKTL	CCATV	GKAWGHPC	CEHCPSR- LECEI	GFLKNQKTG-	ECV
Daphnia FBN	DYRTGPCFTKI -	NNDMCQGGQLQGVV	VCTKQL	CCATV	GKAWGHPC	CEGCPVK- LDCDKGFLRNI	HSQ- QCV	
Louse FBN	DYRTGPCYTI V-	KNDMCLGQLEGI	VCTKNL	CCATV	GKAWGHPC	CEGCPPTT- LDCDQGYLKNI	HSQ- ECL	
Sea anemone FBN	DYRTGPCYTVV-	KNRMCHSQLI	GVVCTRNL	CCATI	GQAWGNPC	CERCPLKPAP	CDRGFLPNYKTK-	ACQ

Figure 7.7 - Alignments of LTBP1 EGF3 and fibrillin hyb1 domains

Alignments adapted from Figures 2.7 and 2.8 and coloured identically. Putative surface residues identified in HADDOCK models are highlighted above the alignments. Surface residues from model A are shown, and residues potentially involved in electrostatic interactions or salt bridges are coloured blue. Of these residues, those that have been shown by mutagenesis to play a significant role in the interaction are underlined in blue, while those that do not affect binding when substituted are underlined in red.

One particularly interesting feature is that tyrosine 1644, an aromatic residue in EGF3 that is conserved in all LTBP sequences except LTBP3 (Figure 7.7), seems to pack between fibrillin1 histidine 229 and threonine 241 (Figure 7.6 C and D). While threonine 241 is conserved since the evolution of LBTPs (except in sea squirt), histidine 229 is not, being replaced by asparagine and glutamine in several sequences. However it's possible that these residues could still pack with the LTBP1 core aromatic residues. The consistent presence of a large linker region between LTBP1 TB3 and EGF3 in various organisms (Figure 2.7) may facilitate the activities of both the individual LTBP1 TB3 and EGF3 domains by allowing them to simultaneously interact with their separate binding sites on fibrillin.

The uncertainties surrounding the structural details of the interaction model means the predicted surface residues may not be accurate. However they do provide many potential targets for further mutagenesis studies to test the LTBP1-fibrillin1 interaction model. Determination of a single accurate structure for FBN^{e2cb1}, which could then be used in docking calculations, may also improve the accuracy of the interaction model, (as structural studies in Chapter 3 demonstrate that the novel interface of EGF3 and hyb1 may not be well represented in the homology model). The positioning of fibrillin1 EGF3 relative to hyb1 in the FBN^{e2cb1} structure has particular significance, as it may affect elements of the LTBP1 TB3 binding site, and some domain orientations may partially block the binding of the LTBP1 EGF3 domain to fibrillin hyb1, if binding occurs in the manner proposed by model A. Figure 7.8 shows an overlay of bent and straight FBN^{e2cb1} structures with the docked homology model. The bent structure does not appear to display any particular conflicting residues with LTBP1 EGF3, but when the straight structure of FBN^{e2cb1} is overlaid, the fibrillin-EGF3 domain appears to clash with the flexible N-terminus of LTBP1-EGF3. Repeating HADDOCK docking with a finalised structure of FBN^{e2cb1} would give a more accurate outcome, and if

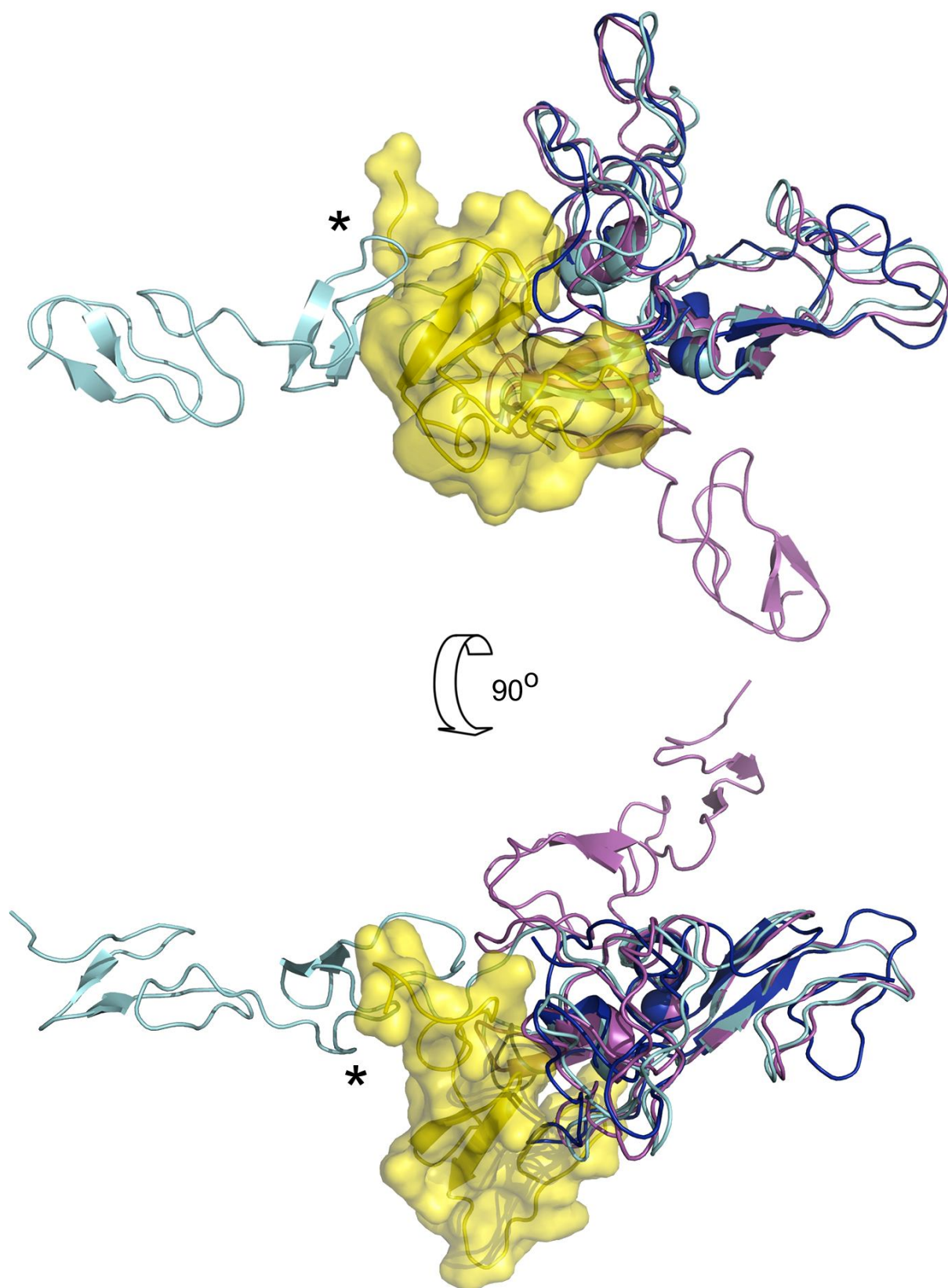


Figure 7.8 - LTBP1 EGF3 docking conflicts with fibrillin EGF3

Three structures are shown overlaid and aligned by their hyb1-cbEGF1 domain pairs. Hyb1-cbEGF1 from docking simulations is shown in dark blue, the straight NMR structure (40-3-3) of FBN^{e2cb1} is shown in light blue, and the bent NMR structure (41-3-13) is coloured magenta. LTBP1 EGF3 is coloured yellow, and an asterisk highlights the potential clash between LTBP1 EGF3 and fibrillin EGF3 in the straight structure of FBN^{e2cb1}.

multibody docking is employed to simultaneously model the interactions of LTBP1 TB3 and EGF3 with FBN^{e2cb1} a near-complete structural model of the LTBP1-fibrillin1 complex may be generated [12].

7.4.4 Context of the fibrillin1-LTBP1 interaction within the matrix

Despite the structural uncertainties that remain, the data generated in this thesis significantly enhance our understanding of the mechanism by which LTBP1 interacts with fibrillin1. The model for how LTBP1 and fibrillin1 may be organised in the ECM is summarised in Figure 7.9. Fibrillin microfibrils form a pre-established rigid scaffold, which the C-terminus of LTBP1 is then able to bind via its TB3 and EGF3 domains. These domains are connected by a sizeable interdomain linker, which may facilitate both interactions occurring simultaneously. It is still unclear whether the linker forms any direct interactions with fibrillin, although the lack of sequence conservation suggests any specific interactions with the linker do not play an important role. It is also unclear which direction the linker may loop around the fibrillin molecule in order to connect LTBP1 TB3 and EGF3 domains, and whether this may protect it from cleavage by extracellular proteases.

Once LTBP1 is bound to fibrillin1, the flexible linker between cbEGF14 and TB3 should still allow a significant range of motion for the rest of the LTBP1 molecule. This may allow the rest of the LTBP1 molecule to interact with other proteins more readily than if LTBP1 was rigid, and unable to move relative to its fibrillin binding site. For example this motion may also allow the LAP-TGF β complex bound to LTBP1 to more readily interact with cell surface $\alpha\beta$ 6 integrin receptors, or interact with other factors bound to fibrillin, like ADAMTSL proteins, which may also play a role in TGF β activation.

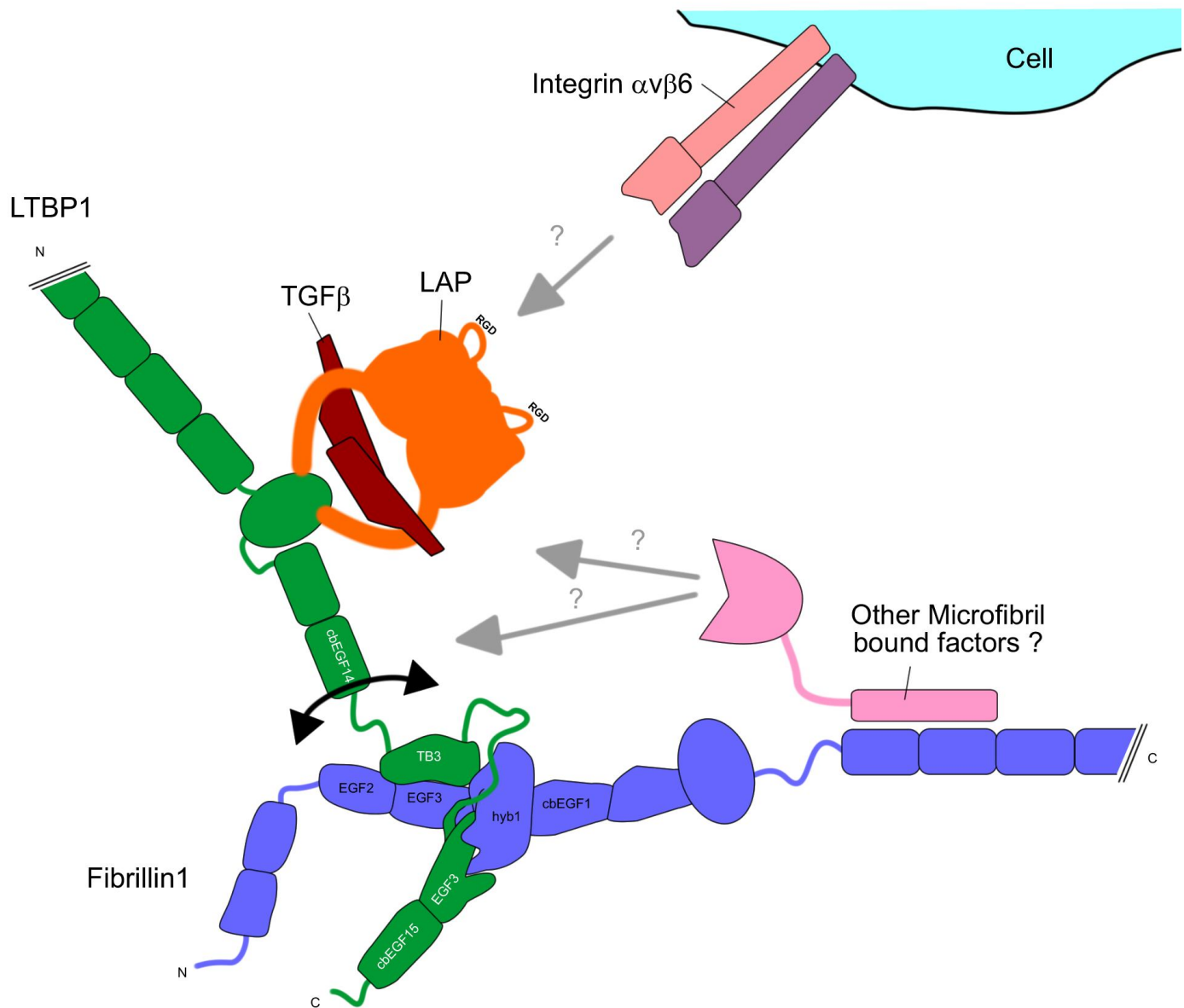


Figure 7.9 - Potential organisation of LTBP1 and fibrillin1 in the extracellular matrix
 LTBP1 is shown in dark green, and fibrillin1 marine blue, the dimeric small latent complex is also shown bound to LTBP, with LAP coloured orange and TGFβ coloured maroon. Black arrows indicate the potential motion of LTBP relative to fibrillin, and grey arrows indicate potential interactions of other factors with LAP or LTBP1.

7.5 Biological implications of the fibrillin1- LTBP1 interaction

There are several hypothesis for the role of the LTBP1-fibrillin1 interaction in TGF β signalling (Section 1.6), and how this might be dysregulated in genetic diseases such as Marfan syndrome. However recently some doubt has been cast on the biological significance of LTBP-fibrillin interactions by studies of mice with a targeted deletion of the hyb1 domain [13]. Mice that are homozygous for this hyb1 deletion show no significant phenotype, suggesting that the hybrid1 domain is non-essential, both for microfibril assembly, and for regulation of growth factors via LTBP. This result also suggests that fibrillin binding to other factors, such as the fibulins (for which the hybrid domain is essential [14]), is also not biologically significant. This result is not conclusive however, as hybrid 1 deletion mice still retain fibrillin2, which may compensate for the loss of the fibrillin1 hyb1 domain. This thesis has shown that the hyb1 domain does not constitute the entire LTBP1 binding site, although it would be surprising if the weak interaction of the fibrillin1 EGF3 region with LTBP1 TB3 alone could compensate for the loss of fibrillin hyb1 and its interaction with LTBP1 EGF3. However, it is also possible that other microfibril bound factors could provide additional contacts with LTBP.

Even with these caveats, the results of these hybrid 1 deletion mice are surprising for several reasons, several human disease mutations exist in the hyb1 domain of fibrillin 1, and while some may cause misfolding and intracellular retention of fibrillin, others, such as a glycine 214 to serine substitution, linked to ectopia lentis [15], and Weill-Marchesani syndrome [16], are not so easily predicted. Also a splice site mutation at the end of exon 6 (the exon consisting exclusively of the hybrid 1 domain) has been reported as a MFS mutation in both UMD and HMGD databases, and in other literature [17]. The patient presenting with this mutation had several features of

classical MFS syndrome, including elongation of the long bones, scoliosis, high arched palate, aortic root enlargement and aortic valve prolapse (Professor Catherine Boileau, personal communication). Details of the exact biochemical outcome of this mutation have not been investigated, but similar splice site mutations in other exons of fibrillin 1 can cause exon skipping [18]. If this is the case with this exon 6 mutation, the expressed fibrillin1 protein should be equivalent to that expressed by the *hyb1* deletion mouse. Alternatively though a cryptic splice site may be used instead, which could disrupt translation and result in nonsense mediated decay of fibrillin mRNA.

As well being the subject of disease mutations in humans, the *hyb1* domain is also highly conserved, and is present in all complete fibrillin sequences known (Chapter 2 Figure 2.1 and 2.8). This surprising lack of phenotype in the *hyb1* deletion mice should then be the subject of particularly close scrutiny. While the genomic deletion of *hyb1* in these mice has reportedly been confirmed by PCR [13], at present no data have been published that unequivocally demonstrates whether all fibrillin1 protein produced in these mice lacks the *hyb1* domain.

If these mice do indeed prove to have no WT fibrillin1, the continued presence of fibrillin2, and a partial LTBP1 binding site in fibrillin1, still indicates that a biological role for the LTBP – fibrillin interactions cannot be ruled out. However it would be strong evidence against the “localisation” model for MFS, whereby a threshold number of binding sites is needed for proper TGF β latency (Figure 1.15). The lethality of FBN1 null mice clearly demonstrates that fibrillin2 cannot compensate for the critical functions of fibrillin1, even if it can provide a limited number of LTBP binding sites.

Without a clear model for how the interaction of LTBP1 and fibrillin1 regulates TGF β signalling *in vivo* it is difficult to interpret how the structure of the fibrillin1 – LTBP1 complex might be adapted to perform its role. However one established mechanism for TGF β activation is traction-mediated activation by integrins (Figure 1.14). In cell culture assays this has been shown to be dependent on LTBP1 anchoring TGF β to an insoluble fibronectin network [19, 20], but it has also been shown in separate cell culture studies that over time LTBP1 loses its co-localisation with fibronectin, and instead co-localises with LTBP4 and fibrillin [21]. It is possible then that LTBP1 interactions with fibrillin are important to act as an anchor for traction-mediated TGF β activation by integrins, in place of fibronectin in mature ECM. With this in mind, the interaction model proposed in this thesis might be considered to loosely resemble a “grappling hook”; specifically the main β -hairpin of the LTBP1 EGF3 domain may act as the “hook”, and the fibrillin hyb1 “PHP” loop and C-terminal β -hairpin as the “purchase” or “ledge” that the “hook” can catch on to (Figure 7.10). The flexible linkers between LTBP1 TB3 and EGF3, and between cbEGF14 and TB3, then serve as the rope-like connections that transmit the force imposed on the rest of the LTBP1 molecule. The role of the LTBP1 TB3 domain may be not to provide a high affinity interaction, but to help organise the “rope”, and the direction of force transmission, by binding to the opposite face of fibrillin.

In this context it could be postulated that genetic diseases affecting the fibrillin matrix, such as MFS, may perturb this traction dependent TGF β activation. MFS mutations may not directly interfere with the binding sites of LTBP, or reduce those sites below a critical threshold, but instead alter overall biomechanics of the fibrillin network to which LTBP is anchored. This could in turn have significant and unpredictable affects on the intensity and localisation of TGF β signals, depending on the resistance provided to

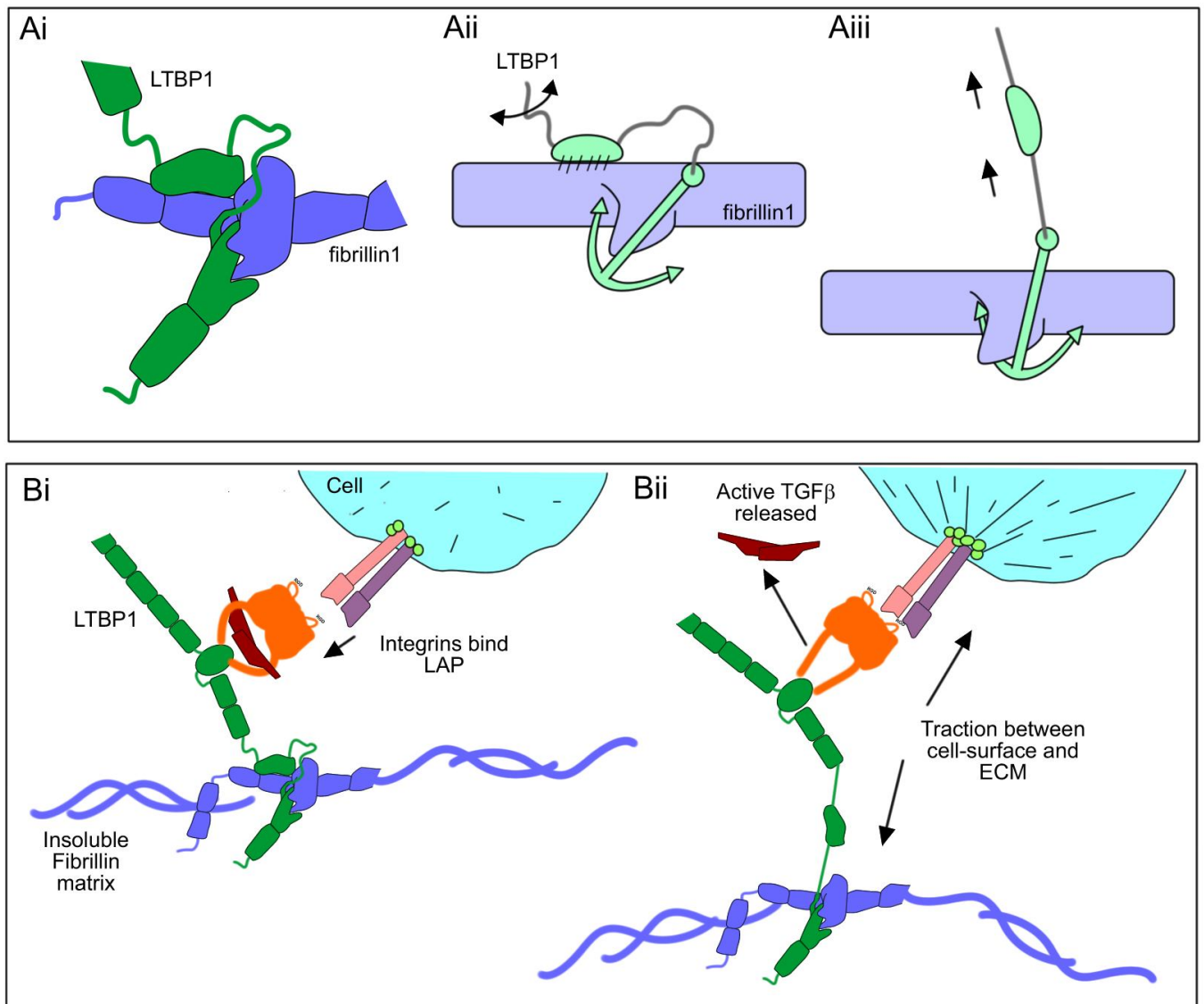


Figure 7.10 - Speculative "Grappling hook" model for the functional significance of the LTBP1 - fibrillin1 complex

Ai) Summary of the structural model of the LTBP1 Fibrillin1 interaction as summarised in Chapter 6 Figure 6.11, LTBP1 coloured dark green and fibrillin1 in marine blue. **Aii)** Interaction model visualised as a "grappling hook" arrangement, with LTBP1 EGF3 performing the role of the "hook", the interdomain linkers the role of a connecting "rope". The LTBP1 TB3 domain also forms contacts with fibrillin that may coordinate the rest of the LTBP molecule in order to provide tension in the correct direction to engage the "hook". **Aiii)** Proposed response of the LTBP1-fibrillin1 complex to tension, such as that imposed by integrin-mediated traction summarised in B). **Bi)** Resting state of LTBP1-fibrillin1-TGF β complex (coloured as in Figure 7.9), in which LAP is exposed for binding by cell surface $\alpha\text{v}\beta 6$ integrins. **Bii)** The anchoring of LTBP to the insoluble fibrillin matrix means that when traction occurs between cells and the ECM, this force is transmitted to LAP, where it has the potential to drive structural changes in LAP that release active TGF β .

integrin based traction. Other fibrillinopathies such as SSS, AD and GD, may disrupt the biomechanics of the fibrillin matrix in a manner distinct from MFS, but which may still cause gross increases TGF β activation apparently similar to MFS, but *in vivo* specific differences in structure and organisation of the fibrillin network may cause discrete changes in the timing or localisation of TGF β signals, which then have significant developmental outcomes.

This functional model is of course highly speculative and it may just be that the multi-domain interaction of LTBP1 and fibrillin1 was the best solution that evolved for specific microfibril localisation of LTBP1. Alternatively some other functional adaptation may become apparent with further understanding of the full biological significance of the LTBP1-fibrillin1 complex.

7.6 Future directions

Collecting further data for determination of a single and accurate NMR structure for the FBN^{e2cb1} fragment of fibrillin is a priority for improving the structural model of the LTBP1-fibrillin1. A structure of FBN^{e2cb1} will also provide valuable information for the mapping of other protein interactions in this region, as well as general insights into the structure of the fibrillin N-terminus. It is hoped that collection of RDCs in alternative alignment media may be able to provide restraints necessary to complete the determination of the FBN^{e2cb1} NMR structure, but if this is unsuccessful small angle X-ray scattering could be employed to examine overall protein shape. Further crystallography could also be attempted, but this might have a better chance of success if FBN^{e2cb1} was re-cloned for expression without the flexible N-terminal residues. The addition of extra domains may also be helpful, if high expression and refolding yields can be maintained. For example a fibrillin1 construct starting with

histidine 118 at the EGF2 N-terminus, running through to valine 391 at the end of the first fibrillin TB domain, might make a good candidate for crystallisation. This is because all the domains in a FBN EGF2-TB1 construct would be predicted to form rigid domain interfaces, and a crystal structure of this region would inform on the structures and relative orientations of cbEGF1, cbEGF2 and TB1 domains as well as the FBN^{e2cb1} region.

In addition to further structural studies, more site-directed mutagenesis experiments may also help reinforce the LTBP1-fibrillin1 interaction model. Double and triple mutants of the LTBP1^{TB3E3} fragment that demonstrate minimal or no interaction with fibrillin, would further support the model, and also help direct other studies, such as generation of a knock-in mouse (discussed below).

A clearer understanding of the kinetics and thermodynamics that underlie the LTBP1-fibrillin1 interaction may also be of interest, and using other techniques to study binding. For example isothermal titration calorimetry or analytical ultracentrifugation could be used to measure K_d in solution, which could be informative, especially if looking at competition or binding-cooperativity with other proteins such as the fibulins. Exploring alternative SPR strategies, such as coupling biotinylated FBN^{e2cb1} to a streptavidin coated chip surface, in the hope of improving binding site presentation and surface homogeneity, may also improve measurements of K_d , if binding consistent with the Langmuir model can be obtained.

Aside from experiments to test and improve the structural model proposed here, there is a major need to improve the understanding of how the fibrillin1-LTBP1 interaction regulates TGF β signalling. Some previous studies have shown eukaryotically

expressed proteins that compete with the LTBP1-fibrillin1 interaction, such as MAGP1 and specific fibrillin1 fragments, can increase pSMAD2/3 levels (a downstream marker of TGF β signalling) when added to cultured cells [22, 23]. This suggests increased TGF β signalling, and supports the LTBP1-fibrillin1 interaction being important for maintaining proper TGF β latency. Repeating these experiments with prokaryotically expressed LTBP1^{TB3E3} and FBN^{e2cb1} fragments, and using binding-deficient mutants as controls, may provide important confirmation of these results, as using prokaryotically expressed protein should avoid the risk of TGF β contamination from eukaryotic protein preparations. Other cell culture experiments using the LTBP1 and fibrillin1 fragments generated in this thesis could also look for effects on LTBP-fibrillin co-localisation, and for effects on TGF β activation in cells derived from Marfan patients, or fibrillin deficient mouse models.

To better understand the role of the LTBP1-fibrillin1 interaction *in vivo*, further information on the hyb1 deletion mouse model would be highly valuable, including crosses with fibrillin2 knock-out mice, and sequencing of genomic DNA or RT-PCR products to confirm hyb1 deletion. Depending on the results of these studies it may also be informative to generate a mouse model in which the C-terminus of LTBP1 is deleted, or where residues crucial for its interaction with fibrillin are substituted by generation of a knock-in mouse. If these mice displayed no phenotype it would be confirmation that the interaction of LTBP1 C-terminus with fibrillin was not biologically significant, despite the extensive conservation of interface residues. However if a phenotype is determined it will make for an interesting comparison with MFS and other mouse models.

If it can be demonstrated that the fibrillin1-LTBP1 interaction is important for TGF β activation by integrins, or other traction mechanisms, it may be of interest to test the importance of the “grappling hook” mode of interaction. Carefully designed atomic force microscopy experiments may be able to examine whether the LTBP1-fibrillin1 interaction is able to resist the traction forces needed for integrin driven unfolding of LAP [24]. If it is possible to specifically bind LTBP1 and fibrillin1 proteins to the AFM tip and a AFM substrate it may even be possible to test the effects of force direction on disrupting the LTBP1-fibrillin1 interaction. This could then illuminate whether the LTBP1 C-terminus genuinely behaves like a “grappling hook”, or just anchors LTBP1 to fibrillin by high binding affinity interactions.

7.7 Concluding remarks

While many questions still remain unanswered over the details and significance of the LTBP1-fibrillin1 interaction, this thesis has advanced our understanding of the individual components of the LTBP1-fibrillin1 complex, and proposed a testable model for their interaction. This information may be of significant help when trying to postulate the mechanism of TGF β regulation by fibrillin, and also potentially be useful for understanding other TGF β independent roles of both proteins.

The work carried out in this thesis also represents a further validation of the dissection methodology for gaining detailed structural understanding of proteins, and protein complexes, that are not amenable to X-ray crystallography. This was achieved here by expressing a number of overlapping protein fragments, and confirming their successful *in vitro* folding, then using a combination of bioinformatics, NMR techniques, SPR, and site directed mutagenesis to investigate their specific structure and interactions.

This structural understanding of the LTBP1-fibrillin1 complex will also help guide further experiments investigating the biology of fibrillins and LTBPs, and the recombinant proteins generated in this thesis may provide useful tools to these ends. Whether targeting the LTBP1-fibrillin1 interaction may have therapeutic potential remains to be determined, and should be informed by a better understanding of the relationship between fibrillin and TGF β signalling. However if disrupting this interaction does prove to be a valid therapeutic target, then the structural information provided here could be highly valuable, for example the knowledge of binding sites could be used to develop monoclonal antibodies specific to the interacting regions of LTBP1 and/or fibrillin1. These may then have the potential to disrupt or enhance TGF β signalling (depending on which is the correct activation model), and the ability to specifically target LTBP/fibrillin-mediated activation may have advantages over direct targeting of TGF β and its downstream effectors.

References

1. Jensen, S.A., et al., *Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins*. Structure, 2009. **17**(5): p. 759-68.
2. Lee, S.S., et al., *Structure of the integrin binding fragment from fibrillin-1 gives new insights into microfibril organization*. Structure, 2004. **12**(4): p. 717-29.
3. Yuan, X., et al., *Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils*. EMBO J, 1997. **16**(22): p. 6659-66.
4. Lack, J., et al., *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, 2003. **334**(2): p. 281-91.
5. Smallridge, R.S., et al., *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, 2003. **278**(14): p. 12199-206.
6. Downing, A.K., et al., *Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders*. Cell, 1996. **85**(4): p. 597-605.
7. Robertson, I., S. Jensen, and P. Handford, *TB domain proteins: evolutionary insights into the multifaceted roles of fibrillins and LTBPs*. Biochem J, 2011. **433**(2): p. 263-76.
8. Reinhardt, D.P., et al., *Initial steps in assembly of microfibrils. Formation of disulfide-cross-linked multimers containing fibrillin-1*. J Biol Chem, 2000. **275**(3): p. 2205-10.
9. Annes, J.P., J.S. Munger, and D.B. Rifkin, *Making sense of latent TGFbeta activation*. J Cell Sci, 2003. **116**(Pt 2): p. 217-24.
10. Ge, G. and D.S. Greenspan, *BMP1 controls TGFbeta1 activation via cleavage of latent TGFbeta-binding protein*. J Cell Biol, 2006. **175**(1): p. 111-20.

11. Tatti, O., et al., *MT1-MMP releases latent TGF-beta1 from endothelial cell extracellular matrix via proteolytic processing of LTBP-1*. Exp Cell Res, 2008. **314**(13): p. 2501-14.
12. Karaca, E., et al., *Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server*. Mol Cell Proteomics, 2010. **9**(8): p. 1784-94.
13. Charbonneau, N.L., et al., *In vivo studies of mutant fibrillin-1 microfibrils*. J Biol Chem.
14. Ono, R.N., et al., *Latent transforming growth factor beta-binding proteins and fibulins compete for fibrillin-1 and exhibit exquisite specificities in binding sites*. J Biol Chem, 2009. **284**(25): p. 16872-81.
15. Sui, R.F., et al., *[Novel mutation of fibrillin 1 gene cause ectopia lentis in a Chinese family]*. Zhonghua Yan Ke Za Zhi, 2004. **40**(12): p. 828-31.
16. De Backer, J., et al., *Utility of molecular analyses in the exploration of extreme intrafamilial variability in the Marfan syndrome*. Clin Genet, 2007. **72**(3): p. 188-98.
17. Matyas, G., et al., *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, 2002. **19**(4): p. 443-56.
18. Evangelisti, L., et al., *A single heterozygous nucleotide substitution displays two different altered mechanisms in the FBN1 gene of five Italian Marfan patients*. Eur J Med Genet. **53**(5): p. 299-302.
19. Annes, J.P., et al., *Integrin alphaVbeta6-mediated activation of latent TGF-beta requires the latent TGF-beta binding protein-1*. J Cell Biol, 2004. **165**(5): p. 723-34.
20. Fontana, L., et al., *Fibronectin is required for integrin alphavbeta6-mediated activation of latent TGF-beta complexes containing LTBP-1*. FASEB J, 2005. **19**(13): p. 1798-808.
21. Dallas, S.L., et al., *Fibronectin regulates latent transforming growth factor-beta (TGF beta) by controlling matrix assembly of latent TGF beta-binding protein-1*. J Biol Chem, 2005. **280**(19): p. 18871-80.
22. Chaudhry, S.S., et al., *Fibrillin-1 regulates the bioavailability of TGFbeta1*. J Cell Biol, 2007. **176**(3): p. 355-67.
23. Massam-Wu, T., et al., *Assembly of fibrillin microfibrils governs extracellular deposition of latent TGF beta*. J Cell Sci. **123**(Pt 17): p. 3006-18.
24. Buscemi, L., et al., *The single-molecule mechanics of the latent TGF-beta1 complex*. Curr Biol. **21**(24): p. 2046-54.

Chapter 8 - Materials and Methods

Bioinformatics and other *In silico* techniques

8.1 Sequence discovery

Fibrillin and LTBP like sequences were detected using basic local alignment search tool (BLAST) algorithms against protein (blastp) and nucleotide (tblastn). Specific libraries used were; the National Center for Biotechnology Information (ncbi) and their online BLAST tools (<http://blast.ncbi.nlm.nih.gov/>), which was used to search broad taxa. Also the genomes of specific organisms, which were only available through the Joint Genome Institute (www.jgi.doe.gov), and Ensembl genome browser (<http://www.ensembl.org>), were also searched with the available online BLAST tools. The amino acid sequence of the human fibrillin1 TB3 domain was primarily used as the search query (as discussed in chapter 2 longer queries with multiple cbEGF domains often gave misleading results), additional query sequences were often employed if no LTBP or fibrillin like proteins were detected initially, such as the human fibrillin1 hyb2-cbEGF10-TB3 domain triplet, which detected the Acorn worm fibrillin partial sequence. The reference sequences detected used in this thesis are summarised in Appendix 4.

8.2 Domain analyses

When a putative fibrillin or LTBP like sequence was discovered, before it was taken forward for further analysis, initial domain annotation was carried out using the “SMART” domain prediction server (<http://smart.embl-heidelberg.de/>) [1]. Where unusual regions of sequence lacking domains were predicted, or the domain structure was inconsistent with the expected fibrillin or LTBP domain structure, re-prediction of the open reading frames from the genomic sequence was attempted using the WISE2 online server (<http://www.ebi.ac.uk/Tools/Wise2/>), using the most closely related

fibrillin or LTBP protein sequence as the search query. These WISE2 re-predictions could sometimes give significant improvements to the completeness of a protein sequence, although they were limited in other cases by the quality and/or size of the genomic sequence scaffolds available.

Once confirmed as fibrillin or LTBP like proteins sequences were then incorporated into alignments, and also used to construct figure 2.1 with further manual sequence inspection and annotation.

8.3 Alignments

In order to examine the conservation of specific regions, and also to construct phylogenetic trees, sequences were aligned using the MEGA5 package of programs [2]. For the LTBP C-terminus and fibrillin N-terminus alignments were generated with the ClustalW program within MEGA 5; Gap opening penalty 5, gap extension penalty 0.5, BLOSUM matrix (residue specific and hydrophobic penalties on, gap separation distance 4, end gap separation off, negative matrix off, delay divergent cut off 30%), Sequences that did not have complete coverage of the regions of interest, or contained potential splicing prediction errors, were not included. Alignments were also manually edited to improve alignment of cysteine residues and also remove potentially bad exon predictions. To generate appropriately coloured figures from the alignments, data was exported from MEGA 5 in FASTA format and opened and annotated using the Pfaat alignment viewer (<http://pfaat.sourceforge.net/>) [3]. Residues were coloured as described in the figure legends, with a chart below the alignment to quantify the level of conservation based on the BLOSUM 62 matrix.

8.4 Phylogenetic trees

Phylogenetic trees were constructed using MEGA 5 from the edited alignments. Numerous trees were constructed from each data set using both Maximum likelihood and Minimum evolution methods. Those trees that seemed most representative of protein evolution were selected for inclusion in figures, and the details of the parameters used to construct these trees are given in Appendix 5.

8.5 Summary tree

In order to summarise the evolutionary history of the fibrillins and LTBP's the information from selected organisms was summarised in a tabular format. The design and methodology behind this format are discussed in Appendix 2.

8.6 Homology modelling

Homology modelling was carried out using Modeller 9 v7 (<http://salilab.org/modeller/>) [4] using alignments of regions of interest and selected structures. Alignments were manually edited to ensure correct positioning of key structural residues. Modelled structures were checked in pymol to ensure correct formation of di-sulphides and other key structural features.

Molecular Biology

8.7 Cloning

Prokaryotic constructs were generated by cloning DNA fragments into expression construct pQE30. Fibrillin fragments were amplified by PCR from human fibrillin1 cDNA. LTBP constructs were amplified by PCR from human LTBP cDNA's kindly provided by Professor Daniel Rifkin and Professor Jorma Keski-oja. PCR conditions are summarised in table 8.1, and the primers used as summarised in table 8.2.

Table 8.1 PCR conditions and thermocycler parameters

	PCR reaction
cDNA	10 ng
10x <i>Pfu</i> reaction Buffer	5 µl
<i>Pfu</i> DNA polymerase	1.25 U
Forward primer	0.15 µM
Reverse primer	0.15 µM
dNTPs	200 µM
QH ₂ O	Make up the volume to 50 µl

Temperature (°C)	Time (minutes)	
95	3	Initial melting
95	1	Melting
55	0.5	Annealing
72	2	Elongation
72	10	Final elongation
4	Hold	



repeat 30 times

Table 8.2 Cloning Primers

“TAGTAG” spacer included to allow restriction digest, restriction sites highlighted in red, codons for IEGR Xa cleavage site included in blue, SA spacer included in green, and introduced stop codons coloured dark red.

Construct	Primer
LTBP1:	
LTBP1 ^{TB3E3} forward	TAGTAGGGATCCATAGAAGGACGATCAGCACGACCGGCTGAGTCAAACG
LTBP1 ^{TB3E3} stop reverse	TAGTAGGTCGACTCAGACACAGGTCATCTTGGC
LTBP1 ^{TB3E3} BirA reverse	TAGTAGGTCGACGACACAGGTCATCTTGGC
LTBP1 ^{cb14TB3} forward	TAGTAGGGATCCATAGAAGGACGATCAGCAGATATGGATGAATGTCAAG
LTBP1 ^{cb14TB3} stop rev	TAGTAGGTCGACTCACACGGGGATGTTACACAG
LTBP1 ^{cb14TB3} BirA rev	TAGTAGGTCGACCACGGGGATGTTACACAG
LTBP1 ^{E3cb15} forward	TAGTAGGGATCCATAGAAGGACGATCAGCAAGCTTTGAGGAGTTACAGG
LTBP1 ^{E3cb15} stop rev	TAGTAGGTCGACTCACTCCAGGTCAGTGTCTTTC
LTBP1 ^{E3cb15} BirA rev	TAGTAGGTCGACCTCCAGGTCAGTGTCTTTC
Fibrillin 1	
FBN ^{e2cb1} fwd	TAGTAGGGATCCATAGAAGGACGATCAGCATCCAGATCCATACAACACTGC
FBN ^{e2cb1} rev stop (Hind III)	TAGTAGAAGCTTCTATTATTCACATTTTGTGACACTTC
FBN ^{e2cb1} BirA rev	TAGTAGGTCGACTTCACATTTTGTGACACTTC

Once generated PCR fragments were prepared for ligation into pQE30 by double digestion with Sal1 and BamH1 restriction enzymes, to create the required sticky ends. This was followed by gel purification using 1% Agarose-Ethidium bromide gels and a gel extraction kit (qiagen). The same restriction and gel purification was carried out on the pQE30 vector.

Ligation of DNA was carried out using T4 ligase (table 8.3), followed by transformation of 2ul of ligation mixture into 100ul of *E. coli* NM554 competent cells by heat shock. NM554 cells were also pre-transformed with pREP4 plasmid encoding the Lac repressor, in order to prevent uncontrolled expression of protein from the pQE30 plasmid. Once transformed, cells were plated onto LB-Agar containing both 25ug/ml Kanamycin and 100ug/ml Ampicillin. Colonies were screened for successful incorporation of insert by PCR, and selected positive colonies grown up in 100ml LB medium for purification of plasmid by MIDI prep protocol (Qiagen).

Table 8.3 Ligation conditions

The reactions were incubated at room temperature for 5 minutes using Roche reagents.

	Reaction mixture	Control #1	Control #2
pQE-30 cut vector	5 µl	5 µl	5 µl
DNA insert	5 µl	-	5 µl
2x ligase buffer	10 µl	10 µl	10 µl
T4 DNA ligase	1 U	1 U	-
DNA dilution buffer	-	5 µl	-
Total volume	21 µl	21 µl	20 µl

Once purified DNA was sent for Sanger sequencing by Source biosciences (<http://www.lifesciences.sourcebioscience.com/genomic-services/sanger-sequencing-service.aspx>), and resulting sequences checked for correct incorporation of sequence into vector and the absence of any mutations. All constructs used in this thesis are summarised in Table 8.4, along with their properties, predicted using the EXPASY “protparam” tool (<http://web.expasy.org/protparam/>).

Table 8.4 Construct table

Predicted masses do not include cysteine protons lost during di-sulphide formation (-6 Da for each EGF, and -8 Da for each TB or hybrid domain)

Construct name	pI with His-tag	pI without His-tag	Number of amino-acids +His	Number of amino-acids	Extinction coefficient with His-tag	Extinction coefficient without His tag	Predicted mass with his tag (reduced)	Predicted Mass without his tag (reduced)
LTBP1 constructs								
TB3E3 BirA	4.86	4.4	189	173	1.484ox 1.444red	1.622ox 1.578red	21748.1	19894.1
TB3E3	4.64		172		1.362ox 1.318red		19653.6	
TB3E3 D1521K	4.77		172		1.361ox 1.317red		19666.7	
TB3E3 D1573R	4.77		172		1.36ox 1.315red		19694.7	
TB3E3 E1625R	4.76		172		1.36ox 1.316red		19680.7	
TB3E3 E1642R	4.76		172		1.36ox 1.316red		19680.7	
TB3E3 D1655R	4.77		172		1.360ox 1.315red		19694.7	
cb14TB3 BirA	4.89		150		1.538ox 1.487red		17107	
cb14TB3	4.61	4.01	133	117	1.387ox 1.328red	1.582ox 1.515red	15012.6	13158.6
e3cb15 BirA	5.47	4.65	141	123	0.861ox 0.814red	0.985ox 0.931red	15914.7	13902.6
e3cb15	5.13	4.37	124	108	0.593ox 0.539red	0.685ox 0.623red	13820.3	11966.3
linker only (tb3-e3)	6.57	4.53	52	36	0.735	1.058	6078.5	4224.5
Fibrillin1 constructs								
e2cb1 (C204S)	8.12	7.94	193	177	0.56ox 0.482red	0.615ox 0.529red	20703.4	18849.4
e2cb1 K138D		7.33	193	177	0.560ox 0.482red	0.616ox 0.529red	20690.4	18836.4
e2cb1 R232E	7.73	7.33	193	177	0.561ox 0.482red	0.616ox 0.53red	20676.4	18822.4
e2cb1 R182E		7.33		177		0.616ox 0.53red		18822.4
e2cb1 (C204S) BirA	8.12	7.95	210	194	0.750ox 0.679red	0.816ox 0.739red	22797.9	20943.9
cb9-hyb2-cb10		4.62		153		0.841ox 0.765 red		16335.6
cb22-TB4-cb23 BirA	5.24		196		0.924ox 0.865		21326.9	
NE3 BirA	8.45	8.33	169	153	1.016ox 0.939red	1.132ox 1.047red	18052.4	16198.4

8.8 Site-directed mutagenesis

To specifically mutate codons in selected fibrillin1 and LTBP1 expression constructs the quikchange lightning site-directed mutagenesis kit was used, following the manufacturers protocol. The primers designed to introduce the desired mutations are shown in table 8.5

Table 8.5 – Primers for site-directed mutagenesis

Construct	Primer 5' > 3'
LTBP1:	
LTBP1 ^{TB3E3} D1521K fwd	CAAATAGAAGAACTAAGGTCTACCAAGATTTGTGCTGGGAAC
LTBP1 ^{TB3E3} D1521K rev	GTTCCACAGCACAAATCTTGGTAGACCTTAGTTTCTTCTATTTG
LTBP1 ^{TB3E3} D1573R fwd	CCCTGAAGGATTGAGATCGCTATGCTCAGCTGTGTAAC
LTBP1 ^{TB3E3} D1573R rev	GTTACACAGCTGAGCATAGCGATCTGAATCCTTCAGGG
LTBP1 ^{TB3E3} E1625R fwd	GTTACAGGCTGAGCGATGCGGCATCCTCAATGG
LTBP1 ^{TB3E3} E1625R rev	CCATTGAGGATGCCGCATCGCTCAGCCTGTAAC
LTBP1 ^{TB3E3} E1642R fwd	CTGTGTGAGGGTCCAGCGAGGTTACACCTGCG
LTBP1 ^{TB3E3} E1642R rev	CGCAGGTGTAACCTCGCTGGACCCTCACACAG
LTBP1 ^{TB3E3} D1655R fwd	GGGTATCACTTGCGTACGGCCAAGATGACCTGTG
LTBP1 ^{TB3E3} D1655R rev	CACAGGTCATCTTGGCCGTACGCAAGTGATACCC
Fibrillin1	
FBN ^{e2cb1} K138D fwd	CACTGTCTATGCCAGGATGGATACATAGGGACTCACTG
FBN ^{e2cb1} K138D rev	CAGTGAGTCCCTATGTATCCATCCTGGCATAGACAGTG
FBN ^{e2cb1} R182E fwd	CCAGTGTGAAAGAGATTACGAGACAGGCCCATGTTTTAC
FBN ^{e2cb1} R182E rev	GTAAACATGGGCCTGTCTCGTAATCTCTTTCACACTGG
FBN ^{e2cb1} R232E fwd	CCAGCCTCACCCCTGCGAACGTGGCTTCATTCC
FBN ^{e2cb1} R232E rev	GGAATGAAGCCACGTTTCGAGGGGTGAGGCTGG
FBN ^{e2cb1} C204S fwd	CAACTCAGCGGGATTGTCTCCACAAAACGCTCTGCTGTG
FBN ^{e2cb1} C204S rev	CACAGCAGAGCGTTTTTGTGGAGACAATCCCGCTGAGTT

Protein production

All proteins used in this thesis were produced prokaryotically and refolded *in vitro*, however the precise extraction, refolding and purification strategies used were varied to suit each construct. Figure 8.1 provides a flow diagram summarising the different methods used in this thesis to generate the required proteins, and the details of each step are discussed below.

8.9 Protein expression- rich mediay

Plasmids were transformed into thermally competent BL21 *E. coli* competent cells, which were already transformed with pREP4 plasmid, containing the Lac repressor gene to inhibit protein expression from the pQE30 plasmid. After transformation by heat shock, cells were initially plated with onto agar plates enriched with YT medium (1.6% tryptone, 1% yeast extract, 85 mM NaCl, pH 7.4) containing 25 µg/ml kanamycin (to maintain pREP4) and 100 µg/ml ampicillin (to maintain pQE30 expression vector). Plates were incubated at 37°C overnight to develop colonies, and then stored at 4°C until needed (used within two weeks of transformation). Single colonies were used to inoculate 100 ml 2 x YT medium with antibiotics in 500ml conical flasks. These cultures were grown overnight at 37°C with shaking at 200rpm. ~25 ml of each starter culture was then used to inoculate 2L flasks, each containing 500ml of 2x YT media (containing 100 µg/ml ampicillin and 25 µg/ml kanamycin). These flasks were then incubated at 37°C (with shaking at 200 rpm) until the turbidity of the cultures (A_{600}) measured between 0.8 to 1.0. Protein expression was then

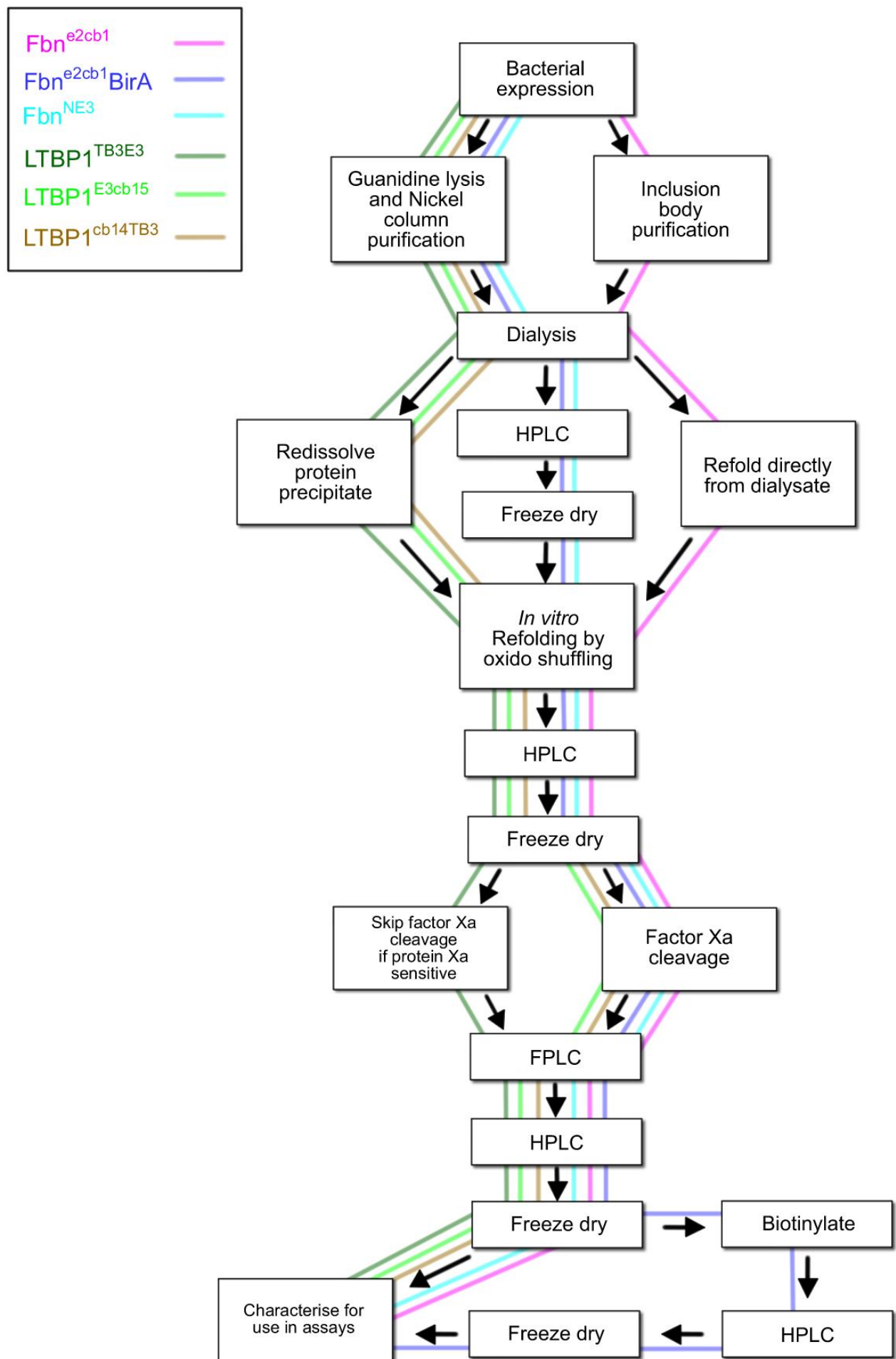


Figure 8.1 - Protein production flow diagram

Diagram summarising the many potential steps in protein production, coloured lines highlight the steps used to make specific proteins in this thesis (legend top left)

induced by addition of IPTG (isopropyl- β -D-thiogalactopyranoside) to a final concentration of 1 mM. After induction cells were grown for ~4-5 hrs, and the cultures were harvested by centrifugation in 500 ml centrifuge bottles at 5,000 rpm for 15 min (Beckman J-2, JLA-10.500 rotor). Bacterial pellets were then frozen at -80°C.

8.10 Protein expression- minimal media

BL21 cells were transformed with pQE30 expression vector as described in 8.9, colonies were then used to inoculate 120 ml starter cultures of minimal media (without isotopically labelled components). Minimal media was prepared according to the recipe in table 8.6.

Table 8.6 – Recipe for minimal media

Volumes given for full 600ml culture, for 120ml starter culture divide all values by 5.

Component	Notes
500ml water	autoclaved in 2l flask
50ml 10xNaKPNaCl	Stock: 60g Na ₂ HPO ₄ , 30g KH ₂ PO ₄ , 5g NaCl/ for 1l, autoclaved
50ml 10x Yeast Nitrogen Base without amino-acids and NH ₃ Cl	Stock: 5.9g in 300ml from “formedium” (DIFCO,#0335-15-19) sterile filtered, , store at 4 °C
6ml E. coli trace elements	To 800ml H ₂ O add 5g of EDTA and pH to 7 with NaOH. Then add the following returning to pH 7 after every addition -0.5g FeCl ₃ , -0.05g ZnCl ₂ , -0.01g CuCl ₂ , 0.01g CoCl ₂ ·6H ₂ O, 0.01g H ₃ BO ₃ make up volume to 1l and autoclave, store at 4 °C
600ul 1M MgSO ₄	sterile filtered, store at 4 °C
600ul 0.1M CaCl ₂	sterile filtered, store at 4 °C
600ul 1M Thiamine	sterile filtered, store at 4 °C
15ml 20% glucose , or 1g ¹³ C glucose	sterile filtered, (add glucose to water not other way around)
3ml 0.2 g/ml NH ₃ Cl or 0.2g/ml ¹⁵ NH ₃ Cl	sterile filtered, store at 4 °C
1.2ml ampicillin 50mg/ml	sterile filtered, store at -20 °C
300ul kanamycin 50mg/ml	sterile filtered, store at -20 °C

Starter cultures took longer to grow than those in rich media, and were generally setup early afternoon on the day before they were needed, and grown overnight at 37°C with shaking at 200 rpm. 2L flasks with 600ml prepared minimal media were then each inoculated with 50ml of each starter culture. These flasks were then incubated at 37°C (with shaking at 200 rpm) until the turbidity of the cultures (A_{600}) measured between

0.8 to 1.0. Protein expression was then induced by addition of IPTG (isopropyl- β -D-thiogalactopyranoside) to a final concentration of 1 mM. After induction cells were grown at a lower temperature of 28 °C (with shaking at 200 rpm) and left to grow for ~20 hours. Cultures were then harvested by centrifugation in 500 ml centrifuge bottles at 5,000 rpm for 15 min (Beckman J-2, JLA-10.500 rotor). Bacterial pellets were then frozen at –80 °C.

8.11 Guanidine-HCl – Nickel column purification

The frozen bacterial pellet was transferred into 30 ml guanidine buffer (6 M guanidine-HCl, 50 mM sodium phosphate, 5 mM β -mercaptoethanol, pH 7.4) and stirred for ~1 hr at room temperature. ~10 g of additional guanidine-HCl was also added to aid lysis. The cell lysate was then sonicated with three 30 second bursts of 20 W (Jencons Ultrasonic Processor), followed by ultra-centrifugation at 40,000 rpm for 30 min (Beckman L7-55 Ultracentrifuge, 60Ti rotor) to remove membranes and other cell debris. A column containing 2-3 ml fast flow chelating Sepharose (Amersham Pharmacia) was setup for affinity purification of the His-tagged protein from guanidine lysate by gravity flow. 3 ml of 1 M NiSO₄ was passed through the column, which was then washed with 50 ml H₂O to remove excess NiSO₄ before equilibration with 10ml guanidine buffer. Supernatant from ultra-centrifugation was loaded onto the column and washed with 10 ml guanidine buffer, before elution of His-tagged protein with 25 ml guanidine buffer containing 50 mM EDTA. The eluted protein was then reduced with 0.3g DTT, buffered with 0.1 M Tris-HCl pH 8.3 and left for at least 1 hr at room temperature to allow for full reduction of the protein. To remove guanidine in preparation for HPLC, the reduced protein solution was acidified to pH 2-3 with HCl and dialysed overnight at room temperature against 2 L of 0.1% trifluoroacetic acid (TFA).

8.12 Inclusion body purification

Some proteins, specifically FBN^{e2cb1} and its mutants, appeared to form highly insoluble inclusion bodies, which were resistant to guanidine lysis, and so gave poorer yields with guanidine extraction. Indications of these insoluble inclusion bodies could be seen in the pellet from ultra-centrifugation, which was dark brown to black with FBN^{e2cb1} extractions, compared to light brown to grey with LTBP extractions. To increase protein yield (particularly for isotopic labelling), an exclusion body extraction was carried out, in which cell membranes and other cell debris were removed by lysis and repeated detergent washes, leaving behind inclusion bodies of relatively pure protein. The buffer recipes for this protocol are summarised in Table 8.7.

Table 8.7- Inclusion body extraction buffers

Lysis buffer	Inclusion Body Buffer 1	Inclusion Body Buffer 3
50 mM Tris pH 8.0	20 mM Tris pH 8.0	10 mM Tris pH 8.0
1 mM EDTA	0.2 M NaCl	0.25% Sodium deoxycholate
25% Sucrose	1% Sodium deoxycholate	1 mM EGTA
	2 mM EGTA	

Bacterial pellet was thawed and resuspended in lysis buffer (20ml per Litre of culture), lysozyme was also added (2.5mg per 500ml culture) to help break cell walls (although this is not essential). Once resuspended cells were lysed by sonication with 5 x bursts of 30 sec at 20 W, swirling mixture between bursts to make sure that the temperature did not increase locally. Lysate was then centrifuged at 12 Krpm (Ja20) for 20 minutes. Supernatant was then discarded and the pellet was re-suspended in Inclusion body buffer 1 (25 mls per litre of culture). Re-suspension was assisted with repeated stirring and sonication. Mixture was centrifuged again at 12 Krpm (Ja20) for 20 minutes and the supernatant discarded. The efficiency of the inclusion body extraction can be monitored throughout the protocol by comparing the ratio of brown inclusion bodies to

white residue, which normally sits on top of the centrifuge pellet, if there is a considerable amount of this white residue after the wash with inclusion body buffer 1, it may be worth repeating this step. Next, pellet was resuspended in inclusion body buffer 2 (25 mls per litre of culture) and stirred for 15-30 minutes on bench, further sonication was used to assist resuspension. Mixture centrifuged again at 12 Krpm (Ja20) for 20 minutes. Re-suspension in inclusion body buffer 2 was repeated three times, or as many as necessary to remove the majority of white residue and give a pure brown pellet.

Inclusion body pellet was then dissolved in 20 ml 8 M urea, with 0.1 mM NaN₃, 1 mM EGTA, 50 mM DTT and 10 mM Tris pH 8.0, using a hand homogeniser to break up the inclusion bodies. Inclusion bodies could be slow to dissolve, and the presence of some remaining white cell debris made re-suspension easier, but increased precipitation during refolding later on. With large quantities of pure inclusion bodies perseverance with the hand homogeniser was required to dissolve them.

Protein-urea solution was then acidified and dialysed against 2 litres of water with 0.1% TFA. With the FBN^{e2cb1} extraction the resulting material was quite pure, as demonstrated by SDS-PAGE, and could be used directly for refolding, or further filtered and purified by HPLC.

8.13 High performance liquid chromatography (HPLC)

HPLC purifications were carried out using a Beckman Gold system equipped with a reverse-phase C8 column (Rainin). All samples were acidified to pH 2-3 and filtered with 0.22 micron syringe filters (Milipore) to remove particulates prior to loading. A 20-80% gradient of buffer B was applied over 30 min (buffer A = 0.1% trifluoroacetic acid;

buffer B = 80% acetonitrile, 0.1% trifluoroacetic acid) and chromatograms collected at 280 nm.

8.14 Re-dissolving proteins precipitated in dialysis

With some low pI proteins, especially the LTBP1 fragments, significant protein precipitation could occur when guanidine was dialysed out at low pH. Precipitate was spun out by centrifugation at 14 Krpm (Ja20) for 30mins, after which the supernatant could be kept and investigated by HPLC to look for any remaining protein in solution. The protein pellet could then be dissolved in ~30ml 1 M Tris at pH 8.3 with 30 mM cysteine, with vortexing, and some sonication if necessary (30 second bursts at 20 W, taking care not to overheat the sample, and keeping it on ice).

8.15 Protein refolding

The contrasting properties of the proteins used in this thesis meant that refolding strategies needed to be optimised for each construct. Oxido-shuffling of disulphides using cysteine and cystine was used in all cases, but other conditions were varied to optimise protein solubility and resulting yield.

For the LTBP1 constructs, LTBP1^{TB3E3}, LTBP1^{cb14TB3}, and LTBP1^{E3cb15}, (both with and without BirA tags), proteins were highly soluble at the high pHs (> pH 8) needed for disulphide shuffling, and no special components to aid protein solubility were required. These refolds were set under the following conditions: ~0.2 mg/ml reduced protein, 100 mM Tris pH 8.3, 3 mM cysteine and 0.3 mM cystine, (with MilliQ filtered water). 5 mM CaCl₂ was also included for LTBP1^{cb14TB3} and LTBP1^{E3cb15} refolds to aid folding of the cbEGF domains. Once set up, refold mix was left for 48 hrs in the cold room, then acidified to pH 2-3, centrifuged at 14 krpm (Ja20) for 30 mins, filtered through 0.22

micron membranes to remove any precipitate, and finally run on HPLC. The high pI of fibrillin FBN^{e2cb1}, and FBN^{NE3} fragments (~pH 8) made their refolding more challenging, and glycerol was required to maintain their solubility. Reduced fibrillin fragments, either re-dissolved from freeze dried material previously subjected to HPLC, or dialysate taken directly from inclusion body purification, were refolded with 50% glycerol, at a protein concentration of ~0.2 mg/ml. 3 mM cysteine and 0.3 mM cystine were present to drive oxido-shuffling, as well as 50 mM CaCl₂ for FBN^{e2cb1} refolds. The refold mix was buffered to pH 8.3 with 100 mM Tris. Refolding components were added in the following order: glycerol, water, protein, cysteine, calcium (mixed thoroughly at this stage), then Tris buffer, and finally cystine. Once setup, refold mix was left for 48 hrs in the cold room, then acidified to pH 2-3 and dialysed against 0.1% TFA overnight, filtered through 0.22 micron membranes to remove any precipitate, and run on HPLC

8.16 Factor Xa cleavage

In order to remove the His-tags from protein fragments, cleavage was carried out with bovine factor Xa enzyme (Novagen), as proteins were engineered with an "IEGR" Xa cleavage site following the N-terminal His-tag. Cleavage of LTBP1 fragments was carried out with a maximum protein concentration of 5 mg/ml, with 50 mM Tris-HCl pH 8.3, 100 mM NaCl, and 10 mM CaCl₂ and incubated at 37°C over night. Xa cleavage of FBN^{e2cb1} and FBN^{NE3} proteins were carried out with a maximum protein concentration of 1.5 mg/ml, buffered with 30mM MES-HCl at pH 6, 100 mM NaCl, and 10 mM CaCl₂ and incubated at 37°C over night. For all reactions one unit of factor Xa was used for every 2 mg of protein. The reaction mixture was then purified by ion exchange FPLC (see Section 8.17) after first diluting ~1:1 with running buffer and centrifuging sample at 14 krpm for 5 mins to remove any precipitate.

8.17 Fast protein liquid chromatography (FPLC)

Ion-exchange chromatography, was carried out on an UPC900/P900 Äkta basic FPLC system. Ion exchange was chosen to provide an alternate surface chemistry to the hydrophobicity of the HPLC reverse phase column, and so further separate out any misfolded protein, or other material, that might be insufficiently resolved from the main sample by HPLC. Different columns with either anionic or cationic surface chemistries were used depending on the predicted pI properties of the proteins. LTBP1 fragments were purified using a “MonoQ” 5/50 GL column (GE Healthcare), which has a cationic surface chemistry capable of binding the anionic LTBP1 fragments at mid to high pH's. Once proteins were injected onto the column a 2.5 % per minute gradient was applied, transitioning from 100% buffer A to 100% buffer B (buffer A: 50mM Tris-HCl pH 8.5, 2 mM CaCl₂; buffer B: 50 mM Tris-HCl pH 8.5, 1 M NaCl, 2 mM CaCl₂) and A₂₈₀ monitored to collect eluting protein fractions. For FPLC purification of FBN^{e2cb1} and FBN^{NE3} protein fragments a “MonoS” 5/50 GL column (GE Healthcare) was used, which has an anionic surface able to bind cationic fibrillin fragments in low pH buffer. For these purifications the buffer A used was a 30 mM citrate solution at pH 4, and buffer B was a pH 4 30 mM citrate solution with 2 M NaCl added. After FPLC all proteins were then acidified to ~pH 2, filtered to remove any precipitate, and desalted by further HPLC purification, before finally being freeze dried.

8.18 Biotinylation

For use in pull down assays recombinant FBN^{e2cb1} “BirA” protein, that had been expressed with a C-terminal BirA recognition peptide (VDLHHILDAQKMVWNHR), was biotinylated using BirA biotin ligase (Avidity *in vitro* biotinylation kit). Reactions were set up as summarised in Table 8.8, and incubated at room temperature overnight.

Biotinylated protein was then acidified and purified by HPLC, with an additional 15-minute equilibration phase before starting the HPLC gradient, in order to elute excess ATP. Protein was subsequently analysed by SDS-PAGE and mass spectrometry to confirm Biotinylation.

Table 8.8 Biotinylation reaction conditions.

The reaction was incubated at 25°C overnight

	Final concentration
BiomixA (0.5 M Bicine buffer, pH 8.3)	10% (v/v)
BiomixB (100 mM ATP, 100 mM magnesium acetate, 500 µM biotin)	10% (v/v)
Tris-HCl pH7	10% (v/v)
BirA	4 µg/ml
Protein	750 µg
QH ₂ O	Make up volume to 1 ml

Biochemical and biophysical assays

8.19 SDS-PAGE

Proteins were analyzed with SDS-PAGE by loading into lab made 1.5 mm thick polyacrylamide gels, which were composed primarily of a 16% separating gel [16% acrylamide, 0.1% SDS, 0.38 M Tris-HCl pH 8.8, polymerized with 0.15% ammonium persulphate (APS) w/v, and 0.15% N,N,N',N'-tetramethylethylenediamine (TEMED)] v/v, and capped with a 6% stacking gel (6% acrylamide, 0.25 M Tris-HCl pH 6.8, 0.1% SDS, 0.2% APS, 0.2% TEMED).

Prior to loading onto the gel, protein samples were heated at 95 °C for 5 minutes, with an equal or greater volume of SDS loading buffer (50 mM Tris pH 6.8, 10% glycerol, 3% SDS, 0.2% bromophenol blue) for non-reducing samples, and for reducing samples the same SDS loading buffer was used with 50 mM β-mercaptoethanol

added. The gels were run for 45 minutes at 180 V in Tris-glycine buffer (30 mM Tris-HCl, 250 mM glycine pH 8.3, 0.1% SDS) and then stained with Coomassie brilliant blue (Sigma) (1.25 g in 1L 10% Methanol, 10% acetic acid) for 1 hr at 50 °C. “Destaining” was carried out overnight at room temperature in 10% methanol, 10% acetic acid aqueous solution on a rocking platform.

8.20 Mass-Spectrometry

Mass of purified proteins was confirmed by Electrospray Ionisation Mass spectrometry (ESI-MS). Prior to analysis, freeze-dried proteins were re-dissolved and samples diluted to 10 µM protein in 100 µl 50% acetonitrile: 50% H₂O solution containing 0.1% formic acid. Samples were loaded onto the autosampler of an A Micromass Platform-II equipped with Agilent HP100 Sample Handling system at a rate of 1µl/min. Collision data was collected for 1 minute, and analysed using the maximum entropy method in the mass lynx software package.

8.21 UV-vis spectroscopy - protein concentration determination

The protein extinction coefficients predicted by the EXPASY protparam tool (<http://web.expasy.org/protparam/>) (Table 8.5) were routinely used for determination of protein concentration, using UV-spectroscopy measurements at 280 nM collected with a shimadzu UV mini 1240 spectrometer. For UV measurements of protein concentration, protein samples were dissolved in 6 M guanidine- 0.1% TFA, and mixed thoroughly prior to data collection. Use of 6 M guanidine ensured against protein aggregation that could otherwise give scattering and inaccurate measurements of concentration.

8.22 Calcium Chelation

Calcium binding by FBN^{e2cb1} was confirmed by a chromophoric chelator competition assay (summarised in Figure 8.2). The high affinity chelator Quin-2 was used as described previously [5] using 20 or 30 μ M protein in calcium free buffer (5 mM Tris-HCl, 150 mM NaCl, pH 7.5). Protein-chelator solution was titrated with calcium buffer (1 mM CaCl₂, 5 mM Tris-HCl, 135 mM NaCl, pH 7.5) in the presence of 30 μ M chromophoric Quin2 chelator, until A_{263} stopped decreasing (measured using a Shimadzu UV mini 1240 spectrophotometer). In the presence of a calcium-binding protein A_{263} of the chelator decreases more slowly due to competition by the protein for calcium binding. The residual calcium concentration in the stock solution of Quin2 chelator was determined so that it could be added to the “[Ca²⁺] total” value. This was done by dissolving chelator in calcium free buffer to a final concentration of 30 μ M and measuring the following:

$$\begin{aligned}A_1 &= A_{263} \text{ of untreated solution} \\A_2 &= A_{263} \text{ after adding } 2 \mu\text{l } 0.5 \text{ M EDTA} \\A_3 &= A_{263} \text{ after adding } 5 \mu\text{l } 1 \text{ M CaCl}_2\end{aligned}$$

The following equation was used to calculate residual calcium concentration

$$[\text{Ca}^{2+}_{\text{initial}}] = [Q] \times (A_2 - A_1)/(A_2 - A_3)$$

Where Q is the chelator concentration (calculated using the extinction coefficient, ϵ = 42000 M⁻¹cm⁻¹ at 239 nm in calcium saturated conditions)

8.23 Fluorimetry

Fluorescence emission spectra were collected from 4 μ M protein samples under both reducing conditions (6 M guanidine-HCl, 50 mM sodium phosphate, 5 mM β -mercaptoethanol, pH 7.4) and in non-reducing conditions (50 mM Tris pH 7.4) over a

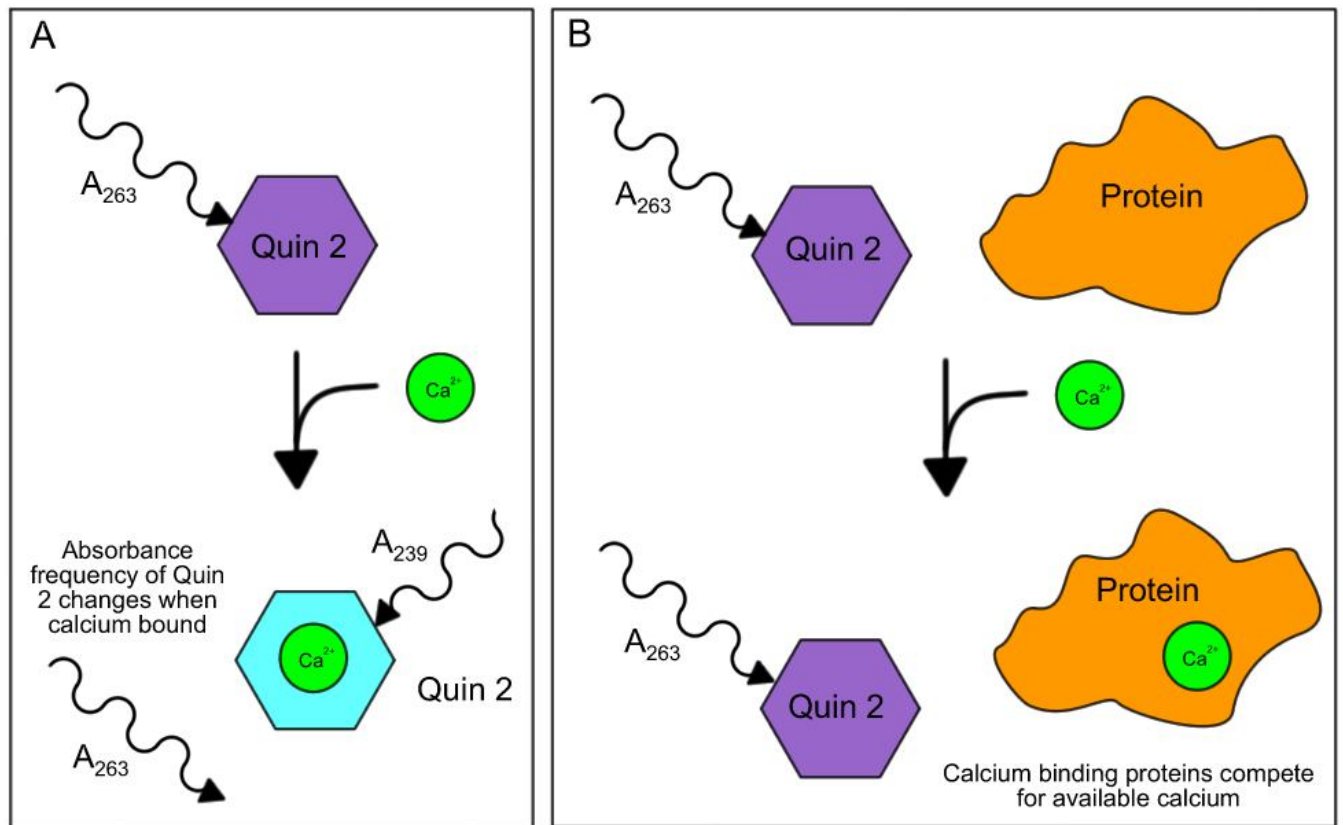


Figure 8.2 - Calcium chelation

Diagram summarising the affect of calcium on the Quin 2 calcium chelator, **A)** when alone and **B)** when in the presence of a high calcium affinity protein.

wavelength range of 250-400 nm. Fluorescence spectra were collected using a Perkin Elmer LS-50B fluorimeter with FL-Winlab software, from a 1 ml sample contained in a quartz fluorimetry cuvette.

8.24 Pull down experiments

Pull-down assays were used to detect protein-protein interactions (Figure 8.3) conditions for these experiments optimised from those previously described [6]. 5.8 μ M his-tagged recombinant LTBP1 or negative control protein was incubated with Nickel – NTA magnetic agarose beads (Qiagen) for half an hour in pull down buffer 1 (10 mM Tris pH 7.5, 150 mM NaCl, 1 mM CaCl_2 , 80 mM imidazole, and 0.05% Tween 20) at room temperature on a rotating platform, then washed 3 times with pull down buffer 1, and a further 3 times with pull down buffer 2 (10 mM MES pH 6, 150 mM NaCl, 1 mM CaCl_2 , 80 mM imidazole, and 0.05% Tween 20). Beads were then incubated with 0.5 μ M biotinylated fibrillin fragments in pull down buffer 2 for an hour at room temperature on a rotating platform. Beads were then washed 3 times with pull down buffer 2, and 3 times with tris buffered saline (TBS), followed by incubation with non-reducing SDS buffer (see section 8.19) for 5 minutes at 95 °C and loaded onto polyacrylamide gels for analysis by SDS-PAGE (see section 8.19) followed by western blotting.

8.25 Western blotting

Once protein samples separated by SDS-PAGE they were transferred onto a nitrocellulose membrane (Whatman Protran) by Western Blotting for 90 minutes at 100V using a Mini Trans-Blot Cell (Biorad), filled to the brim with transfer buffer (25 mM Tris, 190 mM glycine, 20% (v/v) methanol and 0.02% (w/v) SDS). Membranes used to detect biotinylated proteins were blocked using 5% (w/v) non-fat milk powder in TBST (TBS containing 0.1% (v/v) Triton X-100 (Sigma) , incubated at room

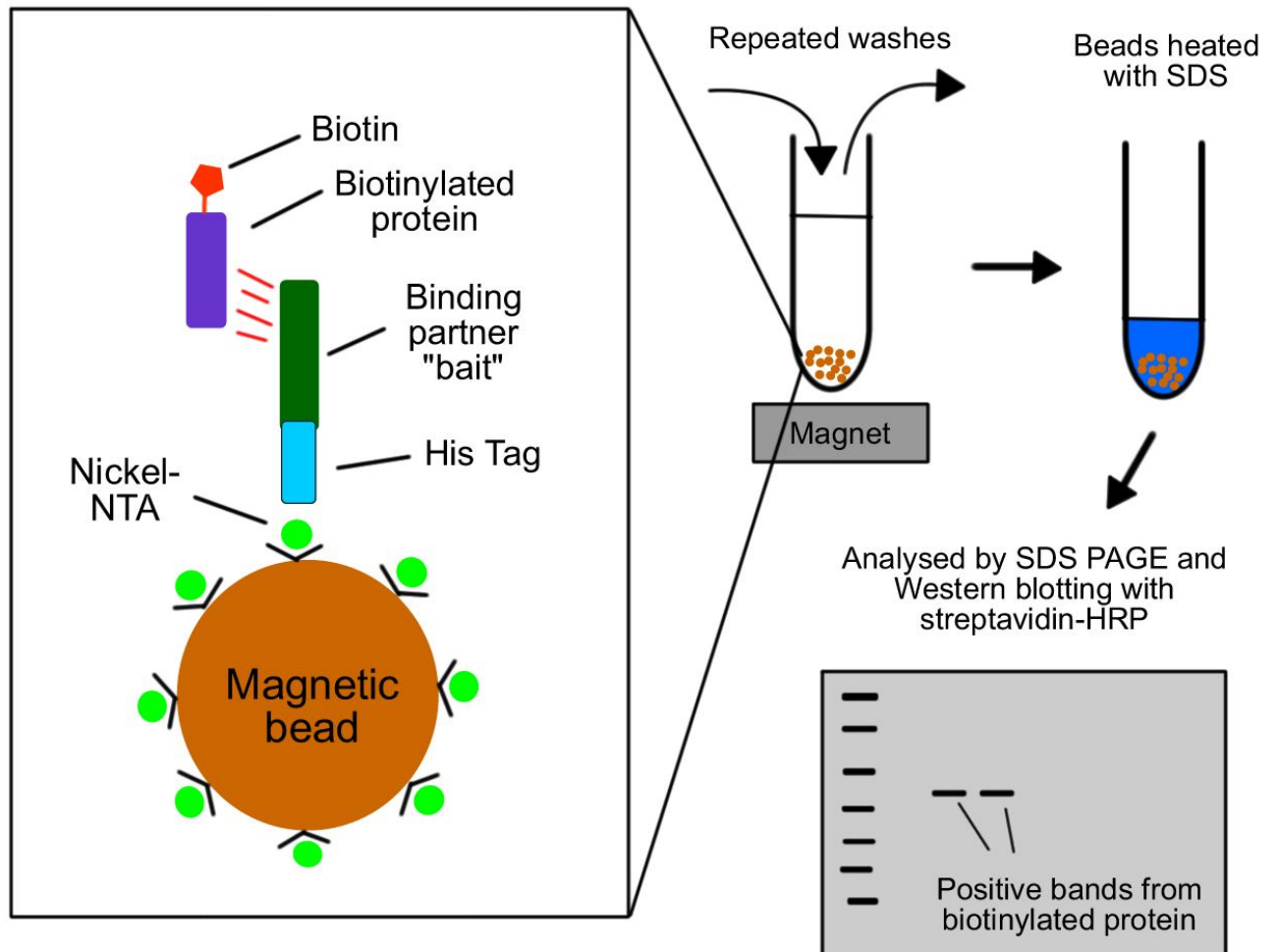


Figure 8.3 - Pull down principle

On the left is shown a detailed picture of how LTBP "bait" proteins are proposed to bind biotinylated fibrillin fragments, and on the right the key pull down steps are summarised, as described in the text

temperature on a rocking platform for 1 hour. Membranes used to detect His-tagged protein, (in order to confirm equal loading of beads with bait protein), were blocked using Qiagen blocking reagent (according to manufacturer's instructions). Membranes were blocked for 1 hour then washed twice in TBST buffer. Biotinylated protein was detected using Streptavidin-Horseradish peroxidase (HRP) conjugate (Sigma, diluted 1:5000 in TBS buffer) and His-tagged protein was detected using an Anti-RGS-His HRP antibody conjugate (Qiagen, diluted 1: 2000 in blocking buffer). Membranes were incubated with their respective detection antibodies for one hour on a rotary platform at room temperature. Membranes were then washed four times with TBST, incubating each wash for five minutes on the rotary platform. Detection of membrane bound HRP was carried out by chemiluminescence using a PerkinElmer Western LightingTM Plus-ECL kit and visualization with Amersham Hyperfilm ECL (GE Healthcare) developed with an automatic Compact-X4 developer (Xograph Imaging systems).

8.26 Surface plasmon resonance (SPR)

SPR studies were carried out using a Biacore T100 instrument (GE Healthcare) to monitor complex formation between fibrillin and LTBP protein fragments (as summarized in Figure 8.4). Fibrillin fragments were initially coupled to the surface of a CM5 (carboxymethylated dextran) sensor chip by amine coupling. ~1000 RU of protein was coupled to each flow cell using the coupling wizard protocol available in the T100 BIAcontrol software. In this wizard the chip was equilibrated with HBS-N buffer (HEPES buffered saline pH 7) and protein pre-concentration onto the chip surface tested (and then regenerated with 50mM NaOH). 0.4 M EDC [1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride] was then mixed 1:1 with 0.1 M NHS (N-hydroxysuccinimide) and injected over the chip to activate the carboxymethyl

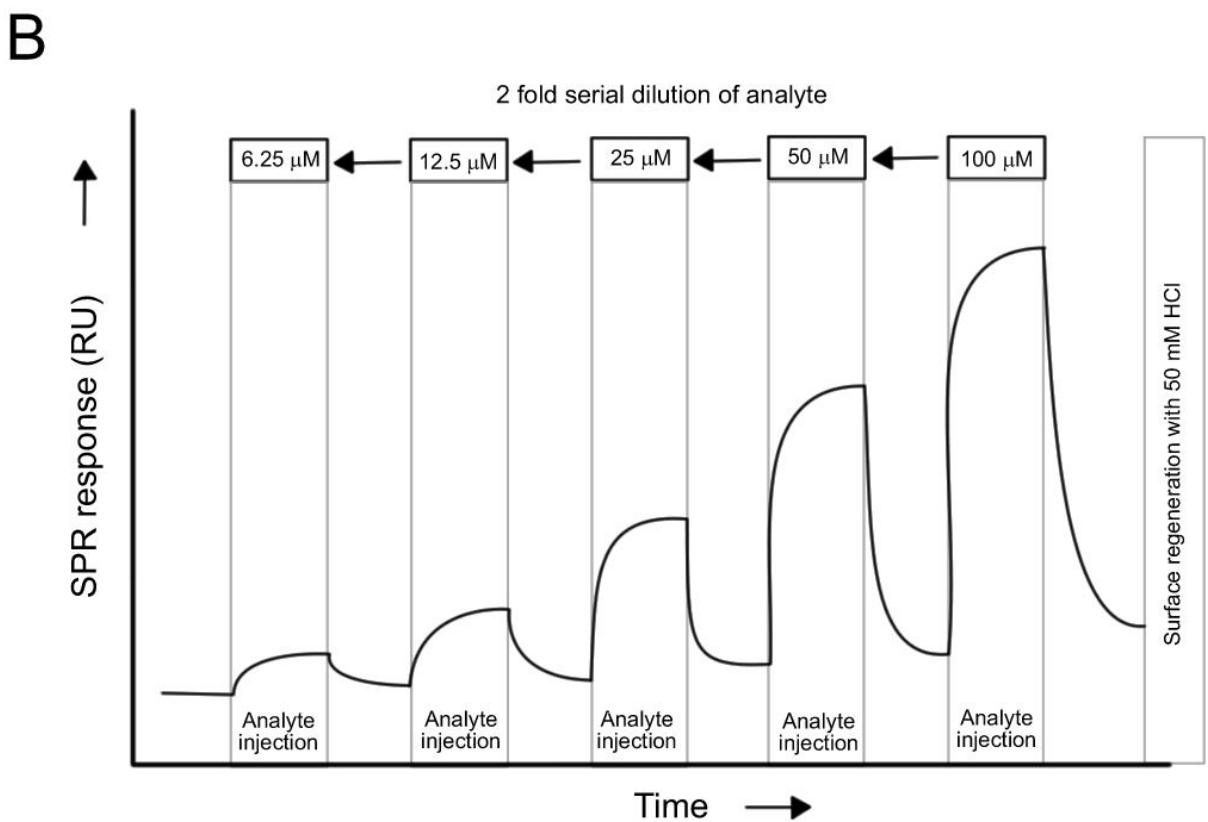
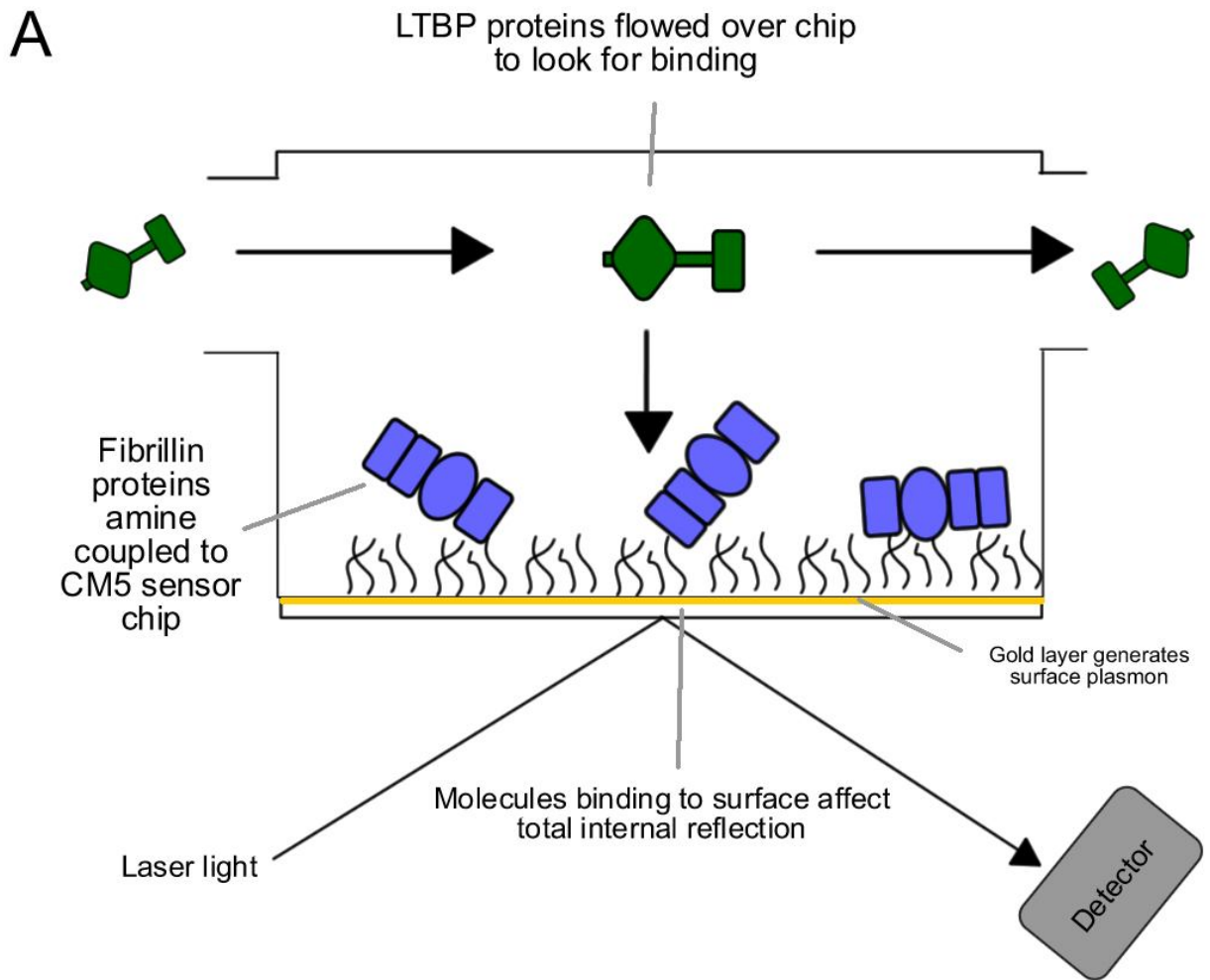


Figure 8.4 - SPR mechanisms

A) Experimental setup demonstrating LTBP proteins as "analytes" flowed over the chip, and fibrillin proteins as "ligands" bound to the CM5 chip surface. **B)** Example results from an idealised single-cycle SPR experiment.

groups of the CM5 chip, so that they can react with exposed protein amine groups. Immediately after activation, fibrillin1 fragments were injected at a concentration of 10 µg/ml in 50 mM NaOAc buffer (pH 5 for FBN^{e2cb1} and FBN^{NE3}, pH 4 for FBN^{cb22TB4cb23}) over the chip surface, in a series of automatically controlled injections to achieve the target 1000 RU. The chip surface was then blocked with 50 µl of 1M ethanolamine at pH 8.0, and final response levels checked to ensure successful coupling had occurred. Flow cell one was used as a blank and treated with EDC-NHS and ethanolamine alone. After coupling, a separate manual run was then carried out in which 50 mM HCl was injected over all flow cells for 1 minute at a flow rate of 20 µl/min, before returning to HBS-N buffer, in order to check stable coupling of proteins to the flow cells.

Sensor chip was then equilibrated with SPR running buffer¹ (50mM Tris pH 7.5, 150mM NaCl, 0.05% (v/v) Tween 20 and 2mM CaCl) and injection of analytes carried out using either single-cycle or multi-cycle biacore programs. Analyte samples were dissolved directly into SPR buffer from freeze-dried samples to generate stock solutions, the protein concentrations in these stock solutions were re-measured by UV-vis spectroscopy², and SPR samples and further serial dilutions made up from these stock samples in 96 well plates. 96 well plates were covered once prepared to prevent sample evaporation. All SPR experiments were performed at 25 °C, with a flow rate of 20 µl per minute, and a 30 second 50mM HCl injection was used for regeneration after each run.

¹ SPR buffer was made fresh for each round of experiments, as although stable for several days, after a week or so some precipitate began to appear. this seemed to be dependent on the presence of both Tween 20 and Calcium. When making up the SPR buffer it was also advisable to add calcium last, and after Tween 20 had been added and thoroughly mixed into the solution. When Tween 20 was added as the last component, some precipitation could be seen at its point of contact with the solution. Buffer was also filtered through a 0.22 micron PVDF membrane prior to use.

² At this stage further concentration of protein using viva-spin columns or similar methods should not be attempted, as Tween-20 detergent also appears to be concentrated, which then interferes with SPR measurements.

The single-cycle method uses five sequential injections of analyte, with buffer injections in-between, but no regeneration until the end of the experiment (Figure 8.4 B). For these injections two fold serial dilutions of a 200 μ l starting solution were used, starting with the most dilute. This single cycle method was used initially, as it allowed collection of a large amount of data in a small time frame, and reduced degradation of the chip surface from frequent regeneration. However repeated experiments showed proteins bound to the chip surface to be quite stable, (and were also relatively stable when stored at 4 °C in water alone). Multi-cycle method SPR experiments were also carried out for selected proteins in chapter 6, which were run under the same conditions as single-cycle experiments, except with regeneration between each injection. Data was analyzed using the BIAevaluation T100 software (Biacore) and Excel (Microsoft). Response values from the blank flow cell (flow cell one) were automatically subtracted from those of other cells, and blank runs of buffer injections alone were also performed, and subtracted from the analyte injection data to improve the resulting sensorgrams. For non-linear regression analyses of K_d the BIAevaluation fitting tool was used, with automatic offset determination.

NMR experiments

8.27 General experiment information

Numerous NMR experiments were carried out on various instruments and under various conditions, which are summarised in Table 8.9. Many of these experiments were performed with the assistance of Professor C. Redfield. Experiments were carried out using a variety of spectrometers operating at different frequencies; a Bruker Avance spectrometer operating at 500 MHz with attached TCI cryoprobe was used for collection of many of the triple resonance experiments for the assignment of

Table 8.9 – Summary of NMR experiments

Experiment	Notes	Frequency			Sweep width			Mixing time	Number of increments		
		F3	F1	F2	F3	F1	F2		F3	F1	F2
FBN ^{e2cb1} experiments											
HSQC-1 for assignment	pH 5.4, 5 mM CaCl, 95% H ₂ O:5% D ₂ O	949.9008	96.263578		11764.71	2844.86			1024	128	
HSQC-2 for assignment	"	500.012	50.671001		6250	1724.14			1024	128	
HSQCs - calcium sensitivity	pH 5.4, 5 mM EDTA or 5mM EDTA +15mM CaCl	500.0687	50.677301		6535.948	1488.10			1024	128	
HSQCs D ₂ O exchange	pH 5.4, 5 mM CaCl, 100% D ₂ O	598.5624	60.658712		7407.407	1785.71			1024	128	
HSQCs e3cb15 titration	pH 5.4, 5 mM CaCl, 100mM NaCl, 95% H ₂ O:5% D ₂ O	500.0687	50.677301		6535.948	1488.10			1024	128	
HSQCs cb14TB3 titration	pH 5.4, 5 mM CaCl, 95% H ₂ O:5% D ₂ O	500.0687	50.677301		6535.948	1488.10			1024	128	
Hetero-nuclear NOE	" 4s saturation time	749.9940	76.004897		9345.794	2232.14			1024	2x 128	
2D NOESY	pH 5.4, 0 and 5 mM CaCl 100% D ₂ O	749.9904			9009	2232.14			1024	256	
2D NOESY (R232E)	pH 5.4, 0 mM CaCl 100% D ₂ O	500.0687			6024.096				1024	256	
¹⁵ N 3D-NOESY 1	pH 5.4, 5 mM CaCl, 95% H ₂ O:5% D ₂ O	949.9008	96.263578	949.9008	11764.71	2844.86		150 ms	512	32	128
¹⁵ N -3D TOCSY	"	949.9008	96.263578	949.9008	11764.71	2844.86		35 ms	512	32	128
¹⁵ N 3D NOESY 2	"	949.8483	96.258263	949.8483	11764.71	2844.86		75 ms	512	32	128

HNCA	"	500.012	50.672001	125.727	6250	1724.138	3521.127		512	28	50
CBCANH	"	500.012	50.672001	125.727	6250	1724.138	7462.687		512	28	50
CBCACON H	"	500.012	50.672001	125.727	6250	1724.138	7462.687		512	28	50
HNCO	"	500.012	50.672001	125.727	6250	1724.138	1886.24		512	28	50
HNCACO	"	500.012	50.672001	125.727	6250	1724.138	1886.24		512	28	50
HBHACON H	"	500.012	50.672001		6250	1724.138			512	28	
HCCH TOCSY	pH 5.4, 5 mM CaCl, 100% D ₂ O	500.012		125.733	6250		8802.817		512	46	128
¹³ C – NOESY 1	pH 5.4, 5 mM CaCl, 100% D ₂ O	500.012		125.733	6250		8802.817	75 ms	512	46	110
¹³ C – NOESY 2	pH 5.4, 5 mM CaCl, 95% H ₂ O:5% D ₂ O	949.8483		238.8479	11764.71		14705.88	75 ms	512	46	110
IPAP	pH 5.4, 5 mM CaCl, 95% H ₂ O:5% D ₂ O or 10% D ₂ O and 2.2% PEG-hexanol	600.1508	60.819671		7434.944	1908.40			1024	2x 128	
LTBP1 ^{cb14TB3}											
HSQC-1 for assignment	pH 5.9, 20 mM CaCl 95% H ₂ O : 5% D ₂ O	500.0687	50.677264		6535.948	1515.15			1024	128	
HSQCs - calcium sensitivity	"	500.0687	50.677264		6535.948	1515.15			1024	128	
HSQCs D ₂ O exchange	pH 5.9, 20 mM CaCl 100% D ₂ O	500.0687	50.677264		6535.948	1515.15			1024	128	
HSQCs protein titration	pH 5.9, 20 mM CaCl 95% H ₂ O : 5% D ₂ O	500.0687	50.677264		6535.948	1515.15			1024	128	
Hetero- nuclear NOE	" (ilim1 300)	500.0687	50.677264		6535.948	1515.15			1024	2x 128	

2D NOESY	pH 7.7, 0 and 10mM CaCl, BirA tagged protein 100% D ₂ O	749.994			9009.009			150 ms	1024	256	
¹⁵ N 3D-NOESY 1	pH 5.9, 20 mM CaCl 95% H ₂ O : 5% D ₂ O	500.0687	50.677264	500.0687	6535.948	1515.15		150 ms	512	32	128
¹⁵ N -3D TOCSY	"	500.0687	50.677264	500.0687	6535.948	1515.15		50 ms	512	32	128
LTBP1 ^{E3cb15}											
HSQC for assignment	pH 5.4 10 mM CaCl, 95% H ₂ O : 5% D ₂ O	598.5624	60.658666		6849.315	1785.71			1024	128	
HSQCs - calcium sensitivity	pH 5.4, 0 and 10 mM CaCl, 95% H ₂ O : 5% D ₂ O	749.994	76.004897		8547.009	2500.00			1024	128	
HSQCs D ₂ O exchange	pH 5.4, 10 mM CaCl, 100% D ₂ O	500.0687	50.677301		5730.659	1672.24			1024	128	
HSQCs protein titration	pH 5.4 10 mM CaCl, 100mM NaCl, 95% H ₂ O : 5% D ₂ O	500.0687	50.677301		5730.659	1672.24			1024	128	
Hetero-nuclear NOE	pH 5.4 10 mM CaCl, 95% H ₂ O : 5% D ₂ O 4s saturation time	749.994	76.004897		8547.009	5000			1024	2x 112	
¹⁵ N 3D-NOESY	pH 5.4 10 mM CaCl, 95% H ₂ O : 5% D ₂ O	598.5624	60.658666	598.5624	6849.315	1785.71		150 ms	512	32	128
¹⁵ N -3D TOCSY	"	598.5624	60.658666	598.5624	6849.315	1785.71		45 ms	512	32	128
HNCA	"	500.012	50.672	125.727	6250	1524.39	3521.127		512	28	50
CBCANH	"	500.012	50.672	125.727	6250	1524.39	7462.687		512	28	50
CBCACON H	"	500.012	50.672	125.727	6250	1524.39	7462.687		512	28	50
HNCO	"	500.012	50.672	125.727	6250	1524.39	1886.24		512	28	50
HNCACO	"	500.012	50.672	125.727	6250	1524.39	1886.24		512	28	50
LTBP1 ^{TB3E3}											

HSQC for assignment	pH 6.9 95% H ₂ O : 5% D ₂ O	598.4844	60.650763		7812.5	1754.39			1024	128	
Hetero-nuclear NOE	"										
	3s saturation time	500.0687	50.677264		6535.948	1515.15			1024	2x128	
2D NOESY stop	pH 7.7 100% D ₂ O	600.1507			7042.254			150 ms	1024	256	
2D NOESY D1655R	"	600.1507			7042.254			150 ms	1024		
¹⁵ N 3D-NOESY 1	pH 6.9 95% H ₂ O : 5% D ₂ O	598.4844	60.650763	598.4844	7812.5	1754.39		150 ms	512	32	
¹⁵ N -3D TOCSY	"	598.4844	60.650763	598.4844	7812.5	1754.39		46 ms	512	32	

FBN^{e2cb1} and LTBP1^{E3cb15} fragments, while other experiments were carried out on several home-built instruments, incorporating “Oxford instruments” magnets, operating at 500-950 MHz. All experiments were carried out at 25 °C unless stated otherwise.

Experiments were processed using the NMRpipe software tool [7], running on a Sun Microsystems UNIX based computer network. Spectra were analysed and figures generated using the CCPN analysis software [8].

8.28 NMR - 2D experiments

A number of 2D NOESY (Nuclear Overhauser Effect Spectroscopy) experiments were carried out [9] primarily to confirm the fold of non-isotopically labelled proteins, but were also used for FBN^{e2cb1} structure determination. The purpose of these experiments is to follow the cross-relaxation peaks between ¹H resonances that are closely positioned in space. The pattern of peaks then reveals structural information about the environment of amino acid protons, and through-space inter- and intra-residue connections.

Numerous 2D HSQC (Heteronuclear Single Quantum Coherence spectroscopy) experiments were also performed to aid the assignment of protein amides [10], and examine amide peak behaviour during titrations and other experiments. HSQC experiments are based on the transfer of magnetization from an ¹H nucleus to a “hetero” nucleus, in this case of an ¹⁵N atom (via an Inensitive Nuclei Enhanced by Polarization Transfer (INEPT) step), and then back to the ¹H nucleus (via a “reverse-INEPT” step). This results in a 2D spectrum with peaks for all ¹H-¹⁵N bonded atom pairs, which for a protein primarily consists of peaks from backbone and side chain amides. The sensitivity of this experiment, and the small number of generally well

dispersed peaks, representing each amino acid of the protein molecule (except prolines), make it a particularly useful experiment for following changes in protein local environment.

8.29 NMR – 3D spectra and assignment

As described in Chapters 3 and 4 protein spectra were assigned using 3D double resonance experiments, and triple resonance experiments as necessary, these experiments were performed with assistance from Professor C. Redfield. ^{15}N -edited TOCSY (TOtal Correlation SpectroscopY) 3D experiments were used to identify intra-residue spin-couplings [11], and ^{15}N -edited NOESY experiments used to examine the through-space connections between residues [11], which in some cases gave sufficient information to allow sequential backbone assignment.

In the case of the FBN^{e2cb1} and LTBP1^{E3cb15} proteins, however, the ^{15}N -edited experiments did not provide sufficient information for comprehensive assignment, and so protein samples were isotopically labelled with both ^{15}N and ^{13}C for use in triple resonance experiments (see minimal media expression Section 8.10). For backbone resonance assignment 3D HNCA, CBCANH, CBCACONH, HNCOC, and HNCACO experiments were performed [12]. The HNCA experiment gives peaks for C^α resonances of residue of (i) and the residue preceding (i-1) for each $^1\text{H}^\text{N}$ - ^{15}N pair. The CBCANH experiment gives peaks for these resonances and also the C^β resonances of residues i and i-1, whereas the CBCACONH just gives peaks for C^α and C^β of the i-1 residue. Sequential connections determined with these three experiments were also further confirmed using HNCOC and HNCACO experiments, where the HNCOC experiment gives peaks for the carbonyl carbon of residue i-1, and the HNCACO experiment gives a carbonyl carbon peak for both residue i and i-1. Using the

information in these experiments it is possible to connect sequential amide resonances, and by using general predicted chemical shift information for different residue types, these sequential assignments could be fitted to regions of the protein's amino acid sequence.

For the FBN^{e2cb1} protein, side chain resonance assignment was also carried out; ¹⁵N-edited TOCSY and NOESY experiments were used in combination with an HBHACONH triple resonance experiment [13], which allowed specific assignment of H^β and H^α resonances of i-1 residues for each ¹H^N-¹⁵N pair. HCCH-TOCSY and ¹³C-edited NOESY experiments [12] were then used to obtain further side chain assignments where possible, although this was frustrated in some cases by regions with multiple over-lapping peaks, or peaks with weak signal intensity.

To determine structural distance restraints from the FBN^{e2cb1}, NOE assignment was carried out using 2D ¹H-¹H, 3D ¹⁵N-edited, and 3D ¹³C-edited NOESY spectra in combination. All assignment was carried out using the CCPN analysis software package [8] for annotation of peaks, and conversion of peak intensities into distance restraints.

8.30 NMR - Residual dipolar couplings

In order to collect residual dipolar coupling (RDC) values IPAP spectra were collected [14], both in isotropic medium (95% H₂O with 5% D₂O, 5 mM calcium, at pH 5.4), and in a 2.2% PEG-hexanol liquid crystal alignment medium [15] (90% H₂O, 10% D₂O, 5 mM calcium at pH 5.4). In order to prepare the protein sample in alignment medium a 10% stock of PEG-hexanol was initially prepared, in which hexaethylene (C12E6) polyethylene glycol was dissolved in H₂O:D₂O by heating and vortexing. Hexanol was

then added in small volumes (1 hexanol: 50 total (v/v) initially then 1:200 sequential additions). On addition of hexanol the stock solution first transitions to a milky colloidal phase and then to a clear but viscous liquid crystal phase, at which point no further hexanol addition was needed (care was needed not to add too much hexanol as the solution could then irreversibly return to a colloidal phase). PEG-hexanol stock solution could then be mixed with the protein sample, if liquid crystal reverted to a colloidal phase at this stage this could be corrected with small extra additions of hexanol. The pH of the sample was checked and adjusted as necessary and IPAP experiments were then performed.

For the analysis of RDCs for determination of D_a and R , the XPLOR protocol CalcDaRh was used with the assistance of D. Yadin, and interpretation of RDC distribution histograms was also used with the assistance of Professor C. Redfield [16]. While these different techniques did give different estimates³ of D_a and R , the use of these different values in refinement calculations did not affect the domain orientations of the structures obtained.

8.31 NMR – Structure determination

Once a sufficient number of NOEs had been assigned for the FBN^{e2cb1} protein, structure calculations were carried out with the assistance of Professor C. Redfield. As described in Chapter 3 (Section 3.8) refinement calculations using a homology model starting structure were used initially for iterative improvement of NOE and other restraints. These calculations were run on a Sun Microsystems network using the XPLOR-NIH structure determination software version 3.8 [17] (later versions were not

³ D_a of 20 and R of 0.55 was initially obtained with the preliminary set of RDCs used for XPLOR analysis, which were the values used for the majority of structure calculations. Interpretation of the whole protein histogram gave a D_a of 25 and R of 0.27, while interpretation of the hyb1-cbEGF1 histogram alone gave a D_a estimate of 22 and R of 0.21.

compatible with the pro-chiral swapping used in these initial calculations [18]). Table 8.10 summarises the force constants and other key parameters used in these calculations. The structure of hyb2, on which the starting homology model structure was based, contains a cis-proline, and in hyb1 there was a proline at the equivalent position (proline 220). To determine whether proline 220 also adopted the cis-form, the differences between C^β and C^δ chemical shifts were measured as 8.6 ppm, which is consistent with a cis proline in this position [19]. For other proline residues where assignments were available $C^\beta - C^\delta$ differences were all consistent with trans prolines.

Table 8.10- Initial refinement parameters

Also used in refinement of simulated annealing structures

Parameter	Value
Distance restraint force constant	50
Dihedral restraint force constant	200
RDC restraint force constant	0.01-0.5
Initial temperature	1500
Number of cooling steps	20000
Bond force constant	1000
Angle force constant	500
Improper force constant	500

Once consistent structures, giving suitably low energies, were obtained from homology model refinement, full simulated annealing calculations from a random template were carried out. This was followed by further refinement, at first without RDCs then with RDCs included. These calculations were performed by Professor C. Redfield using the XPLOR-NIH version 2.29, running on an iMac Duo core system running Mac OSX 10.6.8. The key force constants and other parameters are summarised in Table 8.11.

Table 8.11- Simulated annealing parameters

Parameter	Value
Distance restraint force constant	10-50
Dihedral restraint force constant	5-200
Initial temperature	2000
High temperature steps	30000
Cooling steps	40000 + 20000

To generate the structure figures shown in Chapter 3 structures were aligned, coloured, and rendered using pymol.

8.32 NMR – D₂O exchange

To look at amide solvent accessibility and hydrogen bonding in various proteins, D₂O exchange experiments were carried out. In these experiments ¹⁵N-labelled protein samples were prepared at the desired pH, with calcium added if necessary, and then freeze dried. These samples of protein were then dissolved in 100% D₂O, and successive HSQCs collected to examine the loss of peak intensity. This peak intensity loss occurs due to backbone and side chain amides exchanging with the deuterium oxide solvent, replacing ¹H-¹⁵N atom pairs with ²H-¹⁵N atom pairs, which do not give a signal in HSQC spectra.

8.33 NMR – Heteronuclear NOE

¹H-¹⁵N Heteronuclear NOE experiments were carried out on ¹⁵N-labelled protein samples in order to examine the sub-nanosecond dynamics of specific amides [20]. This experiment specifically looks at motion of N-H bond vectors that is faster than the overall tumbling of the protein. To get specific heteronuclear NOE values, an NOE experiment is collected in which ¹H resonances are saturated, and a blank experiment is collected where ¹H resonances are in equilibrium, and the intensity of peaks in the NOE experiment are divided by that of the peaks in the blank experiment, to give the heteronuclear NOE ratio. Low or negative ratios reflect significant fast timescale movement of an amide, while high ratios (>0.6) reflect structured amides moving on a similar timescale to overall protein tumbling. Blank and NOE experiments were collected together as interleaved experiments and so should be unaffected by any protein degradation or other time dependant sample changes.

Error bars were calculated by taking the standard deviation of random background noise, and recalculating ratios after adding this standard deviation to the NOE peak intensities, and subtracting it from the blank peak intensities (both sets of peak intensity were squared and square-rooted to render them positive values). This method may have slightly overestimated errors when compared to using a Monte-Carlo simulation to add random noise values to peak intensities, but the standard deviation method could easily be performed in Microsoft Excel.

8.34 NMR – HSQC protein titrations

In order to identify specific sites of interaction between specific proteins in Chapter 5, a number of HSQC protein titrations were carried out. Initially an HSQC spectrum was collected from an ^{15}N -labelled sample of the protein alone, and then successive freeze dried aliquots of unlabelled interaction partner were added, with the pH of the NMR sample measured before and after each addition, and adjusted accordingly prior to running each HSQC experiment. For the titration of ^{15}N -LTBP1^{cb14TB3} with FBN^{e2cb1} a significant excess of saturating calcium could not be used due to the low affinity of LTBP1-cbEGF14, so for each aliquot of FBN^{e2cb1} that was added, an equi-molar amount of calcium was added to compensate for the calcium bound by the high affinity site in fibrillin cbEGF1. In all protein titrations significant losses in peak intensity were seen with successive protein additions, and in order to collect spectra with sufficient signal to noise in the later titration points, long experiments were collected, often with 256 scans (see Table 8.9).

Titration data was then analysed using the CCPN analysis software to annotate and measure peak intensities and peak positions, and then Microsoft Excel was used to

calculate peak shifts and intensity changes. To analyse peak shifts, the peak positions in each titration point spectrum were subtracted from the peak position in the previous titration point. These position differences between points (where peaks could be reliably determined) were then added together to get the total peak shift. This was done separately for both ^1H and ^{15}N dimensions, and the total shifts in each dimension were then squared and square rooted to give positive values and combined, (to combine both dimensions the shift in the ^{15}N dimension was divided by six and added to the shift in the ^1H dimension). Standard deviation bars were calculated based on the standard deviation of all recorded shifts for that peak, and combined, again dividing the standard deviation in the ^{15}N dimension by six. To analyse peak intensity changes during the titration, peak intensities in the first HSQC experiment without protein were divided by the peak intensities at each titration point. Errors bars represent the potential error introduced by background noise in the experiment, calculated as for the heteronuclear NOE experiments (Section 8.33 above).

8.35 NMR- HADDOCK

In order to model putative protein complexes consistent with the NMR titration data, the HADDOCK webserver was used [21]. The “expert level” webserver interface was employed in order to allow for the incorporation of fully flexible regions in protein structures during docking calculations. To test whether similar docking results were consistently obtained, multiple runs were carried out, utilising different cut offs for ambiguous interaction restraint determination, different histidine protonation states, and different flexible regions to suggest different levels of structural uncertainty and account for dynamics (as summarised in Tables 5.2 and 5.3). For all calculations carried out the default “expert” HADDOCK webserver docking parameters were used. Final structures were visualised and aligned using pymol, manual visual inspection

was then employed to class the high scoring structures into particular “models” depending on consistent domain orientations, specific salt bridges, and other features.

References

1. Schultz, J., et al., *SMART, a simple modular architecture research tool: identification of signaling domains*. Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5857-64.
2. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods*. Mol Biol Evol, 2011. **28**(10): p. 2731-9.
3. Caffrey, D.R., et al., *PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments*. BMC Bioinformatics, 2007. **8**: p. 381.
4. Eswar, N., et al., *Comparative protein structure modeling using Modeller*. Curr Protoc Bioinformatics, 2006. **Chapter 5**: p. Unit 5 6.
5. Jensen, S.A., et al., *Structure and interdomain interactions of a hybrid domain: a disulphide-rich module of the fibrillin/LTBP superfamily of matrix proteins*. Structure, 2009. **17**(5): p. 759-68.
6. Cordle, J., et al., *A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition*. Nat Struct Mol Biol, 2008. **15**(8): p. 849-57.
7. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes*. J Biomol NMR, 1995. **6**(3): p. 277-93.
8. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: development of a software pipeline*. Proteins, 2005. **59**(4): p. 687-96.
9. Jeener, B.M., P Bachmann, RR Ernst *Investigation of exchange processes by two-dimensional NMR spectroscopy*. The Journal of chemical physics, 1979. **71**(11): p. 4546-4553
10. Geoffrey Bodenhausen, D.J.R., *Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy*. Chemical Physics Letters, 1980. **69**(1): p. 185–189.
11. Weisemann, R., H. Ruterjans, and W. Bermel, *3D triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins*. J Biomol NMR, 1993. **3**(1): p. 113-20.
12. Xu, Y., D. Long, and D. Yang, *Rapid data collection for protein structure determination by NMR spectroscopy*. J Am Chem Soc, 2007. **129**(25): p. 7722-3.
13. Swapna, G.V. and G.T. Montelione, *Sensitivity-enhanced sim-CT HMQC PFG-HBHA(CO)NH and PFG-CBCA(CO)NH triple-resonance experiments*. J Magn Reson, 1999. **137**(2): p. 437-42.
14. Cordier, F., A.J. Dingley, and S. Grzesiek, *A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings*. J Biomol NMR, 1999. **13**(2): p. 175-80.
15. Otting, M.R.a.G., *Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments*. J. Am. Chem. Soc, 2000. **122**(32): p. 7793–7797.
16. Clore, G.M., A.M. Gronenborn, and A. Bax, *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, 1998. **133**(1): p. 216-21.
17. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package*. J Magn Reson, 2003. **160**(1): p. 65-73.
18. Folmer, R.H., et al., *Floating stereospecific assignment revisited: application to an 18 kDa protein and comparison with J-coupling data*. J Biomol NMR, 1997. **9**(3): p. 245-58.
19. Schubert, M., et al., *A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics*. J Biomol NMR, 2002. **24**(2): p. 149-54.
20. Kay, L.E., D.A. Torchia, and A. Bax, *Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease*. Biochemistry, 1989. **28**(23): p. 8972-9.
21. de Vries, S.J., M. van Dijk, and A.M. Bonvin, *The HADDOCK web server for data-driven biomolecular docking*. Nat Protoc, 2010. **5**(5): p. 883-97.

Appendix 1 – NMR assignments and other supplementary NMR data

Table A1.1 – Assigned resonances for fbn^{e2cb1}

		H	N	ND	NE	HA	HB	HG	HD	HE	HZ	HH	C	CA	CB	CG	CD	CE
111	Ser	-	-	-	-	-	-	-	-	-	-	-	170.81	57.5	63.28	-	-	-
112	Ala	8.72	124.31	-	-	4.42	1.41	-	-	-	-	-	177.63	52.64	19.35	-	-	-
113	Ser	8.39	115.23	-	-	4.43	3.88,3.88	-	-	-	-	-	174.61	58.38	63.86	-	-	-
114	Arg	8.42	122.62	-	-	4.4	1.77,1.89	1.64	3.2	-	-	-	176.24	56.13	30.91	27.06	43.29	-
115	Ser	8.35	116.82	-	-	4.46	3.83	-	-	-	-	-	174.45	58.26	63.8	-	-	-
116	Ile	8.18	122.14	-	-	4.18	1.84	1.17,0.84,1.4	0.85	-	-	-	176.07	61.15	38.73	27.28,17.50	13.04	-
117	Gln	8.39	123.56	-	-	4.34	2.05,1.94	2.34	-	-	-	-	175.62	55.6	29.69	33.84	-	-
118	His	8.59	119.9	-	-	4.71	3.16,3.24	-	-	-	-	-	173.82	55.43	29.5	-	-	-
119	Cys	8.51	120.5	-	-	4.71	3.10,3.34	-	-	-	-	-	174.53	56.43	41.55	-	-	-
120	Asn	9.14	121.92	112.16	-	4.82	2.69,2.86	-	7.68,6.85	-	-	-	174.57	53.54	39.23	-	-	-
121	Ile	7.22	117.95	-	-	4.19	1.61	1.03,0.73,1.4	0.65	-	-	-	173.89	60.65	40.27	27.68,17.06	13.69	-
122	Arg	8.45	123.92	-	-	4.31	1.68	1.57,1.44	3.12	-	-	-	175.75	55.64	30.35	27.13	43.29	-
123	Cys	8.61	123.68	-	-	4.69	2.37	-	-	-	-	-	173.68	53.19	39.38	-	-	-
124	Met	8.35	120.49	-	-	4.4	2.01,1.82	2.51	-	1.44	-	-	176.39	55.2	34.83	32.6	-	-
125	Asn	9.07	113.9	117.87	-	4.11	1.93,1.45	-	7.76,7.29	-	-	-	176.66	54.89	36.63	-	-	-
126	Gly	8.76	106.35	-	-	3.71,4.03	-	-	-	-	-	-	175.53	45.81	-	-	-	-
127	Gly	7.37	106.37	-	-	4.39,3.45	-	-	-	-	-	-	171.7	45.55	-	-	-	-
128	Ser	8.4	113.91	-	-	4.72	4.40,3.81	-	-	-	-	-	172.44	57.09	65.22	-	-	-
129	Cys	8.89	123.09	-	-	4.88	3.15,3.14	-	-	-	-	-	174.44	56.25	41.98	-	-	-
130	Ser	8.76	122.2	-	-	4.51	3.59,3.53	-	-	-	-	-	173.99	57.92	64.13	-	-	-
131	Asp	9.09	125.31	-	-	4.23	2.93,2.70	-	-	-	-	-	174.78	56.59	39.58	-	-	-
132	Asp	8.15	114.43	-	-	4.38	2.81,2.80	-	-	-	-	-	174.27	55.12	40.03	-	-	-
133	His	7.84	113.85	-	-	4.89	3.26	-	-	7.25	-	-	172.63	54.57	31.02	-	-	-
134	Cys	9.08	117.5	-	-	4.92	2.73,2.73	-	-	-	-	-	174.11	55.08	41.98	-	-	-

135	Leu	9.27	126.22	-	-	4.65	1.66	-	0.90,0.9	-	-	-	176.83	53.96	41.27	-	24.3	-
136	Cys	8.84	122.73	-	-	4.78	2.70,3.38	-	-	-	-	-	176.71	53.21	37.28	-	-	-
137	Gln	8.63	120.29	-	111.62	4.36	2.38,1.89	2.32,2.56	-	7.40,6.94	-	-	175.17	55.57	29.44	33.83	-	-
138	Lys	8.39	120.43	-	-	4.07	1.83,1.79	1.49,1.43	1.69	3.01	-	-	177.21	58.34	32.22	24.65	29.17	42.01
139	Gly	8.69	112.43	-	-	4.08,3.48	-	-	-	-	-	-	171.69	44.71	-	-	-	-
140	Tyr	7.76	115.59	-	-	5.25	2.61,2.78	-	6.71	6.68	-	-	175.01	57.31	41.92	-	133.58	117.8
141	Ile	9.16	115.74	-	-	4.77	1.97	0.79,1.26,1.78	-	-	-	-	175.06	60.07	42.84	17.51,25.65	13.99	-
142	Gly	7.86	105.61	-	-	5.05,3.89	-	-	-	-	-	-	177.45	43.72	-	-	-	-
143	Thr	8.73	118.24	-	-	3.83	3.86	0.99	-	-	-	-	174.92	66.03	69.21	21.81	-	-
144	His	8.79	113.68	-	-	-	-	-	-	-	-	-	174.96	54.11	28.87	-	-	-
145	Cys	7.93	112.83	-	-	4.08	3.45,3.16	-	-	-	-	-	175.05	54.95	36.31	-	-	-
146	Gly	8.73	104.57	-	-	3.81,3.9	-	-	-	-	-	-	174.05	45.98	-	-	-	-
147	Gln	8.36	120.8	-	111.76	5.1	1.94	2.27	-	7.68,6.99	-	-	172.61	52.12	30.45	33.33	-	-
148	Pro	-	-	-	-	4.69	1.56,1.30	1.66,1.30	3.93,4.34	-	-	-	175.85	61.64	31.37	27.04	50.6	-
149	Val	8.1	119.48	-	-	3.95	1.65	0.75,0.75	-	-	-	-	176.12	62.14	33.45	21.60,21.04	-	-
150	Cys	8.85	128.11	-	-	5.11	3.04,2.80	-	-	-	-	-	175.24	58.96	36.83	-	-	-
151	Glu	8.94	124.22	-	-	3.82	2.01,2.01	2.23,2.37	-	-	-	-	177.68	59.86	29.88	36.82	-	-
152	Ser	8.59	113.15	-	-	4.54	3.88,3.74	-	-	-	-	-	174.21	58.12	63.21	-	-	-
153	Gly	7.64	107.65	-	-	4.02,3.72	-	-	-	-	-	-	172.1	44.52	-	-	-	-
154	Cys	8.14	115.26	-	-	4.64	2.26,2.38	-	-	-	-	-	173.89	53.2	39.87	-	-	-
155	Leu	8.42	122.82	-	-	4.36	1.31,1.65	-	0.79,0.78	-	-	-	177.74	54.39	43.88	-	25.24,21.92	-
156	Asn	8.86	113.04	-	-	3.92	1.89,1.12	-	8	-	-	-	175.43	54.86	36.61	-	-	-
157	Gly	8.51	104.53	-	-	4.11,3.62	-	-	-	-	-	-	175.41	45.19	-	-	-	-
158	Gly	7.56	106.35	-	-	3.57,4.37	-	-	-	-	-	-	171.27	45.4	-	-	-	-
159	Arg	8.33	117.09	-	-	4.68	1.71,1.71	1.44,1.56	3.19	-	-	-	174.53	53.99	33.84	27.34	43.29	-
160	Cys	9.08	125.53	-	-	4.93	3.56,2.99	-	-	-	-	-	175.3	56.59	28.75	-	-	-
161	Val	8.59	121.29	-	-	4.57	2.37	0.96,0.82	-	-	-	-	175.04	61.92	33.5	22.19,18	-	-
162	Ala	8.68	126.07	-	-	4.33	1.44	-	-	-	-	-	172.02	50.95	18.49	-	-	-
163	Pro	-	-	-	-	4.27	2.17,1.87	1.72,1.96	3.46,3.51	-	-	-	176.95	64.78	31.09	24.36	50.12	-
164	Asn	9.07	120.61	112.62	-	4.46	3.17,2.85	-	7.51,7.11	-	-	-	173.27	54.92	37.87	-	-	-

165	Arg	8.4	121.05	-	-	-	1.60,1.46	-	3.19	-	-	-	173.94	55.72	32.87	-	43.43	-
166	Cys	8.85	123.07	-	-	5.33	2.62,2.76	-	-	-	-	-	174.92	54.02	43	-	-	-
167	Ala	9.71	126.5	-	-	4.79	1.43	-	-	-	-	-	177.31	51.31	20.29	-	-	-
168	Cys	8.85	122.24	-	-	5.02	3.50,2.71	-	-	-	-	-	176.4	52.88	38.6	-	-	-
169	Thr	8.73	114.71	-	-	4.48	4.54	1.29	-	-	-	-	173.88	61.57	70.08	22.19	-	-
170	Tyr	8.46	119.26	-	-	4.19	3.25,2.86	-	7.17	6.91	-	-	176.29	61.18	38.14	-	133.19	118.15
171	Gly	7.21	113.58	-	-	2.05,3.51	-	-	-	-	-	-	170.96	44.46	-	-	-	-
172	Phe	7.59	115.27	-	-	5.56	2.89,3.00	-	6.96	7.11	7.04	-	175.63	56.39	42.22	-	-	-
173	Thr	9.25	114.59	-	-	4.92	4.21	1.15	-	-	-	-	172.95	59.9	70.29	19.42	-	-
174	Gly	7.91	108.5	-	-	4.93,3.97	-	-	-	-	-	-	174.59	43.71	-	-	-	-
175	Pro	-	-	-	-	4.41	2.40,2.10	2.1	3.64	-	-	-	178.1	65.24	31.83	28.34	49.81	-
176	Gln	9.04	115.51	-	110.75	4.96	2.39,1.95	2.32,2.10	-	6.75,7.33	-	-	175.75	55.17	30.36	35.48	-	-
177	Cys	8.05	113.31	-	-	43.57,3.18	-	-	-	-	-	-	174.75	55.23	36.54	-	-	-
178	Glu	10.51	118.32	-	-	4.05	2.09,1.92	2.49,2.20	-	-	-	-	176.74	57.47	29.68	35.9	-	-
179	Arg	8.99	123.77	-	-	4.61	1.70,1.99	1.60,1.46	-	-	-	-	173.82	55.68	31	27.44	-	-
180	Asp	8.44	124.79	-	-	4.62	2.82,2.50	-	-	-	-	-	176.5	52.52	41.95	-	-	-
181	Tyr	8.64	123.38	-	-	4.78	2.63,3.18	-	6.86	6.55	-	-	175.8	56.05	38.31	-	133.65	118
182	Arg	8.13	122.76	-	-	4.18	2.02	-	-	-	-	-	177.33	57.44	31.57	-	-	-
183	Thr	8.3	109.69	-	-	5.04	4.01	1.16	-	-	-	-	174.56	59.93	71.78	22.13	-	-
184	Gly	8.7	106.97	-	-	3.33,4.16	-	-	-	-	-	-	168.79	45	-	-	-	-
185	Pro	-	-	-	-	-	-	1.59,1.79	2.90,3.36	-	-	-	-	-	-	27.06	48.87	-
186	Cys	-	-	-	-	5.23	2.71	-	-	-	-	-	173.53	53.38	38.7	-	-	-
187	Phe	9.03	127.66	-	-	5.01	2.69,3.42	-	7.07	7.25	7.01	-	176.6	58.04	40.11	-	-	-
188	Thr	9.26	115.89	-	-	4.53	4.65	1.28	-	-	-	-	173.56	62.85	69.04	23.3	-	-
189	Val	7.69	116.96	-	-	4.43	20.90,0.90	-	-	-	-	-	173.77	60.37	35.96	20.81,	20.77	-
190	Ile	8.46	121.86	-	-	4.46	1.78	2	0.77,0.92,1.3	0.78	-	-	175.87	60.14	39.18	18.06,27.13	14.1	-
191	Ser	8.73	120.51	-	-	4.75	-	-	-	-	-	-	-	58.35	-	-	-	-
192	Asn	8.57	122.54	112.61	-	4.89	2.63,2.91	-	6.80,7.8	-	-	-	-	-	39.63	-	-	-
193	Gln	7.29	116.91	-	-	4.33	1.68	-	-	-	-	-	-	-	-	-	-	-
195	Cys	8.54	122.71	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

196	Gln	8	119.82	-	111.61	4.04	1.96,2.25	2.45,2.32	-	6.89,7.68	-	-	176.72	58.41	28.57	33.82	-	-
197	Gly	8.97	114.11	-	-	4.29,3.65	-	-	-	-	-	-	174.1	45.1	-	-	-	-
199	Leu	-	-	-	-	4.66	1.51,1.65	-	0.87,0.86	-	-	-	-	53.24	43.5	-	23.5	-
200	Ser	-	-	-	-	4.57	3.69,3.26	-	-	-	-	-	174.71	57.42	64.87	-	-	-
201	Gly	8.74	112.59	-	-	4.15,3.82	-	-	-	-	-	-	173.34	45.26	-	-	-	-
202	Ile	8.01	112.28	-	-	4.33	1.75	0.67,1.52,0.8	0.7	-	-	-	174.62	60.33	39.25	18.05,	26.42	13.56
203	Val	8.13	127.02	-	-	4.53	1.83	0.73,0.72	-	-	-	-	175.45	60.84	33.82	20.97,20.5	-	-
204	Ser	8.92	117.13	-	-	5.05	3.50,4.17	-	-	-	-	-	176.22	57.64	67.6	-	-	-
205	Thr	8.59	112.87	-	-	5.02	4.5	1.24	-	-	-	-	175.77	60.92	70.54	21.82	-	-
206	Lys	8.8	126.53	-	-	1.99	0.81,-0.9	1.28,1.41	0.53,0.83	2.63,2.83	-	-	178.45	59.78	30.65	29.48	25.63	42.06
207	Thr	8.02	112.01	-	-	3.62	3.83	1.11	-	-	-	-	176.75	65.97	68.87	21.43	-	-
208	Leu	7.16	119.01	-	-	3.87	1.29,1.70	1.47	0.87,0.78	-	-	-	178.75	57.88	41.62	27.01	24.34	25.0
209	Cys	7.85	116.14	-	-	4.17	2.88,3.35	-	-	-	-	-	177.35	58.49	38.29	-	-	-
210	Cys	8.66	113.74	-	-	4.63	2.72,3.24	-	-	-	-	-	175.46	53.56	33.81	-	-	-
211	Ala	7.71	118.45	-	-	4.4	1.51	-	-	-	-	-	177.21	53.5	20.41	-	-	-
212	Thr	7.29	105.08	-	-	4.83	4.29	1.11	-	-	-	-	176.69	59.2	70.21	21.41	-	-
213	Val	9.23	114.59	-	-	4.51	2.53	1.14,0.74	-	-	-	-	176.98	60.96	33	17.94,	22.28	-
214	Gly	8.56	107.22	-	-	3.89	-	-	-	-	-	-	174.34	46.42	-	-	-	-
215	Arg	9.5	119.32	-	-	4.53	1.91,1.41	1.69	3.26	-	-	-	176.7	55.8	31.96	28.99	42.91	-
216	Ala	8.53	122.15	-	-	5.27	1.49	-	-	-	-	-	174.74	52.22	23.22	-	-	-
217	Trp	9.2	120.54	-	127.94	5.61	3.33,2.88	-	6.68	7.35,9.51	7.12,6.91	6.86	175.85	56.79	34.3	-	-	-
218	Gly	8.14	113.05	-	-	3.69,2.21	-	-	-	-	-	-	171.95	44.09	-	-	-	-
219	His	7.61	113.15	-	-	4.82	2.89,3.08	-	7.27	-	-	-	172.85	53.29	31.47	-	119.99	-
220	Pro	-	-	-	-	4.67	2.05,2.38	1.94,1.84	3.40,3.61	-	-	-	175.87	62.89	34.29	25.38	50.22	-
221	Cys	8.08	118.73	-	-	-	-	-	-	-	-	-	173.8	56.1	41.72	-	-	-
222	Glu	9.53	123.58	-	-	4.65	2.18,2.20	2.50,2.53	-	-	-	-	176.16	54	32.39	36	-	-
223	Met	8.97	121.17	-	-	4.61	2.06	2.66,2.88	-	-	-	-	177.88	55.47	32.48	32.6	-	-
224	Cys	8.35	122.39	-	-	-	-	-	-	-	-	-	173.11	52.17	-	-	-	-
225	Pro	-	-	-	-	4.47	2.40,1.76	2.05	3.52	-	-	-	177.13	62.76	32	27.65	50.52	-
226	Ala	8.94	126.69	-	-	4.1	1.54	-	-	-	-	-	178.81	55.46	18.53	-	-	-

227	Gln	7.96	113.89	-	110.33	4.88	1.79,2.19	2.36,2.30	-	7.42,6.89	-	-	173.63	52.29	30.01	32.85	-	-
228	Pro	-	-	-	-	4.56	1.91,2.24	-	3.75,3.53	-	-	-	174.8	62.01	30.48	-	50.67	-
229	His	8.02	119.15	-	-	4.58	3.1	-	-	-	-	-	173.45	54.75	30.41	-	-	-
230	Pro	-	-	-	-	4.17	2.17,1.93	-	-	-	-	-	63.97	-	-	-	-	-
231	Cys	-	-	-	-	4.88	2.66,3.23	-	-	-	-	-	173.51	52.89	40.83	-	-	-
232	Arg	8.21	119.08	-	-	4.46	1.78,1.99	-	3.16	-	-	-	175.26	55.4	31.38	-	-	-
233	Arg	8.51	117.25	-	-	4.14	1.99,1.93	-	-	-	-	-	176.71	58.57	30.39	-	-	-
234	Gly	9.32	108.43	-	-	3.23,4.64	-	-	-	-	-	-	173.86	44.5	-	-	-	-
235	Phe	8.41	118.26	-	-	5.28	3.26,2.58	-	6.94	7.3	7.15	-	173.34	56.17	42.93	-	132.21	132.78
236	Ile	8.8	112.55	-	-	4.79	1.68	0.94,0.71,1.3	0.71	-	-	-	174.33	57.99	41.77	17.75,26.21	13.66	-
237	Pro	-	-	-	-	4.45	1.70,2.04	2.11,1.87	3.71,3.62	-	-	-	174.01	62.73	31.52	27.66	51.08	-
238	Asn	8.86	123.57	114.11	-	4.58	2.72,3.04	-	7.11,7.92	-	-	-	177.7	53.85	39.98	-	-	-
239	Ile	8.71	125.31	-	-	3.94	1.94	1.30,0.96,1.4	0.9	-	-	-	176.07	63.69	38.26	17.69,27.79	13.65	-
240	Arg	8.21	118.63	-	-	4.4	1.97,1.97	1.60,1.72	3.18	-	-	-	178.26	57.97	31.03	27.56	43.78	-
241	Thr	8.02	106.09	-	-	4.4	4.41	1.14	-	-	-	-	176.68	61.69	70.91	21.01	-	-
242	Gly	8.35	110.78	-	-	4.22,3.68	-	-	-	-	-	-	172.96	45.46	-	-	-	-
243	Ala	7.67	121.67	-	-	4.31	1.31	-	-	-	-	-	177.85	51.75	19.7	-	-	-
244	Cys	8.74	119.9	-	-	4.54	2.75,2.57	-	-	-	-	-	174.75	56.85	41.11	-	-	-
245	Gln	9.16	123.98	-	111.52	4.94	2.15,2.16	2.36,2.52	-	7.47,7.02	-	-	174.45	54.27	32.44	33.54	-	-
246	Asp	9.01	122.77	-	-	4.2	4.49,2.31	-	-	-	-	-	175.62	55.54	43.36	-	-	-
247	Val	7.63	125.19	-	-	3.83	1.51	0.67,0.54	-	-	-	-	173.65	60.95	33.47	22.34,20.45	-	-
248	Asp	8.89	127.64	-	-	4.9	2.98,2.32	-	-	-	-	-	175.88	51.09	38.73	-	-	-
249	Glu	10.13	126.09	-	-	3.22	1.28,1.09	2.29	-	-	-	-	176	63.12	28.33	38.66	-	-
250	Cys	7.19	109.59	-	-	4.17	3.14,3.17	-	-	-	-	-	176.86	54.31	36.3	-	-	-
251	Gln	7.15	115.75	-	111.71	4.22	2.05,1.76	2.37,2.19	-	7.60,6.76	-	-	177.22	56.5	30.29	34.1	-	-
252	Ala	8.48	119.06	-	-	4.17	1.27	-	-	-	-	-	177.28	53.8	20.34	-	-	-
253	Ile	7.61	116.12	-	-	4.45	2.08	1.32,0.92,0.6	0.68	-	-	-	172.2	57.83	37	18.36,26.89	13.1	-
254	Pro	-	-	-	-	4.33	1.92,2.29	2.12,1.94	3.68	-	-	-	177.32	63.97	31.84	27.78	-	-
255	Gly	8.53	108.74	-	-	3.63,4.27	-	-	-	-	-	-	174.7	45.16	-	-	-	-

256	Leu	7.36	119.7	-	-	4.16	1.43,1.51	1.53	0.42,0.3	-	-	-	178.24	55.67	42.3	26.32	25.22,24.63
257	Cys	9.18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
260	Gly	-	-	-	-	4.64,3.76	-	-	-	-	-	-	172.44	45.02	-	-	-
261	Asn	8.52	118.09	110.85	-	5.12	5.58,2.76	-	6.86,7.48	-	-	-	172.64	52.15	41.04	-	-
262	Cys	8.97	126.25	-	-	5.36	3.23,2.51	-	-	-	-	-	174.03	57.01	37.98	-	-
263	Ile	9.66	131.18	-	-	4.39	1.89	1.15,0.84,1.4	0.83	-	-	-	174.99	60.38	40.77	17.52,27.49	13.56
264	Asn	9.28	126.22	120.62	-	4.95	2.65,2.98	-	8.07	-	-	-	175.01	54.87	39.66	-	-
265	Thr	7.82	116.07	-	-	4.73	4.25	1	-	-	-	-	174.12	59.5	70.91	21.8	-
266	Val	8.84	123.22	-	-	3.59	2.13	1.06,1.03	-	-	-	-	176.47	64.88	29.85	21.67,	22.57
267	Gly	9.09	120.53	-	-	2.91,4.23	-	-	-	-	-	-	173.62	46.89	-	-	-
268	Ser	6.63	112.1	-	-	4.37	3.41,3.86	-	-	-	-	-	173.77	55.93	63.1	-	-
269	Phe	9.14	117.01	-	-	5.07	2.99,3.74	-	6.9	7.21	7.18	-	172.59	57.26	39.77	-	132.62
270	Glu	9.77	119.76	-	-	4.49	1.85,1.93	2.09,2.08	-	-	-	-	174.52	54.66	33.41	35.67	-
271	Cys	8.88	119.84	-	-	5.61	2.75	-	-	-	-	-	52.62	36.82	-	-	-
272	Lys	-	-	-	-	4.71	1.69,1.84	1.36,1.49	1.68	2.96	-	-	176.58	54.34	33.23	23.19	29.09 42.08
273	Cys	9.12	124.55	-	-	5.19	2.59,2.78	-	-	-	-	-	172.25	51.64	36.55	-	-
274	Pro	-	-	-	-	4.49	1.73,2.37	1.72,2.01	3.21,3.94	-	-	-	175.25	62.69	32.26	27.69	50.8
275	Ala	8.43	121.92	-	-	4.18	1.4	-	-	-	-	-	179.34	53.73	18.11	-	-
276	Gly	8.9	110.44	-	-	4.14,3.64	-	-	-	-	-	-	173.85	45.07	-	-	-
277	His	8.24	117.5	-	-	4.93	3.08,2.89	-	-	-	-	-	172.82	54.32	30.82	-	-
278	Lys	9.01	119.5	-	-	4.59	1.75,1.76	1.37,1.39	1.67,1.68	3.00,3.00	-	-	175.37	55.17	35.8	24.36	29.08 41.97
279	Leu	8.71	125.85	-	-	4.21	1.40,1.82	1.16	0.87,0.63	-	-	-	176.13	55.51	42.71	27.13,33	25.99,23
280	Asn	8.75	127.54	113.39	-	4.78	2.69,3.44	-	8.12,7	-	-	-	176.3	52.68	38.61	-	-
281	Glu	9.23	125.54	-	-	3.92	2.01,2.15	2.33,2.28	-	-	-	-	177.55	58.84	29.63	36.38	-
282	Val	8.26	118.76	-	-	3.9	2.27	0.96,1.04	-	-	-	-	177.4	65.3	31.93	22.30,21.08	-
283	Ser	7.61	111.3	-	-	4.41	3.79,3.93	-	-	-	-	-	175.07	57.99	64.86	-	-
284	Gln	8.64	115.04	-	-	3.59	2.82,2.24	-	-	-	-	-	173.21	57.12	25.31	-	-
285	Lys	7.48	114.09	-	-	4.76	1.80,1.67	1.43,1.32	-	2.99	-	-	176.34	54.82	35.81	25.19	41.86
286	Cys	8.56	120.26	-	-	4.74	2.56,2.85	-	-	-	-	-	174.01	55.31	39.43	-	-
287	Glu	8.87	129.49	-	-	4.37	2.00,2.05	2.16,2.29	-	-	-	-	181.29	58.08	32.15	36.81	-

Figure A1.1 – $\text{fbn}^{\text{e2cb1}}$ amide peak intensities

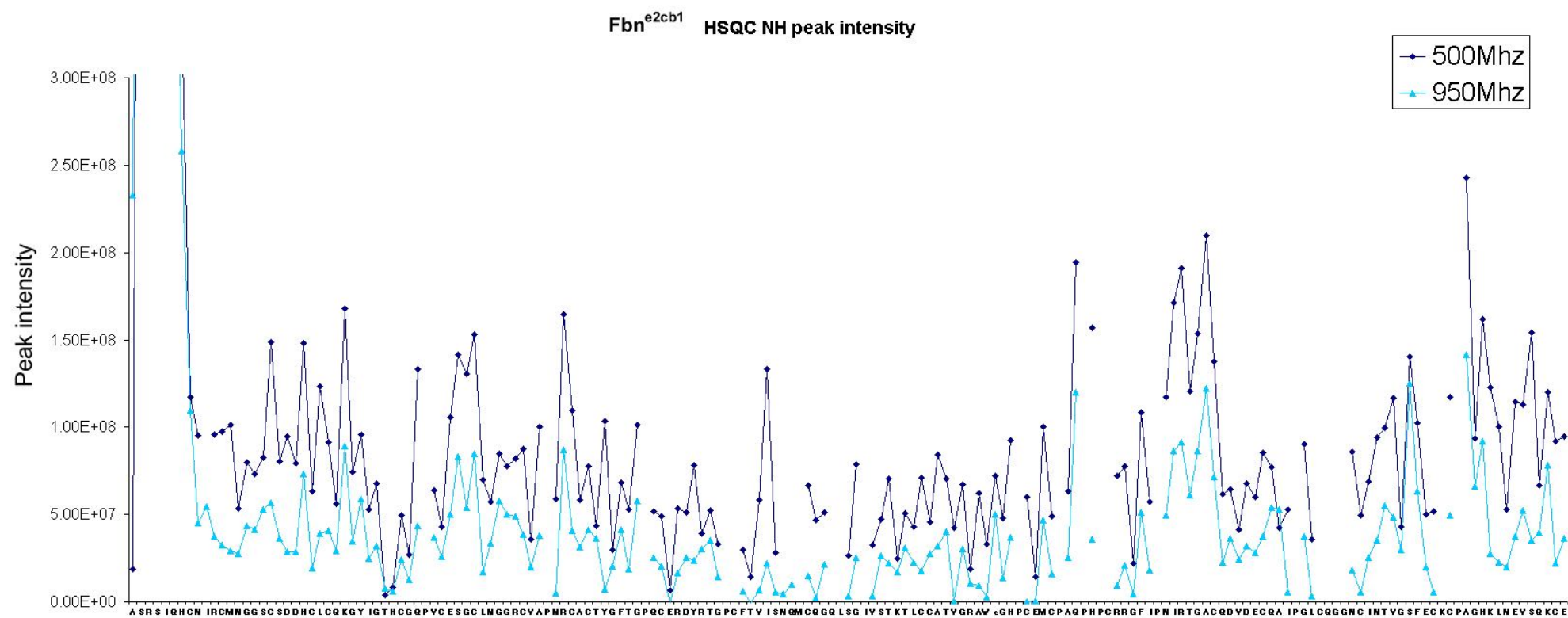


Table A1.2 – RDC values collected for fbn^{e2cb1} residues in 2.2% PEG -hexanol

Residue number	NH RDC in 2.2% PEG-hexanol
119	-21.46
120	-22.416
121	15.243
122	-7.468
125	-34.783
126	-30.498
127	-1.927
130	9.352
131	0.215
133	-0.709
135	-13.324
137	-5.684
139	-0.386
141	13.698
142	-33.694
146	-11.708
151	-21.143
152	-34.819
153	32.688
157	-25.044
158	-13.167
159	0.372
164	-6.994
167	-7.841
172	6.067
174	-37.842
182	-6.087
183	-27.497
184	-5.01
188	7.959
189	-14.689
191	-5.53
195	2.899
197	2.478
203	-17.829
204	-14.691
205	-11.378
208	-14.198
209	-1.958
210	15.038
212	-10.753
213	-4.85
214	11.758
216	-5.286
217	10.12
217ne	-15.4129
218	9.509
222	-17.283
223	8.63
226	0.525

227	21.569
229	-18.666
232	-3.051
233	-18.993
234	-20.865
235	-15.968
236	-7.396
239	3.838
241	17.988
242	-12.07
243	-15.936
244	-22.759
245	-5.73
246	-2.684
247	-11.424
248	19.413
249	-25.849
250	-18.501
253	7.619
255	-3.764
256	24.789
261	-2.183
263	-12.28
268	-12.621
269	-5.467
270	-13.105
271	-24.883
273	-22.156
277	-12.955
278	-17.778
280	-29.036
284	-3.998
286	-18.199
287	-22.023

Table A1.3 - Assigned resonances for LTBP1^{cb14TB3}

88% of backbone amides, 85% H^αprotons, and 45% of H^β protons assigned.

		H	N	NE	HA	HB	HG	HD	HE	HZ	HH
1468	Asp	8.4	119.44	-	4.59	2.47,2.73	-	-	-	-	-
1469	Met	8	122.38	-	4.46	-	-	-	-	-	-
1470	Asp	8.81	123.88	-	4.84	2.48,2.84	-	-	-	-	-
1471	Glu	9.64	126.5	-	3.8	1.27	-	-	-	-	-
1472	Cys	8.32	115.87	-	4.53	3.23	-	-	-	-	-
1473	Gln	7.43	117.91	-	4.06	-	-	-	-	-	-
1474	Asp	7.39	120.35	-	4.98	2.99,2.44	-	-	-	-	-
1476	Ser	8.53	113.83	-	4.32	3.94	-	-	-	-	-
1477	Ser	7.82	117.32	-	4.11	3.36,3.64	-	-	-	-	-
1480	Asp	9.1	121.71	-	4.34	3.04,2.91	-	-	-	-	-
1481	Gly	7.17	105.5	-	3.66,4.72	-	-	-	-	-	-
1482	Gln	8.8	119.98	-	4.62	-	-	-	-	-	-
1483	Cys	8.94	127.05	-	5.28	2.67,3.12	-	-	-	-	-

1484	Val	9.41	129.67	-	4.39	2.05	1.02,1.02	-	-	-	-
1485	Asn	9.23	126.96	-	4.88	3.00,2.66	-	-	-	-	-
1486	Thr	8.08	117.24	-	4.68	4.29	1.04	-	-	-	-
1487	Glu	8.83	122.46	-	4.01	-	-	-	-	-	-
1488	Gly	9.02	115.27	-	4.18,3.88	-	-	-	-	-	-
1489	Ser	7.05	114.7	-	4.61	3.89,3.58	-	-	-	-	-
1490	Tyr	8.98	119.59	-	5.04	3.58,2.94	-	6.89	6.66	-	-
1491	Asn	9.4	118.12	-	4.77	2.06,2.27	-	-	-	-	-
1492	Cys	8.34	120	-	5.33	2.7	-	-	-	-	-
1493	Phe	8.95	124.14	-	4.88	2.96,3.16	-	-	-	-	-
1494	Cys	8.79	121.74	-	5.12	2.67,2.35	-	-	-	-	-
1495	Thr	9.15	121.36	-	4.28	3.99	1.31	-	-	-	-
1496	His	8.99	128.85	-	4.75	3.09,3.24	-	-	-	-	-
1497	Pro	-	-	-	3.73	-	-	-	-	-	-
1498	Met	8.87	123.2	-	4.54	2.18	2.67	-	-	-	-
1499	Val	8.81	113.5	-	4.52	2.12	0.89,0.74	-	-	-	-
1502	Ala	8.04	115.74	-	-	1.4	-	-	-	-	-
1503	Ser	7.11	117.86	-	4.88	-	-	-	-	-	-
1504	Glu	8.33	123.69	-	5.13	-	-	-	-	-	-
1505	Lys	9.68	118.29	-	4.67	-	-	-	-	-	-
1511	Ala	8.39	123.76	-	4.25	1.39	-	-	-	-	-
1512	Glu	8.42	119.66	-	4.32	-	-	-	-	-	-
1513	Ser	8.3	116.67	-	4.45	-	-	-	-	-	-
1514	Asn	8.51	120.96	-	-	2.79	-	-	-	-	-
1515	Glu	8.35	120.79	-	4.26	-	-	-	-	-	-
1516	Gln	8.35	121.07	-	4.33	-	-	-	-	-	-
1517	Ile	8.14	122.21	-	4.15	-	-	-	-	-	-
1518	Glu	8.48	124.51	-	4.32	-	-	-	-	-	-
1519	Glu	8.5	122.35	-	4.32	-	-	-	-	-	-
1520	Thr	8.14	114.92	-	4.34	-	1.19	-	-	-	-
1521	Asp	8.41	122.55	-	4.71	-	-	-	-	-	-
1522	Val	7.95	118.66	-	4.19	-	-	-	-	-	-
1523	Tyr	8.27	123.93	-	4.68	-	-	-	-	-	-
1524	Gln	8.05	120.52	-	5.01	-	-	-	-	-	-
1525	Asp	9.41	119.96	-	4.7	-	-	-	-	-	-
1526	Leu	9.4	121.49	-	4.14	-	-	-	-	-	-
1527	Cys	9.21	120.56	-	5.32	3.59,2.64	-	-	-	-	-
1528	Trp	8.73	127.99	130.38	4.65	2.85,3.41	-	7.15	9.58	7.54	7.16
1529	Glu	9.25	121.65	-	4.34	2.38	-	-	-	-	-
1530	His	8.22	114.25	-	5.03	2.78,3.20	-	-	-	-	-
1531	Leu	8.06	126.27	-	4.74	-	-	-	-	-	-
1532	Ser	8.45	119.57	-	4.51	-	-	-	-	-	-
1533	Asp	8.9	120.5	-	4.2	2.67	-	-	-	-	-
1534	Glu	7.51	116.31	-	4.17	2.17,1.69	-	-	-	-	-
1535	Tyr	7.74	115.15	-	3.77	3.24,3.15	-	6.97	6.85	-	-
1536	Val	7.4	117.35	-	4.04	-	-	-	-	-	-
1537	Cys	8.45	129.98	-	4.48	-	-	-	-	-	-
1538	Ser	8.92	115.45	-	4.62	3.95,3.88	-	-	-	-	-
1539	Arg	9.18	113.69	-	4.23	-	-	-	-	-	-
1541	Leu	8.46	122.97	-	4.34	-	-	-	-	-	-
1542	Val	8.07	125.08	-	3.66	-	-	-	-	-	-

1543	Gly	8.62	113.78	-	4.00,3.66	-	-	-	-	-	-
1544	Lys	7.03	118.25	-	4.51	-	-	-	-	-	-
1545	Gln	8.58	123.83	-	4.35	-	-	-	-	-	-
1546	Thr	9.1	110.81	-	5.33	-	-	-	-	-	-
1547	Thr	9.09	115.89	-	5.06	-	-	-	-	-	-
1548	Tyr	8.63	124.91	-	4.68	-	6.74	6.56	-	-	-
1549	Thr	7.8	113.27	-	3.28	3.86	-	-	-	-	-
1550	Glu	7.49	120.36	-	3.75	-	-	-	-	-	-
1551	Cys	8.33	114.47	-	-	-	-	-	-	-	-
1552	Cys	8.42	115.03	-	4.63	-	-	-	-	-	-
1553	Cys	8.48	115.34	-	4.8	2.66,2.85	-	-	-	-	-
1554	Leu	7.65	116.64	-	4.53	-	-	-	-	-	-
1555	Tyr	7.61	115.83	-	5.14	3.28,3.11	-	7.04	6.77	-	-
1556	Gly	9.02	108.51	-	3.89,4.82	-	-	-	-	-	-
1557	Glu	8.43	115.39	-	4.4	-	-	-	-	-	-
1558	Ala	8.31	120.99	-	5.15	1.46	-	-	-	-	-
1559	Trp	9.77	119.96	130.58	5.66	3.04,3.40	-	6.86	10.36,7.16	7.28,7.52	7.1
1560	Gly	9.13	114.61	-	3.06,3.62	-	-	-	-	-	-
1561	Met	8.76	121.15	-	4.55	2.17	-	-	-	-	-
1562	Gln	8.99	116.24	-	4.41	-	-	-	-	-	-
1563	Cys	8.23	122.75	-	3.66	3.38,2.36	-	-	-	-	-
1564	Ala	8.84	123	-	4.54	1.77	-	-	-	-	-
1565	Leu	8.51	122.12	-	4.21	-	-	-	-	-	-
1566	Cys	7.63	121.41	-	5.12	2.58,2.99	-	-	-	-	-
1568	Leu	8.6	121.29	-	4.18	-	-	-	-	-	-
1569	Lys	8.59	122.8	-	-	-	-	-	-	-	-
1570	Asp	7.58	115.56	-	4.52	2.81,2.69	-	-	-	-	-
1571	Ser	7.9	115.1	-	4.66	-	-	-	-	-	-
1572	Asp	9.2	124.04	-	4.51	2.72	-	-	-	-	-
1573	Asp	8.39	118.43	-	4.37	2.68	-	-	-	-	-
1574	Tyr	7.97	122.23	-	3.71	3.13,2.84	-	6.89	6.84	-	-
1575	Ala	7.7	121.38	-	3.97	1.61	-	-	-	-	-
1576	Gln	7.88	114.89	-	4.03	-	-	-	-	-	-
1577	Leu	7.68	120.43	-	4.03	-	0.55,0.95	-	-	-	-
1578	Cys	7.97	111.72	-	4.61	1.90,2.97	-	-	-	-	-
1579	Asn	7.76	119.47	-	4.58	2.80,2.92	-	-	-	-	-
1580	Ile	8	121.31	-	4.43	-	-	-	-	-	-
1581	Pro	-	-	-	4.44	2.24,1.94	-	-	-	-	-
1582	Val	7.73	124.09	-	4.02	2.07	0.9	-	-	-	-

Table A1.4- Assigned resonances for LTBP1^{E3cb15}

90% backbone amides, 81% carbonyl carbons, 94% α -carbons, 87% β -carbons, assigned.

		H	N	C	CA	CB
1615	Ser	-	-	174.34	58.44	63.95
1616	Ala	8.58	125.9	177.6	52.72	19.55
1617	Ser	8.29	115.57	174.44	58.26	64.14
1618	Phe	8.29	122.3	176.07	58.76	39.49
1619	Glu	8.4	121.81	176.76	57.42	30.25

1620	Glu	8.23	122.12	176.88	57.13	30.21
1621	Leu	8.22	123.33	177.76	55.42	42.44
1622	Gln	8.34	121.23	176.15	56.11	29.44
1623	Ala	8.23	125.02	178.2	53.3	19.24
1624	Glu	8.41	119.87	176.78	57.14	30.22
1625	Glu	8.25	121.33	176.07	-	-
1626	Cys	8.19	119.04	-	-	-
1627	Gly	-	-	174.00	46.31	-
1628	Ile	7.44	120.79	176.47	61.32	38.76
1629	Leu	8.21	126.45	177.29	56.33	-
1630	Asn	8.37	117.97	175.31	53.92	38.6
1631	Gly	7.97	107.49	173.44	45.6	-
1632	Cys	8.48	115.99	173.86	55.04	-
1633	Glu	8.03	121.89	176.18	57.57	-
1634	Asn	8.65	117.45	172.21	54.35	36.57
1635	Gly	6.84	104.6	170.55	45.75	-
1636	Arg	7.92	118.04	174.71	53.8	33.5
1637	Cys	9.05	125.93	173.96	56.37	42.75
1638	Val	8.95	125.93	174.09	60.56	35.71
1639	Arg	8.63	125.61	176.18	56.03	-
1640	Val	8.21	122.47	175.93	60.24	33.98
1641	Gln	8.72	123.02	177.05	59.48	27.7
1642	Glu	8.52	117.43	175.76	56.81	29.27
1643	Gly	7.62	109.66	171.81	45.27	-
1644	Tyr	8.41	120.85	174.83	57.39	42.05
1645	Thr	9.12	116.29	172.15	59.57	69.87
1646	Cys	7.73	118.08	173.19	52.27	40.37
1647	Asp	9.17	126.12	176.65	52.83	41.74
1648	Cys	8.47	124.01	175.45	53.74	37.52
1649	Phe	8.47	123.07	175.32	55	38.53
1650	Asp	8.6	120.48	176.62	56.56	40.86
1651	Gly	8.74	111.14	173.48	44.69	-
1652	Tyr	8.42	118.05	174.67	56.98	41.97
1653	His	9.78	116.65	173.55	53.6	32.18
1654	Leu	8.76	126.65	176.19	56.25	-
1655	Asp	-	128.33	-	53.06	-
1656	Thr	8.56	118.06	-	64.7	-
1661	Cys	-	-	174.11	55.52	42.43
1662	Val	9.38	121	175.84	59.63	35.2
1663	Asp	9.1	124.31	176.05	56.35	43.16
1664	Val	7.8	127.13	174.07	62.01	33.42
1665	Asn	9.13	126.92	175.83	50.39	36.56
1666	Glu	9.78	127.48	176.59	61.71	28.25
1667	Cys	7.51	111.93	175.71	53.14	35.76
1668	Asp	7.05	119.78	176.6	54.66	41.52
1669	Glu	8	121.15	177.33	57.7	30.45
1670	Leu	8.18	121.58	178.16	56.43	42.04
1671	Asn	8.29	117.9	175.76	54.58	38.9
1672	Asn	8.21	117.75	175.9	54.02	38.38
1673	Arg	8.05	119.58	176.65	57.74	30.08
1674	Met	8.16	117.95	175.86	55.63	32.62

1675	Ser	8.11	115.8	174.73	58.32	63.69
1676	Leu	8.08	124.59	177.32	57.13	42.68
1677	Cys	8	114.78	174.01	52	39.97
1678	Lys	8.4	123.15	175.33	55.76	34.33
1679	Asn	9.03	122.88	172.95	53.97	36.47
1680	Ala	7.13	115.76	174.42	50.81	24.15
1681	Lys	8.86	121.09	174.47	54.43	35.1
1682	Cys	8.74	126.59	174.18	57.92	36.03
1683	Ile	9.68	132.18	174.68	59.61	39.49
1684	Asn	9.04	127.61	175.08	54.37	39.52
1685	Thr	7.9	117.44	173.75	59.67	70.85
1686	Asp	8.73	123.77	175.59	55.72	40.3
1687	Gly	9.23	119.85	174.22	46.37	-
1688	Ser	6.64	113.32	173.55	56.18	63.04
1689	Tyr	9.04	118.7	172.7	57.48	39.12
1690	Lys	9.66	120.19	174.13	54.61	36.5
1691	Cys	8.52	119.06	173.9	52.16	39.19
1692	Leu	9.08	127.02	176.15	53.2	43.6
1693	Cys	8.71	124.09	175.22	53.8	37.88
1694	Leu	7.66	124.18	173.69	53.73	40.71
1695	Pro	-	-	178.08	64.37	31.6
1696	Gly	8.84	112.13	173.5	45	-
1697	Tyr	8.44	120.93	174.55	56.44	40.88
1698	Val	9.39	114.71	173.77	58.72	33.31
1699	Pro	-	-	176.79	64.6	32.56
1700	Ser	8.53	119.63	173.49	57.36	66.74
1701	Asp	8.72	118.38	175.99	55.32	40.07
1702	Lys	8.51	122.87	174.96	52.93	33.47
1703	Pro	-	-	176.53	64.29	32.06
1704	Asn	8.29	114.4	173.4	54.1	37.64
1705	Tyr	7.75	120.13	175.29	56.68	40.36
1706	Cys	7.72	125.07	172.12	56.64	46.98
1707	Thr	9.62	118.84	172.1	58.59	71.67
1708	Pro	-	-	177.27	63.59	31.8
1709	Leu	8.07	124.25	176.96	56.16	42.9
1710	Asn	8.46	119.91	175.54	53.28	38.6
1711	Thr	8.04	115.13	174.51	62.29	69.54
1712	Ala	8.16	126.19	177.71	52.81	19.18
1713	Leu	7.98	120.85	177.02	55.38	42.57
1714	Asn	8.31	119.58	175.03	53.3	38.71
1715	Leu	8.19	123.1	177.53	55.34	42.42
1716	Glu	8.33	121.75	176.47	56.72	30.2
1717	Lys	8.24	122.6	176.33	56.37	33.42
1718	Asp	8.43	122.41	176.44	54.69	41.21
1719	Ser	8.23	116.56	174.4	58.71	64.1
1720	Asp	8.44	122.74	176.08	54.56	41.05
1721	Leu	8.05	122.46	176.58	55.32	42.63
1722	Glu	7.85	126.58	181.23	58.06	31.24

Table A1.5- Assigned resonances for LTBP1^{TB3E3}
78%- backbone amides assigned

		H	N	NE	HE
1507	Ser	8.22	116.15	-	-
1508	Ala	8.26	125.76	-	-
1509	Arg	8.24	121.3	-	-
1511	Ala	8.42	124.3	-	-
1512	Glu	8.46	119.74	-	-
1514	Asn	8.48	120.88	-	-
1515	Glu	8.33	120.8	-	-
1516	Gln	8.34	121.27	-	-
1517	Ile	8.12	122.41	-	-
1518	Glu	8.47	124.93	-	-
1519	Glu	8.5	122.78	-	-
1520	Thr	8.09	114.41	-	-
1521	Asp	8.44	122.62	-	-
1522	Val	7.93	118.05	-	-
1523	Tyr	8.22	124.23	-	-
1524	Gln	8.01	120.37	-	-
1525	Asp	9.4	120.06	-	-
1526	Leu	9.41	121.79	-	-
1527	Cys	9.23	120.85	-	-
1528	Trp	-	-	131.72	9.59
1528	Trp	8.7	129.02	-	-
1529	Glu	9.24	122.16	-	-
1530	His	8.19	114	-	-
1531	Leu	8.06	127.06	-	-
1533	Asp	8.89	120.61	-	-
1534	Glu	7.49	116.2	-	-
1535	Tyr	7.77	114.98	-	-
1536	Val	7.4	117.22	-	-
1538	Ser	8.95	115.27	-	-
1539	Arg	9.17	113.18	-	-
1541	Leu	8.46	123.49	-	-
1542	Val	8.08	125.69	-	-
1543	Gly	8.62	113.42	-	-
1544	Lys	7.02	118.29	-	-
1545	Gln	8.58	124.48	-	-
1546	Thr	9.08	110.16	-	-
1547	Thr	9.04	115.71	-	-
1548	Tyr	8.62	125.48	-	-
1549	Thr	7.83	112.94	-	-
1550	Glu	7.48	120.57	-	-
1551	Cys	8.33	114.17	-	-
1552	Cys	8.41	114.68	-	-
1553	Cys	8.46	114.97	-	-
1554	Leu	7.65	116.47	-	-
1555	Tyr	7.59	115.54	-	-
1556	Gly	8.98	107.47	-	-
1557	Glu	8.4	115.28	-	-
1559	Trp	9.76	120	131.65	10.36

1560	Gly	9.15	115.08	-	-
1561	Met	8.68	120.59	-	-
1562	Gln	8.81	115.38	-	-
1563	Cys	8.22	122.33	-	-
1564	Ala	8.86	123.44	-	-
1566	Cys	7.63	121.81	-	-
1568	Leu	8.61	121.58	-	-
1569	Lys	8.61	123.14	-	-
1570	Asp	7.59	115.34	-	-
1571	Ser	7.86	114.66	-	-
1572	Asp	9.23	124.67	-	-
1573	Asp	8.39	118.38	-	-
1574	Tyr	7.96	122.61	-	-
1575	Ala	7.7	121.66	-	-
1576	Gln	7.91	114.66	-	-
1577	Leu	7.72	120.68	-	-
1578	Cys	7.97	110.97	-	-
1579	Asn	7.76	119.4	-	-
1582	Val	8.29	120.32	-	-
1583	Thr	8.07	115.85	-	-
1584	Gly	8.34	110.93	-	-
1589	Tyr	8.13	119.78	-	-
1590	Gly	8.27	110.62	-	-
1591	Arg	8.16	120.51	-	-
1592	Asp	8.36	120.49	-	-
1593	Ala	8.03	123.57	-	-
1594	Leu	8.12	120.59	-	-
1595	Val	7.86	119.94	-	-
1596	Asp	8.2	123.3	-	-
1597	Phe	8.18	121.66	-	-
1598	Ser	8.23	117.28	-	-
1599	Glu	8.28	122.37	-	-
1600	Gln	8.16	120.29	-	-
1601	Tyr	8.19	122.22	-	-
1602	Thr	7.97	120.19	-	-
1604	Glu	8.41	119.93	-	-
1605	Ala	8.07	124.3	-	-
1606	Asp	8.14	121.34	-	-
1608	Tyr	8.25	118.88	-	-
1609	Phe	7.73	121.03	-	-
1610	Ile	7.84	123.15	-	-
1611	Gln	8.21	123.88	-	-
1612	Asp	8.27	121.53	-	-
1613	Arg	8.06	120.74	-	-
1614	Phe	8.2	120.19	-	-
1615	Leu	8.03	122.68	-	-
1616	Asn	8.28	118.79	-	-
1617	Ser	8.1	115.59	-	-
1621	Leu	8.24	123.41	-	-
1622	Gln	8.41	121.19	-	-
1623	Ala	8.33	125.17	-	-

1624	Glu	8.4	119.4	-	-
1628	Ile	7.36	120.26	-	-
1629	Leu	8.23	126.47	-	-
1630	Asn	8.39	117.48	-	-
1631	Gly	7.94	107.15	-	-
1632	Cys	8.5	115.33	-	-
1633	Glu	8.08	121.67	-	-
1634	Asn	8.58	117	-	-
1635	Gly	6.81	104.16	-	-
1636	Arg	7.9	117.53	-	-
1637	Cys	9.05	125.54	-	-
1638	Val	8.89	125.24	-	-
1639	Arg	8.64	125.02	-	-
1640	Val	8.18	121.55	-	-
1641	Gln	8.69	122.48	-	-
1642	Glu	8.52	116.88	-	-
1643	Gly	7.59	109.15	-	-
1644	Tyr	8.37	120.33	-	-
1645	Thr	9.1	115.78	-	-
1646	Cys	7.68	117.54	-	-
1647	Asp	9.16	125.71	-	-
1648	Cys	8.38	123.45	-	-
1649	Phe	8.49	122.43	-	-
1650	Asp	8.54	119.58	-	-
1651	Gly	8.66	112.04	-	-
1652	Tyr	8.23	119.19	-	-
1653	His	9.75	116.63	-	-
1654	Leu	8.79	125.33	-	-
1655	Asp	9.11	128.62	-	-
1656	Thr	8.58	117.45	-	-
1660	Thr	7.4	109.97	-	-
1661	Cys	7.99	122.94	-	-
1662	Val	9.32	125.73	-	-

Table A1.6 - FBN^{e2cb1} assignment completeness:

Atom	% Assigned
H total	88%
H ^N	93.9%
H ^α	93.5%
H ^β	92.0%
N total	69%
N ^H	87%
C total	81%
C ^α	94.9%
C ^β	93.5%
C ^{CO}	92.7%

Appendix 2 - Representing protein evolution

Suggested here is a methodology for condensing large Phylogenetic trees of complex protein families into a species-focused graphical representation of protein homology, so as to more easily follow the emergence, loss, and diversification of specific protein isoforms during evolution.

Genome sequencing projects are producing a wealth of valuable information about the evolution of life and its constituent biochemical systems. Careful examination of this data can often give useful clues about the functional mechanisms used by these systems, for example by looking at sequence conservation. However with such a large quantity of data available it is often difficult and time consuming for researchers to sift out the data relevant to their own work.

Comparisons of protein sequence homology frequently rely on Phylogenetic trees to represent degrees of variation between different proteins, from both different species and of protein isoforms from the same species. While this is a very clear and informative way to represent this data, when looking at large protein families across multiple species these trees can quickly become complex and difficult to follow.

Proposed here is a simple methodology for taking the information from a phylogenetic tree, constructed by any method, and condensing it into a species defined tabular format. How specific features of a phylogenetic tree can be condensed into this format are summarised in Figure A2.1. To focus attention on species-species variation the species from which protein sequences have been gathered define each column, these columns are then populated by lines representing the proteins present in that species. These lines are positioned vertically according to the similarity of the protein in the

adjacent columns, allowing the viewer to easily follow the emergence, loss, and diversification of specific protein isoforms. Species columns can be arranged by their own predicted phylogenetic relationships, so that the most likely evolutionary history of protein isoforms can be more easily followed.

To demonstrate the use of this methodology an example phylogenetic tree was constructed from TGF-beta antagonists using proteins listed in the Interpro database as containing DAN (IPR004133) or sclerostin (IPR008835) domains. Figure A2.2 shows a full and condensed phylogenetic tree produced from these proteins and Figure A2.3 is a representation generated by the methodology described here. (For simplicity the example tree here only contains proteins from interpro searches, and is not a comprehensive analysis of DAN/Sclerostin protein evolution)

The disadvantage of representing protein relationships in this way is that information is lost relating to the “quantitative” homology between proteins, and statistical validation of the phylogenetic tree. However a trade off must sometimes be made in the process summarising complex data into a more accessible format.

These figures are also only as good as the data used to construct them, so careful attention must be paid to the construction of the alignments and phylogenetic tree from which they are derived, and rigorous blast searches must be carried out in all included organisms if a comprehensive view of protein evolution is to be achieved.

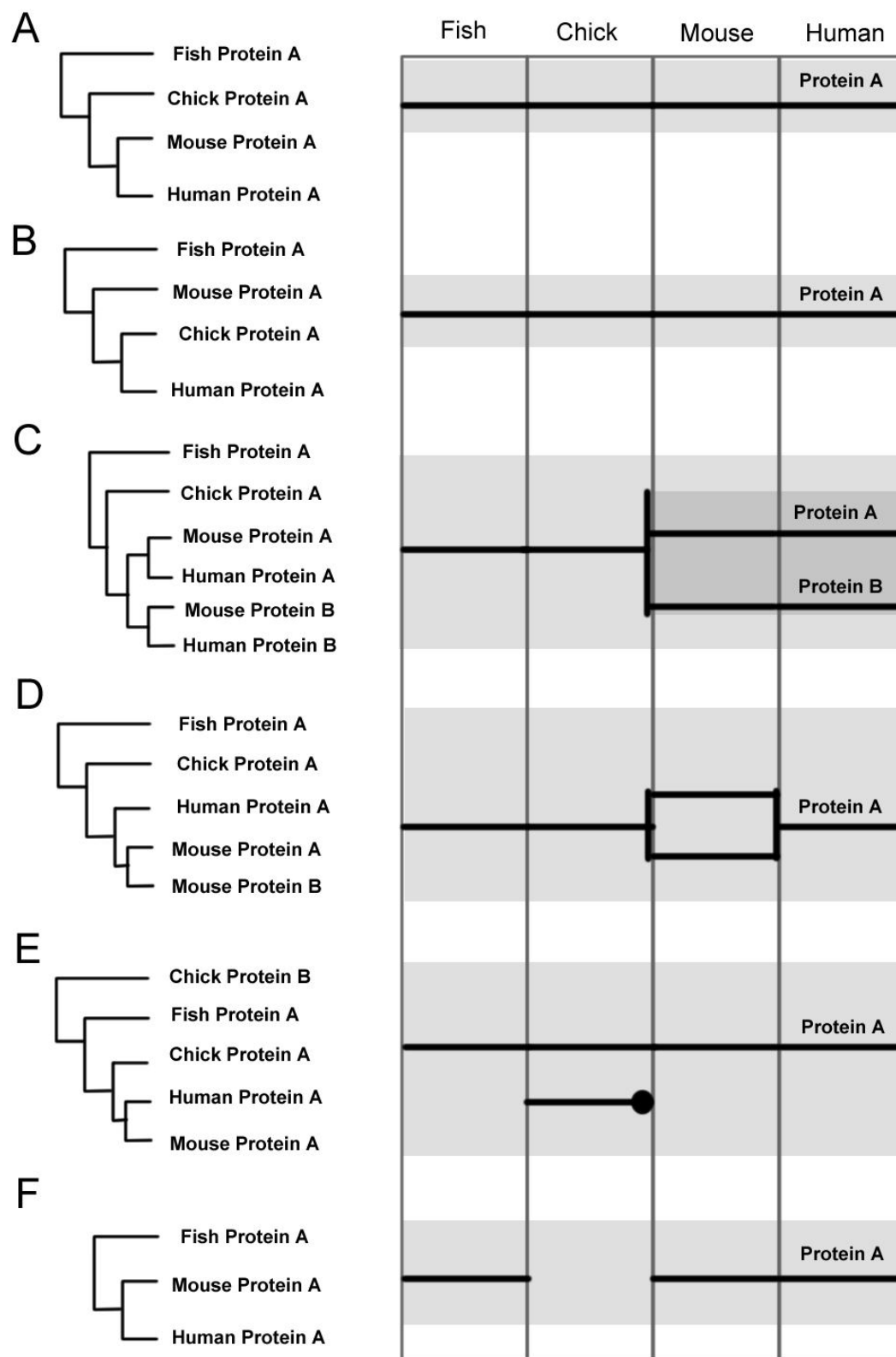
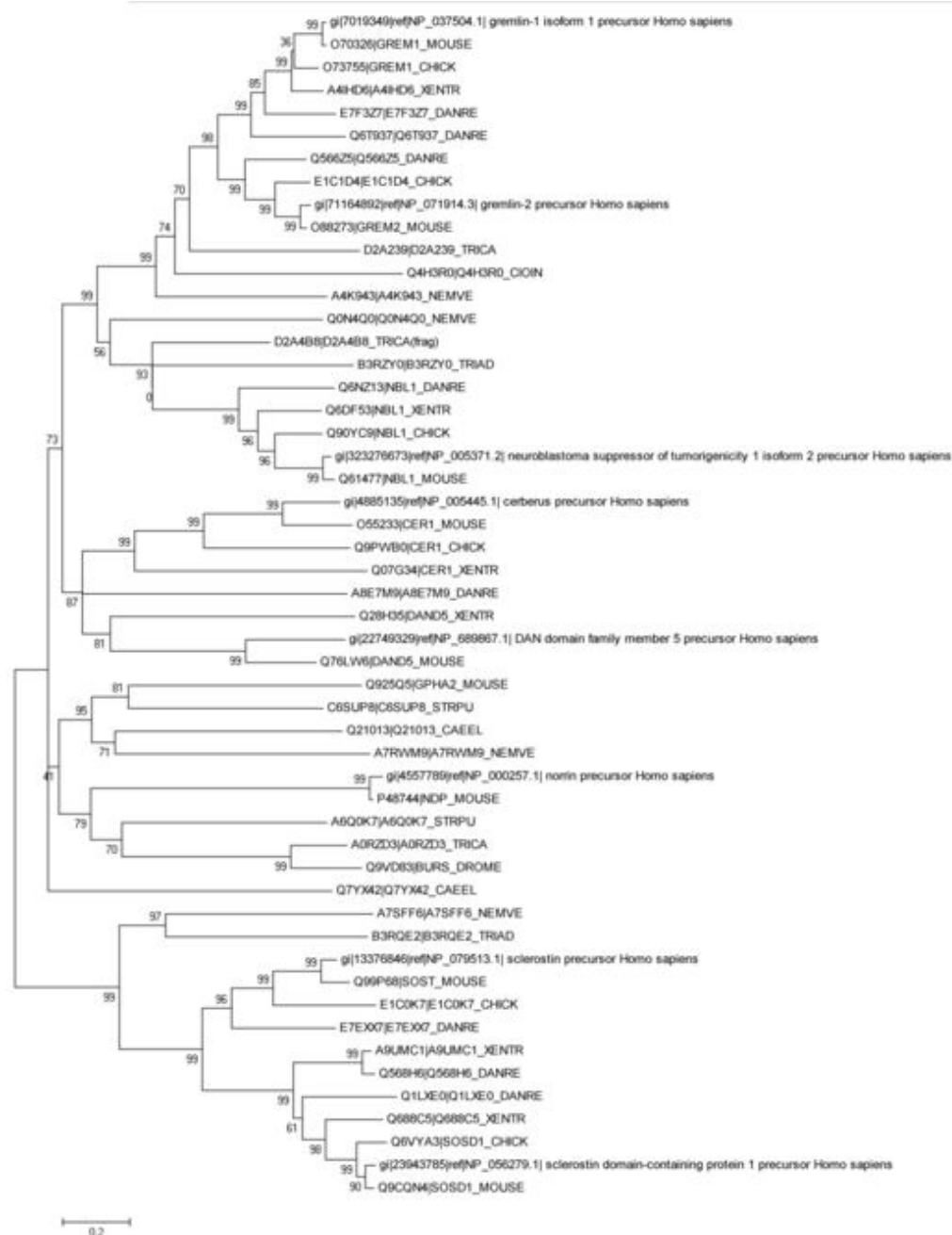


Figure A2.1- Examples of how to condense phylogenetic tree features into a tabular format.

A



B

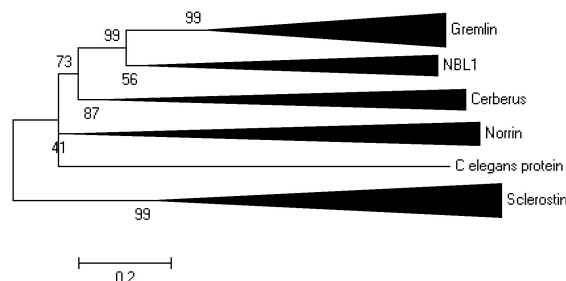


Figure A2.2 A)- Full phylogenetic tree generated using the Mega 5 software package, from sequences aligned by clustalW. (CHICK-Gallus gallus, XENTR-Xenopus Tropicalis, DANRE-Danio Rerio, TRICA-Tribolium castenum, CION-Ciona intestinalis, NEMVE-Nematostella Vectensis, TRIAD- Trichoplax Adhaerens, DROME- Drosophila Melanogaster, STRPU- Strongylocentrotus purpuratus, CAEEL- Caenorhabditis elegans. Figure 2 B) Tree condensed by manually chosen sub-tree groups.

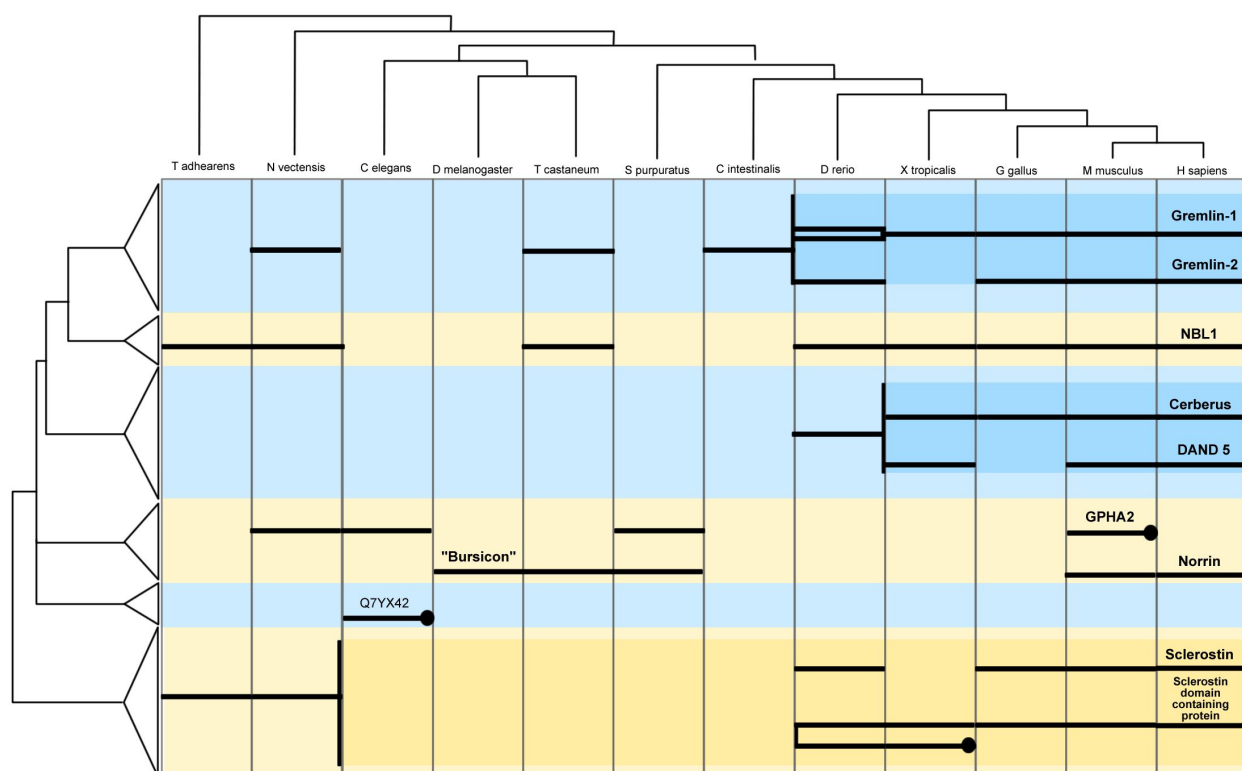


Figure A2.3- Protein evolution table for Sclerostin and DAN domain proteins, manually generated from the data in figure 2 using the methodology proposed here. Above the table the tree structure represents the evolutionary relationships between various species that define the table columns (according the tree of life web project). To the left of the table a tree equivalent to that in figure 2B summaries the overall homology between the groups of proteins in the table. In the table itself groups of more closely homologous proteins are distinguished by use of a different colour background, increasing in intensity to distinguish subgroups.

Appendix 3 – Extensive tables of relevant mouse models

Table A3.1- fibrillin mouse models

Mutation / disruption	Heterozygous mouse phenotype	Homozygous mouse phenotype	Notes	Refs
mgΔ FBN1 - del exon 19-24 (cbEGF8-TB3)	-Morphologically and histologically similar to WT littermates -fertile -Normal lifespan -defective alveolar septation giving progressive increase in alveolar size after birth compared to WT	-No gross abnormalities at birth -die within 3 weeks of birth from cardiovascular complications -aortic dilatation/dissection -fragmentation of elastic fibres in aorta -No skeletal changes -alveolar distension and coalescence -defective alveolar septation, (more pronounced than in heterozygotes)	-10 fold reduction in mutant fbn1 mRNA (due to neomycin cassette), fbn2 mRNA unaffected. -Shortened protein can be detected in fibroblast medium. -large reduction of fbn1 in -/- tissues, but normal elastin staining seen in neonatal skin -Normal Microfibrils can still be seen with EM -increased active TGFβ seen in +/- and -/- alveoli [1], TGFβ neutralising antibody restores normal alveolar septation.	[2] [3]
mgR FBN1 – Hypomorph- 5 fold lower expression of WT FBN1	-Normal	-Average survival 3.8 months -Wk 0 normal vascular anatomy-wk 6 calcification -Wk 8 monocytic infiltration, fragmentation of elastic lamellae -Wk 9 thickening of tunica intima ->Aortic aneurysm -Severe Kyphosis and overgrowth of ribs	-Macrophages seen clustering at sites of Elastin fragmentation and large numbers found in advanced aneurysms (may secrete MMP-12 which degrades fbn1) (-compound mgR/ mgΔ heterozygotes die prenatally) -suggests aneurysm occurs when fbn levels fall below a specific threshold. -MMP2 -/- mgR/mgR mice have less aortic dilatation, lower pErk1/2 + pSMAD, + less active TGFβ and more LTBP1 [4]	[5]
C1039G FBN1 – C1039G miss sense - (C1039Y mutation associated with classical MFS in humans [6, 7])	-Display normal life span and do not suffer aortic dissection -Aortae normal until ~2months then progressive elastic fibre fragmentation and deposition of amorphous matrix (MMP-2+9 inc) -Occasional calcification, but inflammation and hyperplasia not seen (unlike mgR +Δ) -Gradual PN development of kyphosis and	-Die soon after birth due to failure of the aortic wall (but do not display antecedent calcification or inflammation [8])	-Both +/- and -/- mice show reduced fibrillin deposition in culture but normal mRNA levels -Crosses of +/- mice with transgenic mice expressing WT human fibrillin rescues the phenotype -Cross of +/- with transgenic C1663R MFS FBN also partially rescues the phenotype -Aberrant TGFβ signalling observed in mitral valves [9], muscle [10], and Aorta [11]. TGFβ antagonism shown to restore normal development in these tissues	[12]

	elongated ribs -distal airspace widening			
FBN1 - transgenic over expression of human fbn1 C1663R protein from YAC	-No mouse phenotype, and both wt and C1663R FBN1 detected forming fibrillar structures in mouse fibroblast cultures with human specific mAb (contrasts with previous transfected cell culture results [13]) -Immunoelectron microscopy shows incorporation of both human WT and mutant fbn1 in microfibril like structures			[12]
mgN FBN1 -null allele	-viable and fertile -"subclinical" abnormalities in aortic wall	-All die within 2 weeks of birth -aortic aneurysm, -impaired pulmonary function, -enlarged alveoli -diaphragmatic collapse -fragile skin + organs -neonates show poorly developed medial layer, after death elastin fragmented + "wavy" but no excessive matrix accumulation or hyperplasia -significant increase of MMP9 protein in aorta -endothelial lining often detached and abnormal lamellae not just in lesions but throughout aorta (unlike other models)	-Stronger VSMC staining in -/- foetal aortas> fbn1 loss interferes with aortic maturation? -3x more phosphorylated SMAD2 in aorta in westerns, +inc activators of fibrosis, but no fibrosis. -Normal levels of elastin crosslinking -Almost all fbn1 -/- and fbn2 -/-, and half fbn1 +/- fbn2 -/- mice die in utero (after E14.5) with more severe vascular problems, and show v little elastin deposition in vessels. -RS EM showed structurally different microfibrils in mgN/mgN mice (asymmetric beads compared to symmetric beads in WT (wider bead spacing?))	[14]
TSK FBN1 -in frame duplication (cbEGF7-cbEGF24 (includes RGD in TB4)	-Thick inelastic skin, firmly bound to subcutaneous tissues, with excessive collagen and GAG accumulation -enlarged heart due to myocardial fibrosis -Lung fibrosis and elastolysis -enlarged but correctly proportioned skeleton -normal lifespan	In utero death around day 8	Accumulation of abnormal microfibrils observed (green et al 1976) SMAD3 deletion in Tsk mice decreased CCN3 overexpression suggesting TGF β involvement [15] IL-4 disruption rescues homozygotes from embryonic death [16] and TGF β 1 +/- TSK +/- mice do not develop skin fibrosis [17] (result also seen with the intracellular adhesion molecule ICAM-1 [18]) -TSK/MgR mice display both TSK and MFS traits, and die within first 3 weeks from vascular complications [19] (TSK fibrillin can also co-polymerise with WT fbn1)	[20]

GT-8 FBN1- truncation after cbEGF17 -(C-term GFP tagged)	-Progressive fragmentation of elastic lamellae from 2 months of age -cultured fibroblasts only produce sparse fibrillin matrix, and also GFP labelled fibrils -some embryonic lethality of Gt8/+ fbn2-/- pups, the rest born dead or dying. (contrast with mgN where 50% of these pups were viable)	-Die primarily between P9 and P18 -Cultured fibroblasts unable to assemble a fbn matrix (cultures didn't express fbn2) -EM shows absence of "basket" like microfibril structures surrounding elastic fibres in skin	-Secreted GT fibrillin levels in medium similar to WT -tissue sections Show GFP incorporation in typical microfibrillar patterns -most fluorescence in aorta, becomes progressively more intense during PN life, not visible in other tissues for 1 month after birth -ImmunoEM shows direct incorporation of truncated fbn1 into microfibrils (in GT8/GT8), but periodicity in fbn1 staining lost, and progressive loss of fbn 1 staining and microfibril bundles observed -exposure of fbn2 epitopes (but mRNA levels not affected)	[21]
H1Δ FBN1- Hybrid 1 domain deletion (MFS causing exon skipping mutation may be equivalent)	Normal	Normal, with normal staining of fibrillin1 fibres in tissues and normal limited exposure of fbn2 epitopes -no visible difference in microfibril ultra-structure by EM -no significant elastic fibre fragmentation even after 1 yr		[21]

FBN1- WMS del TB1-EGF4	-Displays similar phenotype to homozygous mice, but not as severely	-Survives normally and does not display major MFS features (eg elastic lamellae normal) -Reduced long bone and digit bone growth after 3-4weeks, but growth normalised again after 5months -Develop skin fibrosis -Excessive collagen deposition, wider dermal layer reduced hypodermal fat - Hyper muscularity	-Alterations in microfibril ultrastructure in both +/- and -/- mice: reduced periodicity for immunogold labelling + large microfibril accumulations present in skin -Elastic fibres have “moth-eaten” appearance -Reduced ADAMTSL-6 immunostaining in skin -Thicker “bundles” of fibrillin fibres in mouse model cultures than human WMS cultures -Westerns for pSMAD2/3 show no changes compared to WT at multiple time points + no increase in TGFβ responsive <i>ctgf</i> , <i>pai1</i> or <i>postn</i> expression	[22]
FBN1- cre-lox conditional FBN1- null	-Ubiquitous knockdown by crossbreeding <i>fbn</i> ^{lox} /+ with <i>fbn</i> +/- CMV mice phenocopy <i>fbn</i> -/- mice as expected. - <i>fbn</i> KO in skeletal progenitor cells (<i>Prx1</i> -Cre) and <i>fbn</i> KO in early pre-osteoblasts (<i>Osx</i> -Cre) both show lower bone mass + worse trabecular architecture than <i>mgR/mgR</i> mice			[23]
FBN2- Null		-Normal aortic wall -Display “bilateral syndactyly” (digit fusion) -Contractures of wrist + metacarpal joints in new born mice- disappear within a few days -Fbn1 losses arrangement in digital and inter-digital arrays -no evidence of impaired lung branching	-Reduced apoptosis seen in inter digital tissue -BMP4 beads implanted in developing “autopod” of -/- mice don't increase local <i>Msx1/2</i> mRNA (unlike wt) - <i>Msx</i> , <i>BMP4</i> , and <i>fgf8</i> expression pattern in -/- normal - <i>fbn2</i> +/- and <i>BMP7</i> +/- displays both soft tissue syndactyly and polydactyly, (combined defects of both double null mice)	[24] [14]

Table A3.2- LTBP and other mouse models

Mutation / disruption	Heterozygous mouse phenotype	Homozygous mouse phenotype	Notes	refs
LTBP1- deletion of first shared exon		Viable and fertile -some mild craniofacial defects such as a shortened mandible -Slight shortening of long bones -Less fibrogenesis in liver- lower hydroxyproline content in liver + reduced pSMAD2	Despite exon being used in both isoforms it is possible that it is skipped and LTBP1L like protein is still produced in its absence	[25]
LTBP1L – deletion of long isoform only		~40% embryonic lethality -Rest die shortly after birth from cardiac defects: persistent truncus arteriosus (PTA) and interrupted aortic arch (IAA), -Hypoplasia of endocardial cushions in early embryonic valvulogenesis, and hyperplastic valves in late valvulogenesis.	-Defects are associated with the faulty function of cardiac neural crest cells (CNCCs) -Reduced pSMAD2 levels seen in specific tissues eg in valves, but not myocardium from E12.5 -Attenuated Epidermal Mesenchymal Transition seen <i>in vitro</i> - prolonged EMT thought to result in valve hyperplasia observed	[26] [27]
LTBP1 cb12-TB2-cb13 transgenic expression under control of keratin 14 promoter	- Staining showed protein in outer root sheet of follicle and basal epidermis -Short hair seen in transgenic mice (max 30% shorter at P15), but same hair morphology -Skin histology in tg mice showed shorter period of hair shaft formation (anagen) and early induction of hair shaft involution “catagen” (P15 instead of 18) -Some increase in pSMAD2/3 nuclei seen in tg skin, depending on age -None of the above seen with control tg mouse with TB2 domain modified so as not to bind TGFβ			[28]
LTBP2 null		-Initial mouse model suggesting death at implantation [29] has not been reproduced -Recently generated null mice may instead demonstrate ectopia lentis but not glaucoma [30]		
LTBP3 null		-Multiple skeletal defects: -Craniofacial malformations by P10 -Kyphosis (by 2 mo) -Over extension of mandible -ossification of skull synchondroses - Osteosclerosis and osteoarthritis - Osteopetrosis with high bone mass due to reduced bone resorption -Growth retardation - Spleen and thymus involution and death of ~1/2 mice between weeks 3 and 4 when on a 129SvEV inbred background -Alveolar septation defective [31]	-reduced levels of LTBP3 may impair TGFβ signalling in bone and cartilage (?)- this may compromise osteoclast function and reduce bone turnover. -Thymus and spleen defect caused by elevated corticosterone levels, restored by injection of AMG, an inhibitor of steroid synthesis -Increased levels of surfactant B producing cells in lungs -Normal appearance of elastic fibres in alveoli, also normal number of myofibroblasts -Reduced nuclear pSMAD2/3 in alveoli between P4 and P6, but not at other time points -Increased lung cell proliferation and apoptosis seen days 4-10	[32] [33] [31] [34] [35]
LTBP4 disruption with “gene trap” intronic		-Cardiomyopathy develops after week 12, with ventricular dilatation -Colorectal cancer (lesions similar	-Major reduction in LTBP4 levels demonstrated by western -Significant Reduction in extracellular TGFβ1 deposition	[38] [39] [40]

integration- “nearly null”		to SMAD3 KO mouse [36]) -Rectal prolapse by week 4 -Pulmonary emphysema present at birth -Lung and colon elastic fibres highly fragmented but normal in heart -Defective alveolar septation, evident from E18.5 -Aberrant elastin deposition evident from E14.5 (before fibres even visible in WT mice) -EM shows abnormal elastic fibre ultrastructure, with dense globules of elastin sitting adjacent to microfibril bundles instead of intermeshed with microfibrils as in WT	in colon, lung, heart, but not in kidney. -pSMAD levels also decreased in lung and colon but not kidney>increased c-myc induction> cellular proliferation -However increase in nuclear pSMAD2 seen in alveoli of 7 day old mice -Alveolar septation defect reversed by lowering TGFβ2 levels (with TGFβ2 +/- cross) but elastic fibres still fragmented -Only very small increase in TGFβ expression at P7 -Emphysema partially rescued by sestrin 2 (antioxidant protein) inactivation [37]	
LTBP3xLTBP4 Combined null		-Early lethality compared to single null mice -Don't display any novel defects not seen in individual nulls -Lung abnormalities exacerbated and increased septation deficiency distinguishable by P7 -	-Significant increase in apoptosis seen at P0.5 (not seen in single nulls) -proliferation the same -Abnormal distribution of type 2 cells and myofibroblasts in lung -Decreased incorporation of fnb 1 and 2 in alveolar matrix	[41]
TGFβ1	-Normal and gain weight at same rate as WT.	-Die after 3-4 weeks due to multi-organ inflammation, with massive infiltration of lymphocytes and macrophages, especially the heart and lungs. -only represent 10% of offspring instead of 25%> high embryonic lethality -Mice exhibit “dishevelled” appearance, and wasting, but no gastro-intestinal complications -No significant inflammatory lesions seen in small intestine, thymus, most connective tissues, brain, skin, or bone marrow. -Lack langerhans cells in epidermis -Defective haematopoiesis and vasculogenesis	-Do not develop gastrointestinal tumours (unless exposed to pathogens or combined with Rag2 -/- mutation [42, 43], where benign tumours develop) -Elimination of CD4+ and CD8+ T cells prevents inflammation and prolongs survival [44]	[45] [46] [47]
TGFβ1 RGD>RGE		-Phenocopy TGFβ1 null mice including, multi organ inflammation, vasculogenesis defects, and lack of langerhans cells	-TGFβ rge/rge TGFβ3 -/- mice also display brain vascular defects similar to those in Itgβ8 -/- mice [48]	[49]
TGFβ1 Cys33>Ser (Cannot covalently bind LTBP)		-Multi organ inflammation, and shortened life span, but not as severe as TGFβ1 null mice -Lack of skin Langerhans cells -Gastrointestinal tumours	-TGFβ signalling appears significantly reduced, less active TGFβ in sera, with reduced pSMAD, and increased c-Myc and Mitosis (Ki-67) in stomach tissue	[50]
TGFβ2	-No	- Multiple developmental defects		[51]

	apparent embryonic lethality	that do not overlap with other TGF β null mice - 2/3rds die shortly before or during birth -Congenital heart defects- small and thin walled aorta, ventricular septation defects, hypercellular myocardium. -Craniofacial defects- general reduction in bone size and less ossification. Shortened mandible (Retrognathia), lacking various processes. Complete cleft palate, with defective formation of palatal shelves (clear at E18.5) -Other skeletal defects- shortened forelimbs, spina bifida, abnormal ribs and sternum. -eye defects with hypercellular infusion of posterior chamber and thinner corneal stroma -Inner ear defects -Urogenital defects- testicular ectopia and hypoplasia. Degeneration of kidney tubular epithelium -Lungs exhibit no major defects, except collapsed airways in PN animals -Brain, gut and tooth bud normal -(Thyroid glands normal)		
TGF β 3		-No prenatal lethality -Offspring did not suckle and died within 20h of birth -Similar birth weight to WT animals but deteriorated rapidly and showed respiratory distress -Severely disrupted lung development including defective alveolar septation, mesenchymal thickening an hypercellularity, fewer SPC cells, and pleural haemorrhage -Cleft palate, although at E14.5 shelves had grown out and were adherent, but did not fuse. -No skeletal or cardiovascular abnormalities, or changes in tooth bud or whisker follicle development, despite high TGF β 3 expression in these tissues.	TGF β 2-/- TGF β 3 -/- double nulls display early embryonic death, with failure of midline fusion [52, 53] TGF β 2-/- TGF β 3 -/+ also still display severe malformations while TGF β 2-/+ TGF β 3 -/- mice do not TGF β 1 knocked into the TGF β 3 locus gives partial but not complete rescue of palate closure [54], suggesting specific activities attributed to TGF β 3 protein	[55] [56]
Transgenic expression of C33S TGF β 1 in heart	-Similar levels of active TGF β observed in both atria and ventricles but fibrosis only seen in atria -Delayed wound healing after myocardial injury.			[57]
Integrin β 6 -/-		-Juvenile baldness due to macrophage skin infiltration, and increased inflammation of other tissues such as lungs		[58] [59]
Integrin β 8 -/-		-Embryonic or perinatal lethal with	Itgb8-/- which are also treated	[60]

		severe vascular defects -insufficient vascularisation of placenta and yolk sac -those that survive to birth suffer major intracerebral haemorrhage -Backcrossing onto ICR background improved survival to 2 months [59], showed general weakness	with an anti Itgb6 antibody develop severe auto immunity and lack langerhans cells, (as with TGFβ1 null mice)	[59]
Integrin β8 -/- + β6 -/-		-Most die at mid gestation - ~2/3 surviving mice develop cleft palate. -histology otherwise unremarkable except for slight delay in alveolar development.		[59]
Integrin β8 specific Cre- Lox KO		Leukocyte or dendritic cell specific KO caused wasting and inflammation of liver and colon		[61]
SMAD3 null		Develop metastatic colon cancer		[36]
fibronectin	Normal	Die early in embryo-genesis		[62] [63]

1. Neptune, E.R., et al., *Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome*. Nat Genet, 2003. **33**(3): p. 407-11.
2. Pereira, L., et al., *Targetting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome*. Nat Genet, 1997. **17**(2): p. 218-22.
3. Lima, B.L., et al., *A new mouse model for marfan syndrome presents phenotypic variability associated with the genetic background and overall levels of Fbn1 expression*. PLoS One. **5**(11): p. e14136.
4. Xiong, W., et al., *MMP-2 Regulates Erk1/2 Phosphorylation and Aortic Dilatation in Marfan Syndrome*. Circ Res, 2012.
5. Pereira, L., et al., *Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1*. Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3819-23.
6. Schrijver, I., et al., *Cysteine substitutions in epidermal growth factor-like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes*. Am J Hum Genet, 1999. **65**(4): p. 1007-20.
7. Liu, W.O., et al., *Denaturing HPLC-identified novel FBN1 mutations, polymorphisms, and sequence variants in Marfan syndrome and related connective tissue disorders*. Genet Test, 1997. **1**(4): p. 237-42.
8. Ramirez, F., et al., *Fibrillin microfibrils: multipurpose extracellular networks in organismal physiology*. Physiol Genomics, 2004. **19**(2): p. 151-4.
9. Ng, C.M., et al., *TGF-beta-dependent pathogenesis of mitral valve prolapse in a mouse model of Marfan syndrome*. J Clin Invest, 2004. **114**(11): p. 1586-92.
10. Cohn, R.D., et al., *Angiotensin II type 1 receptor blockade attenuates TGF-beta-induced failure of muscle regeneration in multiple myopathic states*. Nat Med, 2007. **13**(2): p. 204-10.
11. Habashi, J.P., et al., *Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of Marfan syndrome*. Science, 2006. **312**(5770): p. 117-21.
12. Judge, D.P., et al., *Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome*. J Clin Invest, 2004. **114**(2): p. 172-81.
13. Eldadah, Z.A., et al., *Expression of a mutant human fibrillin allele upon a normal human or murine genetic background recapitulates a Marfan cellular phenotype*. J Clin Invest, 1995. **95**(2): p. 874-80.
14. Carta, L., et al., *Fibrillins 1 and 2 perform partially overlapping functions during aortic development*. J Biol Chem, 2006. **281**(12): p. 8016-23.
15. Lemaire, R., et al., *Antagonistic effect of the matricellular signaling protein CCN3 on TGF-beta- and Wnt-mediated fibrillinogenesis in systemic sclerosis and Marfan syndrome*. J Invest Dermatol, 2010. **130**(6): p. 1514-23.

16. Kodera, T., et al., *Disrupting the IL-4 gene rescues mice homozygous for the tight-skin mutation from embryonic death and diminishes TGF-beta production by fibroblasts*. Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3800-5.
17. McGaha, T., et al., *Lack of skin fibrosis in tight skin (TSK) mice with targeted mutation in the interleukin-4R alpha and transforming growth factor-beta genes*. J Invest Dermatol, 2001. **116**(1): p. 136-43.
18. Matsushita, Y., et al., *Intercellular adhesion molecule-1 deficiency attenuates the development of skin fibrosis in tight-skin mice*. J Immunol, 2007. **179**(1): p. 698-707.
19. Gayraud, B., et al., *New insights into the assembly of extracellular microfibrils from the analysis of the fibrillin 1 mutation in the tight skin mouse*. J Cell Biol, 2000. **150**(3): p. 667-80.
20. Siracusa, L.D., et al., *A tandem duplication within the fibrillin 1 gene is associated with the mouse tight skin mutation*. Genome Res, 1996. **6**(4): p. 300-13.
21. Charbonneau, N.L., et al., *In vivo studies of mutant fibrillin-1 microfibrils*. J Biol Chem.
22. Sengle, G., et al., *Microenvironmental regulation by fibrillin-1*. PLoS Genet, 2012. **8**(1): p. e1002425.
23. Cook, J.R., et al., *Generation of Fbn1 conditional null mice implicates the extracellular microfibrils in osteoprogenitor recruitment*. Genesis, 2012.
24. Arteaga-Solis, E., et al., *Regulation of limb patterning by extracellular microfibrils*. J Cell Biol, 2001. **154**(2): p. 275-81.
25. Drews, F., et al., *Disruption of the latent transforming growth factor-beta binding protein-1 gene causes alteration in facial structure and influences TGF-beta bioavailability*. Biochim Biophys Acta, 2008. **1783**(1): p. 34-48.
26. Todorovic, V., et al., *Long form of latent TGF-beta binding protein 1 (Ltbp1L) is essential for cardiac outflow tract septation and remodeling*. Development, 2007. **134**(20): p. 3723-32.
27. Todorovic, V., et al., *Long form of latent TGF-beta binding protein 1 (Ltbp1L) regulates cardiac valve development*. Dev Dyn, 2011. **240**(1): p. 176-87.
28. Mazzieri, R., et al., *Expression of truncated latent TGF-beta-binding protein modulates TGF-beta signaling*. J Cell Sci, 2005. **118**(Pt 10): p. 2177-87.
29. Shipley, J.M., et al., *Developmental expression of latent transforming growth factor beta binding protein 2 and its requirement early in mouse development*. Mol Cell Biol, 2000. **20**(13): p. 4879-87.
30. Todorovic, V. and D.B. Rifkin, *LTBPs, more than just an escort service*. J Cell Biochem, 2012. **113**(2): p. 410-8.
31. Colarossi, C., et al., *Lung alveolar septation defects in Ltbp-3-null mice*. Am J Pathol, 2005. **167**(2): p. 419-28.
32. Chen, Y., et al., *Growth retardation as well as spleen and thymus involution in latent TGF-beta binding protein (Ltbp)-3 null mice*. J Cell Physiol, 2003. **196**(2): p. 319-25.
33. Dabovic, B., et al., *Bone abnormalities in latent TGF-[beta] binding protein (Ltbp)-3-null mice indicate a role for Ltbp-3 in modulating TGF-[beta] bioavailability*. J Cell Biol, 2002. **156**(2): p. 227-32.
34. Dabovic, B., et al., *Bone defects in latent TGF-beta binding protein (Ltbp)-3 null mice; a role for Ltbp in TGF-beta presentation*. J Endocrinol, 2002. **175**(1): p. 129-41.
35. Dabovic, B., et al., *Osteopetrosis-like phenotype in latent TGF-beta binding protein 3 deficient mice*. Bone, 2005. **37**(1): p. 25-31.
36. Zhu, Y., et al., *Smad3 mutant mice develop metastatic colorectal cancer*. Cell, 1998. **94**(6): p. 703-14.
37. Wempe, F., et al., *Inactivation of sestrin 2 induces TGF-beta signaling and partially rescues pulmonary emphysema in a mouse model of COPD*. Dis Model Mech, 2010. **3**(3-4): p. 246-53.
38. Sterner-Kock, A., et al., *Disruption of the gene encoding the latent transforming growth factor-beta binding protein 4 (LTBP-4) causes abnormal lung development, cardiomyopathy, and colorectal cancer*. Genes Dev, 2002. **16**(17): p. 2264-73.
39. Dabovic, B., et al., *Dual functions for LTBP in lung development: LTBP-4 independently modulates elastogenesis and TGF-beta activity*. J Cell Physiol, 2009. **219**(1): p. 14-22.
40. Koli, K., et al., *Disruption of LTBP-4 function reduces TGF-beta activation and enhances BMP-4 signaling in the lung*. J Cell Biol, 2004. **167**(1): p. 123-33.

41. Dabovic, B., et al., *Control of lung development by latent TGF-beta binding proteins*. J Cell Physiol, 2011. **226**(6): p. 1499-509.
42. Engle, S.J., et al., *Elimination of colon cancer in germ-free transforming growth factor beta 1-deficient mice*. Cancer Res, 2002. **62**(22): p. 6362-6.
43. Engle, S.J., et al., *Transforming growth factor beta1 suppresses nonmetastatic colon cancer at an early stage of tumorigenesis*. Cancer Res, 1999. **59**(14): p. 3379-86.
44. Bommireddy, R., et al., *Elimination of both CD4+ and CD8+ T cells but not B cells eliminates inflammation and prolongs the survival of TGFbeta1-deficient mice*. Cell Immunol, 2004. **232**(1-2): p. 96-104.
45. Kulkarni, A.B., et al., *Transforming growth factor beta 1 null mutation in mice causes excessive inflammatory response and early death*. Proc Natl Acad Sci U S A, 1993. **90**(2): p. 770-4.
46. Dickson, M.C., et al., *Defective haematopoiesis and vasculogenesis in transforming growth factor-beta 1 knock out mice*. Development, 1995. **121**(6): p. 1845-54.
47. Caux, C., et al., *Respective involvement of TGF-beta and IL-4 in the development of Langerhans cells and non-Langerhans dendritic cells from CD34+ progenitors*. J Leukoc Biol, 1999. **66**(5): p. 781-91.
48. Mu, Z., et al., *TGFbeta1 and TGFbeta3 are partially redundant effectors in brain vascular morphogenesis*. Mech Dev, 2008. **125**(5-6): p. 508-16.
49. Yang, Z., et al., *Absence of integrin-mediated TGFbeta1 activation in vivo recapitulates the phenotype of TGFbeta1-null mice*. J Cell Biol, 2007. **176**(6): p. 787-93.
50. Yoshinaga, K., et al., *Perturbation of transforming growth factor (TGF)-beta1 association with latent TGF-beta binding protein yields inflammation and tumors*. Proc Natl Acad Sci U S A, 2008. **105**(48): p. 18758-63.
51. Sanford, L.P., et al., *TGFbeta2 knockout mice have multiple developmental defects that are non-overlapping with other TGFbeta knockout phenotypes*. Development, 1997. **124**(13): p. 2659-70.
52. Dunker, N. and K. Krieglstein, *Tgfbeta2 -/- Tgfbeta3 -/- double knockout mice display severe midline fusion defects and early embryonic lethality*. Anat Embryol (Berl), 2002. **206**(1-2): p. 73-83.
53. Dunker, N. and K. Krieglstein, *Reduced programmed cell death in the retina and defects in lens and cornea of Tgfbeta2(-/-) Tgfbeta3(-/-) double-deficient mice*. Cell Tissue Res, 2003. **313**(1): p. 1-10.
54. Yang, L.T. and V. Kaartinen, *Tgfb1 expressed in the Tgfb3 locus partially rescues the cleft palate phenotype of Tgfb3 null mutants*. Dev Biol, 2007. **312**(1): p. 384-95.
55. Kaartinen, V., et al., *Abnormal lung development and cleft palate in mice lacking TGF-beta 3 indicates defects of epithelial-mesenchymal interaction*. Nat Genet, 1995. **11**(4): p. 415-21.
56. Taya, Y., S. O'Kane, and M.W. Ferguson, *Pathogenesis of cleft palate in TGF-beta3 knockout mice*. Development, 1999. **126**(17): p. 3869-79.
57. Nakajima, H., et al., *Atrial but not ventricular fibrosis in mice expressing a mutant transforming growth factor-beta(1) transgene in the heart*. Circ Res, 2000. **86**(5): p. 571-9.
58. Huang, X.Z., et al., *Inactivation of the integrin beta 6 subunit gene reveals a role of epithelial integrins in regulating inflammation in the lung and skin*. J Cell Biol, 1996. **133**(4): p. 921-8.
59. Aluwihare, P., et al., *Mice that lack activity of alphavbeta6- and alphavbeta8-integrins reproduce the abnormalities of Tgfb1- and Tgfb3-null mice*. J Cell Sci, 2009. **122**(Pt 2): p. 227-32.
60. Zhu, J., et al., *beta8 integrins are required for vascular morphogenesis in mouse embryos*. Development, 2002. **129**(12): p. 2891-903.
61. Travis, M.A., et al., *Loss of integrin alpha(v)beta8 on dendritic cells causes autoimmunity and colitis in mice*. Nature, 2007. **449**(7160): p. 361-5.
62. George, E.L., et al., *Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin*. Development, 1993. **119**(4): p. 1079-91.
63. Watt, F.M. and K.J. Hodivala, *Cell adhesion. Fibronectin and integrin knockouts come unstuck*. Curr Biol, 1994. **4**(3): p. 270-2.

Appendix 4 – Sequences used in evolutionary analyses

Table A4.1 – Sequences used in evolutionary analysis of fibrillins and LTBP3

Organism	Database used	Reference of protein sequence used (order as in figure 4)
<i>H. sapiens</i>	NCBI	P35555.2 FBN1 NP_001990.2 FBN2 NP_115823.3 FBN3 NP_996826.1 LTBP1 NP_000419.1 LTBP2 NP_001123616.1 LTBP3 NP_001036009.1 LTBP4
<i>B. Taurus</i>	NCBI	NP_776478.1 fibrillin-1 DAA27725.1 fibrillin 2-like XP_001254850.2 fibrillin-3 DAA13624-1 LTBP3
<i>E. caballus</i>	Ensembl	ENSECAP00000018025 <i>LTBP-1</i>
<i>M. Musculus</i>	NCBI	NP_032019 NP_034311 NP_064303 NP_038617 NP_032546 NP_001107021
<i>Canis familiaris LTBP-1</i>	Ensembl	ENSCAFP00000008768
<i>Wild boar</i>		ENSSSCP00000002566 LTBP2
<i>Dolphin</i>		ENSSSCP00000002566 LTBP1 ENSTTRP00000012372 LTBP2 ENSTTRP00000002108 LTBP3
<i>Monodelphis domestica</i>	Ensembl	ENSMODP00000019266 <i>LTBP-1</i>
<i>O. anatinus</i>	Ensembl	ENSOANP00000020466 FBN1 ENSOANP00000010191 FBN3 ENSOANP00000003699 LTBP1
<i>G. gallus</i>	NCBI	XP_413815.2 XP_424715.2 XP_418173.2 XP_419510 XP_426444
<i>T. guttata</i>	NCBI + Ensembl	XP_002196620.1 fibrillin 1 XP_002189202.1 fibrillin 2 ENSTGUP00000000998/ XP_002196137.1 fibrillin3 XP_002191623.1 LTBP1 ENSTGUP00000012510 LTBP2
<i>Turkey</i>		ENSMGAP00000011575 LTBP1 ENSMGAP00000013379 LTBP2
<i>A. carolinensis</i>	Ensembl	ENSACAP00000005193 FBN1 ENSACAP00000002293 FBN2 ENSACAP00000014508 FBN3 ENSACAP00000009073 <i>LTBP1</i> ENSACAP00000016291 LTBP2 ENSACAP00000002537 LTBP3
<i>X. tropicalis</i>	JGI	C_scaffold_221000019 C_scaffold_7000008 Ext_fgenesh1_pg.C_5550008 C_scaffold_98000018. gw1.409.46.1. gw1.146.1.1
<i>Xenopus laevis</i>	NCBI	AAN76497.1 <i>LTBP-1</i>
<i>L. chalumnae</i>	Ensembl	ENSLACP00000010136 FBN1 ENSLACP00000002844 FBN2 ENSLACP00000018340 FBN3 ENSMGAP00000011575 LTBP1 ENSLACP00000021326 LTBP2

		ENSLACP00000019176 LTBP4 ENSLACP00000002664 LTBP3
<i>O. niloticus</i>	Ensembl	ENSONIP00000016782 FBN1 ENSONIP00000002223 FBN3 ENSONIP00000014122 LTBP1 ENSONIP00000015138 LTBP2 ENSONIP00000001442 LTBP4 ENSLACP00000002664 LTBP3
<i>O. latipes (medaka fish)</i>	Ensembl	ENSORLP000000025211 <i>LTBP-1</i>
<i>G. aculeatus</i>	Ensembl	ENSGACP000000002952 <i>LTBP-1</i>
<i>D. rerio</i>	NCBI	NP_001129262.1 ACI29314.1 + XP_686500.4 (overlapping sequences combined) XP_691109.4 WISE2 re-prediction based on XP_002665326.1 XP_001342567.2 XP_001921093.1 XP_691374.3 XP_690670.3 XP_690560.3
<i>T. nigroviridis</i>	UNIPROT NCBI	Q4SHN1 FBN1 CAG08034.1 FBN3 Q4RLT5 LTBP Q4SXB6 + Q4RDX0 LTBP4 Q4SFI1 LTBP3
<i>P. marinus</i>	Ensembl	ENSPMAP000000002691 FBN ENSPMAP00000010095 LTBP1 ENSPMAP000000005378 LTBP4
<i>C. intestinalis</i>	NCBI	XP_002119988.1
<i>B. floridae</i>	JGI	estExt_gwp.C_10660003 FBN scaffold_132000005
<i>S. kowalevskii</i>		XP_002741297.1 fibrillin NP_001164705.1 LTBP
<i>S. purpuratus</i>	NCBI	WISE2 re-prediction of XP_787529.2 XP_784696.2
<i>T. castaneum</i>	NCBI	XP_974344.2
<i>A. mellifera</i>	NCBI	XP_001122812.1
<i>A. pisum</i>	NCBI	XP_001945971.1
<i>P. humanus corporis</i>	NCBI	XP_002428778.1
<i>D. pulex</i>	JGI	240066 SNAP_00008124
<i>I. scapularis</i>	NCBI	WISE2 re-prediction based on XP_002412357.1
<i>L. gigantea</i>	JGI	C_sca_28000110 + e_gw1.28.29.1
<i>C. teleta</i>	JGI	estExt_fgenesh1_pg.C_1980013
<i>H. magnipapillata</i>	NCBI	XP_002167158.1
<i>P. carnea</i>	NCBI	AAA91336
<i>N. vectensis</i>	JGI	WISE2 reprediction based on fgenesh1_pg.scaffold_4621000001 e_gw.113.12.1.

Appendix 5 – Parameters used in phylogenetic tree construction

For the phylogenetic tree of fibrillin sequences the following parameters were used: Tree based on alignments of whole protein constructed with MEGA 5 (aligned with the MEGA 5 BLOSUM matrix (Gap opening penalty 5, gap extension penalty 0.5, BLOSUM matrix (residue specific and hydrophobic penalties on, gap separation distance 4, end gap separation off, negative matrix off, delay divergent cut off 30%)), and manually edited to improve alignment of cysteine residues and also remove potentially bad exon predictions. Tree constructed using MEGA 5 and the following settings: Analysis - Phylogeny Reconstruction, Statistical Method - Maximum Likelihood, Test of Phylogeny - Bootstrap method, number of Bootstrap Replications – 100, Substitution Model: Substitutions Type - Amino acid, Model/Method - Jones-Taylor-Thornton (JTT) model, Rates among Sites - Uniform rates, Gaps/Missing Data Treatment - Partial deletion, Site Coverage Cutoff (%) – 95, Tree Inference Options, ML Heuristic Method - Nearest-Neighbour-Interchange (NNI), Initial Tree for ML- Make initial tree automatically, No. of Seqs : 43, No. of Sites : 1618, No Of Bootstrap Replications = 100

For the phylogenetic tree of LTBP sequences these parameters were used: Tree based on of alignment of the proteins from the first EGF of the tandem array (EGF3 in most LTBPs, starting with residue 874 in human LTBP1) to the C-terminus with MEGA 5 (aligned with the MEGA 5 BLOSUM matrix (Gap opening penalty 5, gap extension penalty 0.5, BLOSUM matrix (residue specific and hydrophobic penalties on, gap separation distance 4, end gap separation off, negative matrix off, delay divergent cut off 30%)), and manually edited to improve

alignment of cysteine residues and also remove potentially poor exon predictions. Tree constructed using MEGA 5 and the following settings: Analysis -Phylogeny Reconstruction, Scope - All Selected Taxa, Statistical Method - Minimum Evolution method, Test of Phylogeny - Bootstrap method, No. of Bootstrap Replications 500, Substitution Model: Substitutions Type - Amino acid, Model/Method - Poisson model, Rates among Sites - Uniform rates, Pattern among Lineages - Homogeneous, Gaps/Missing Data Treatment - Pairwise deletion, Tree Inference Options; ME Heuristic Method - Close-Neighbour-Interchange (CNI) Initial Tree for ME - Obtain initial tree by Neighbour-Joining, No. of Sites : 1605, No Of Bootstrap Replications = 500