

UNIVERSITY OF KENT

Masters by Research Thesis

The role of talin splice variants in health and disease

Author: W J S Edwards

Degree in biochemistry

Academic advisor: B T Goult

Word count: 28,771

2022

Contents

Abstract 7

Copyright statement 8

Acknowledgements 8

Abbreviations 8

Chapter 1: introduction 10

1.1 Talin, the extracellular matrix and their interactive nature 10

1.1.1 Cell adhesion 10

1.1.2 The extracellular matrix 10

1.1.3 The focal adhesion 12

1.1.4 The structure of talin 1 head and rod domains 16

1.1.2 The effect of mechanical force on talin 19

1.1.5 Vinculin 21

1.1.6 Vinculin allows continued survival of adhesions 23

1.1.7 The pathogenic effects of faulty talin 23

1.1.8 Altered ECM in cancer 23

1.2 Genetic splicing and errors therein 25

1.2.1 Overview of the concept of gene splicing 25

1.2.2 splicing regulation 27

1.2.3 Diseases caused by splicing errors 27

1.3 The discovery and importance of exon 17b 28

1.4 Systemic capillary leak syndrome (SCLS) 28

1.4.1 Background of disease 28

1.4.2 The lack of a documented SCLS inheritance pattern 30

1.5 Intrabodies and antibodies 30

1.5.1 Overview of antibodies and their production 30

1.5.2 Phage display 31

1.5.3 The origin of the intrabody 32

1.6 Project Aims 33

Identification and characterisation of a cancer associated novel exon in the TLN1 gene 33

Investigation into the effects of an exon skipping event correlated with SCLS 34

Testing the viability of a new, ethical antibody alternative 34

Chapter 2: Materials and methodologies 35

2.1 Media and buffers 35

2.2 Plasmid constructs (SCLS 17b and intrabodies) and primers 36

2.3 Production of bacterially expressed protein 38
2.3.1 Transformation of plasmids into *E. coli* 38
2.3.2 Protein expression in *E.coli* 38
2.3.3 Expression of ^{15}N labelled protein 38
2.3.4 Subcloning (digestion ligation) 39
2.3.5 Agarose / DNA gels 39
2.3.6 SDS-PAGE gels 40
2.4 Protein extraction and purification 40
2.4.1 Cell harvesting 40
2.4.2 Cell lysis 40
2.4.3 His tagged protein purification 40
2.4.4 Sample Dialysis 41
2.4.5 Ion chromatography (Q and S) 41
2.4.6 GST chromatography (gravity flow) 42
2.4.7 Measuring protein concentration (Bradford/nanodrop) 42
2.5 Biochemical characterisation of protein 43
2.5.1 Bioinformatics (UniProt alignment etc.) 43
2.5.2 Structural modelling (AlphaFold) 43
2.5.3 Circular dichroism 43
2.5.4 Size exclusion chromatography 43
2.6 Structural biochemistry 44
2.6.1 Nuclear magnetic resonance (NMR) 44
Chapter 3 Results 45
3.1 A novel exon in the TLN1 gene and its role in cancer pathogenicity 45
3.1.1 The discovery and initial work on exon 17b 45
3.1.2 The R1R2-17b insertion 49
3.1.3 The implications of 17b going undetected until now 49
3.1.4 Predictive modelling of the exon 17b insertion 50
3.1.5 17b Biophysical characterisation 57
3.1.7 17b Biochemical characterisation/functionality 63
3.1.8 The RIAM binding site in R1R2 remains functional 64
3.1.9 R1R2-17b retains vinculin binding capability 67
3.1.10 R1R2-17b shows increased affinity for vinculin 69
3.1.11 Production of 'stretchable' R1R2-WT/17b 72
3.1.12 Altered force loading capability of R1R2-17b 77
17b discussion 79

3.1.13 17b is functionally and structurally different from the WT 79
3.1.14 Predicted effect of R1R2-17b on focal adhesions 80
3.1.15 The role of R1R2-17b in cancer motility and pathogenicity 82
3.1.16 R1R2-17b may allow cells to change the mechanosensitive ability of their focal adhesions in response to injury related matrix stiffening 85
3.1.17 Usage of R1R2-17b as a risk stratification method for breast cancers 88
Chapter 3.2 The role of the systemic capillary leak syndrome deletion in the pathogenicity of the disease 88
3.2.1 Discovery of a talin exon skipping event in SCLS patients 89
3.2.2 SCLS predictive modelling 90
3.2.3 SCLS biophysical characterisation 95
3.2.4 SCLS biochemical characterisation/functionality 97
3.2.5 Initial size exclusion analysis of R12R13-SCLS 97
3.2.6 Comparison of vinculin binding between R12R13-SCLS and WT 98
SCLS discussion 101
3.2.7 R13-SCLS less stable than WT, but the VBS not more active/less cryptic 101
3.2.8 Preclusion of R12R13-SCLS from actin binding assay 101
3.2.9 Reduced stability in SCLS talin leads to increased vinculin localisation 102
3.2.10 Over-active talin outcompetes cadherin junctions for stabilising vinculin 105
3.2.11 R13-SCLS causes weakened adherens junctions leading to capillary leakage 105
Chapter 3.3 intrabodies 108
3.3.4 The process of intrabody evolution 108
3.3.5 Intrabody functionality 110
3.3.6 Validation of XPH 238 binding talin 1 R2 110
3.3.7 Validation of XPH 241 binding talin 1 R11 111
3.3.8 Validation of XPH 242 binding talin 1 R11 112
3.3.9 Mapping of epitopes 116
3.3.10 Mapping the epitope of XPH238 117
3.3.11 Mapping the epitope of XPH241 121
3.3.12 Mapping the epitope of XPH242 124
3.3.13 Validating intrabody specificity 127
Intrabodies discussion 131
3.3.14 Functionality as an antibody alternative 131
3.3.15 Usage of intrabodies in future talin experiments 131
Chapter 4: Future work 134
4.1 Generation of crystal structures for R1R2-17b and Intrabodies 134

4.2 Acquiring a Kd of the intrabodies using micro-scale thermophoresis 134
4.3 R1R2-17b GFP pulldown in cancer lysate 135
4.4 Investigation of the relationship between exon 17b and CLSTN1 135
Chapter 5: Final conclusions 136
Bibliography 138

Figure 1: Simplified depiction of a focal adhesion 12
Figure 2: Process of the formation of a mature focal adhesion 15
Figure 3:Simplified depiction of talin 1 with annotated binding sites 18
Figure 4: Mechanical unfolding of talin and exposed/cryptic binding sites 20
Figure 5: Structure and activation of vinculin 22
Figure 6: Alternate mRNA splicing overview 26
Figure 7: Overview of systemic capillary leak syndrome 29
Figure 8: Intrabody structure and hypervariable region 33
Figure 9: BioMedX data on the novel exon 17b 49
Figure 10: R1R2-WT/17b sequence alignment 50
Figure 11: Reliability of AlphaFold2 R1R2-17b prediction 52
Figure 12: Predictive modelling of talin 1 WT/17b 54
Figure 13: Alterations to talin 1 R1R2 helices due to exon 17b 56
Figure 14: Circular dichroism comparison of R1R2-WT/17b 59
Figure 15: The effect of DTT on reducing R1R2-WT dimerisation 61
Figure 16: Biophysical characterisation of R1R2-17b 62
Figure 17: R1R2-WT/17b NMR spectra overlay 63
Figure 18: R1R2-17b RIAM binding 67
Figure 19: R1R2-17b initial vinculin binding gel filtration data 69
Figure 20: R1R2-17b vinculin binding functionality 71
Figure 21: pGEX-Stretch experimental method 73
Figure 22: pGEX-Stretch subcloning process 75
Figure 23: pGEX-Stretch subcloning results 76
Figure 24: comparison of force loading capacity of R1R2-WT and 17b via single molecule stretching 78
Figure 25: Predicted effect of exon 17b expression in focal adhesion formation 81
Figure 26: R1R2-17b in vitro data 83
Figure 27: Expression of exon 17b to downregulate ECM force signals 87
Figure 28: SCLS Exon 54 skipping mutation family pedigree 90
Figure 29: Reliability of AlphaFold2 R12R13 WT/SCLS prediction 92
Figure 30: Predictive modelling of the effect of exon 54 skipping 94
Figure 31: Circular dichroism comparison R12DD-WT/SCLS 96
Figure 32: Gel filtration baseline data for R12DD-WT/SCLS 98
Figure 33: Vinculin binding ability between R12DD-WT/SCLS 100
Figure 34:In vitro localisation of FA components in WT/SCLS expressing cells 104
Figure 35: Predicted effect of exon 54 skipping in FA and adherens junction stability 107
Figure 36: Phage display process 109
Figure 37:NMR spectra comparing binding ability of intrabodies XPH238,421,242 115
Figure 38: XPH242 binding and identification of slow exchange interaction 116

Figure 39: Reliability of AlphaFold2 XPH238 prediction 119
Figure 40: Analysis of XPH238 binding and identification of R2 binding site 120
Figure 41: Reliability of AlphaFold2 XPH241 prediction 122
Figure 42: Analysis of XPH241 binding and identification of R11 binding site 123
Figure 43: Reliability of AlphaFold2 XPH242 prediction 125
Figure 44: Analysis of XPH242 binding and identification of R11 binding site 126
Figure 45: Intrabody specificity, inability to bind other talin domains 128
Figure 46: intrabody specificity, lower affinity for talin domain splice variants 130
Figure 47: Use of intrabodies as 'reader' molecules 133

Abstract

Talin is a protein crucial to the function of cell-matrix focal adhesions, which allows a cell to interact with its environment. In this thesis, I show that alternate splicing of talin, whether by mutation or by cassette exon splicing contributes to the pathogenicity of multiple diseases, by affecting the form and function of focal adhesion complexes. Two variations in alternate talin splicing were examined, the loss of exon 54 and the insertion of novel exon 17b.

My work, along with the work of my collaborators on both splice variants, has revealed that the novel talin exon 17b, is highly expressed in many cancers, and alters the structure of the R1 and R2 rod domains of talin, resulting in focal adhesions that display altered force loading and mechano-sensing properties. The data presented here suggests that exon 17b is expressed by cancers to gain drug resistance and evade the immune system. I also propose a function for exon 17b in non-cancerous human cells, that the exon is spliced into talin 1 to reduce the formation of excess scar tissue by deactivating myofibroblasts.

I have also worked with collaborators who have discovered a family who suffer from systemic capillary leak syndrome, due to an autosomal dominant mutation, causing an alternate splicing event of talin 1. My work in analysing the impact of exon 54 removal on the R13 rod domain of talin, has shown the resulting deletion creates an over-active talin that 'leeches' vinculin from surrounding cell-cell adherens junctions. Preventing vascular tissue reorganisation, resulting in rapid loss of fluid from the vascular system into surrounding tissue.

The thesis also discusses my work with a research group from Bordeaux who have developed an artificial, bacterially expressed antibody alternative. Using nuclear magnetic resonance, I show that these 'intrabodies' bind their targets with high specificity and affinity, showing potential as a replacement for monoclonal antibodies.

Copyright statement

All work shown here (unless stated to be from collaborators or other sources) has been solely produced by myself.

Acknowledgements

I would like to thank Professor Ben Goult for his invaluable help as both my supervisor and with the research presented in this paper.

I would also like to thank the many people with whom I got to collaborate on the contained work with.

Finally, I would like to thank Matt, Jack, Valentina, Yasumi, Neil and especially Lorena (for my endless questions) from the Goult lab.

Abbreviations

AA: Amino acid

AJ: Adherens junction

BRCA: Breast cancer

BSA: Bovine serum albumin

CCLE: Cancer cell line encyclopaedia

CD: Circular Dichroism

COADREAD: Colorectal adenocarcinoma

ECM: Extracellular matrix

EGFR: Epidermal growth factor receptor

ErbB: Erythroblastic oncogene B

FA: Focal adhesion

Fn3: Fibronectin type III domain

GFP: Green fluorescent protein

GSEA: Gene set enrichment analysis

GST: Glutathione S-transferase

IAC: Integrin adhesion complexes

IBS: Integrin binding site

IPTG: Isopropyl β-D-1-thiogalactopyranoside

kD: Dissociation constant

mAbs: Monoclonal antibodies

MEF: Mouse embryonic fibroblasts

MST: Micro scale thermophoresis

NEB: New England Biolabs

NMR: Nuclear magnetic resonance

NSCLC: Non-small cell lung cancer

OD: Optical density

PGs: Proteoglycans

RFP: Red fluorescent protein

SCLS: Systemic capillary leak syndrome

SDS PAGE: Sodium dodecyl-sulphate polyacrylamide gel electrophoresis

SDS: Sodium dodecyl sulphate

SEC: Size exclusion chromatography

TEMED: N Tetramethylethylenediamine

TEV: Tabaco etch virus

TGF-β: Transforming growth factor beta

TM : Melting temperature

Chapter 1: introduction

1.1 Talin, the extracellular matrix and their interactive nature

1.1.1 Cell adhesion

The eukaryotic cell and species composed of them are defined by their ability to coordinate and operate as a single organism and form tissues, organs, and large, complex biological structures (apart from unicellular eukaryotes such as protozoa). They are also defined by their ability to form connections, called cell adhesions. These adhesions can be to other cells or to the non-cellular structure that surrounds eukaryotic tissue, the extracellular matrix (ECM) (Khalili and Ahmad, 2015). These adhesions that are formed between the cell and the ECM are composed of many specialised proteins that act as an intermediary between the cellular cytoskeleton and the surrounding cellular environment. Perhaps most crucial to these cell-matrix adhesions is the protein talin. This large, multidomain protein spans the whole adhesion, between the cell membrane bound integrins to the F-actin fibres of the cytoskeleton. It is talin, its role in the cell, and the consequences of its variation that this project will centre around.

1.1.2 The extracellular matrix

The environment surrounding cells in human and animal tissues is a complex matrix of non-cellular protein components, and It is to this extracellular matrix (ECM) that eukaryotic cells must form adhesions to hold them in place and form tissues. The first of two main protein families found within the ECM are the tough fibrous proteins such as collagen, laminin, elastins and fibronectins (Frantz, Stewart and Weaver, 2010). These fibrous proteins form a strong ‘scaffold’ which cells can interact with and adhere to form ordered tissues. The second major type of protein found in the ECM are the proteoglycans (PGs). Proteoglycans

do not form an ordered structure, but instead fill the gaps between the fibrous scaffold by forming a hydrated gel which serves to both physically support the fibrous proteins as well chemically interacting with the cells within the ECM, with PGs being able to modulate certain cellular processes. As these PGs are a buffer region between cells (and help maintain a homeostatic environment) they are often important in cellular signalling pathways, and have also been shown to play a role in everything from cell migration to cell proliferation (Schaefer and Schaefer, 2010).

Of the fibrous proteins that make up the scaffold, the majority of the structure is composed of collagens, which make up 30% of the structural protein present in the ECM (Frantz, Stewart and Weaver, 2010)). While fibronectins do not make up the bulk of the ECM as a whole, they are intimately involved in cellular adhesion. Fibronectins have a similar structure to that of talin, with both being composed of repeating subunits that are 'mechanically sensitive', able to stretch/unfold to many times their original length in response to force. The unfolding of these proteins is due to cellular tension/traction produced by the adhesion of cells to the ECM (Smith *et al.*, 2007). The motility and force applied by cells adhered to fibronectin unfold its compact regions, exposing previously cryptic integrin binding sites (IBS) that are normally buried within the structure. It is these integrin binding sites that allow the interaction between cells and the structural 'scaffold' portion of the ECM.

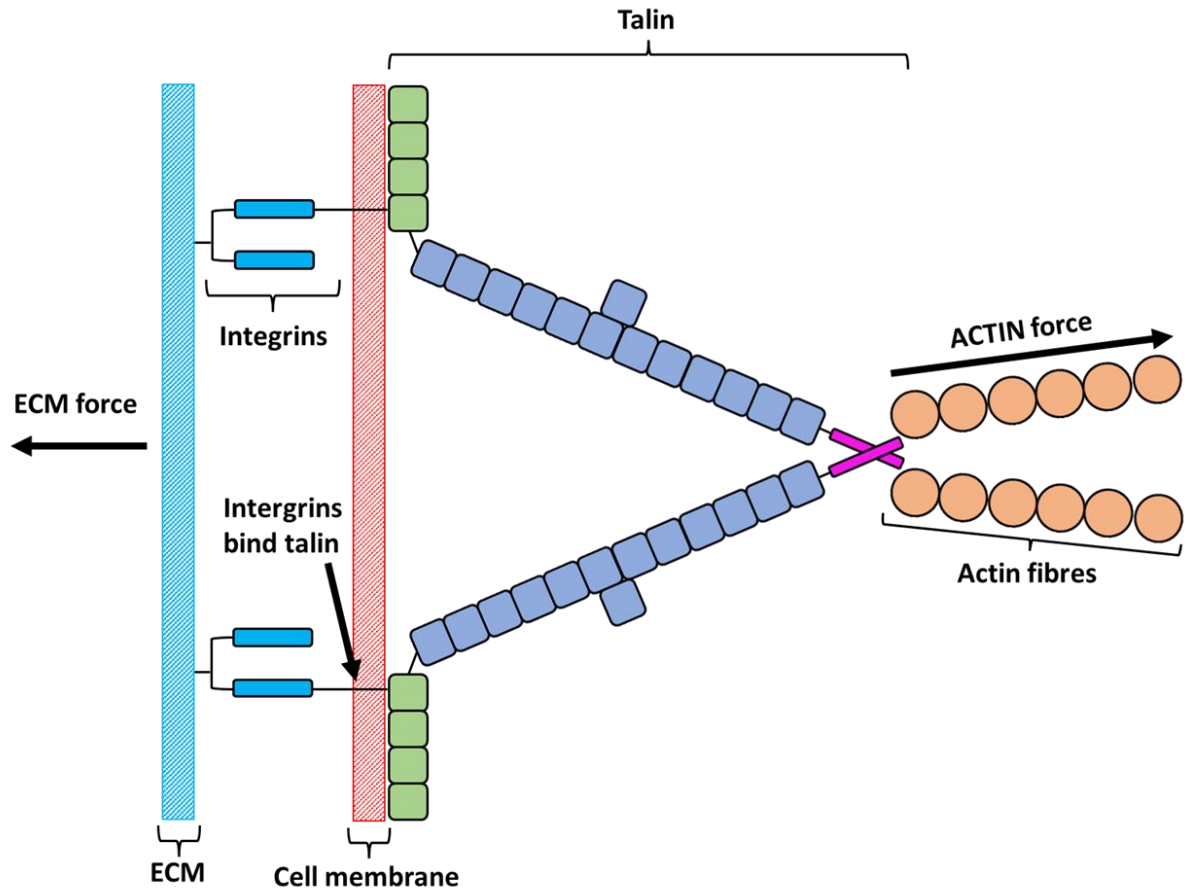


Figure 1: Simplified depiction of a focal adhesion

Cartoon showing a simplified eukaryotic focal adhesion complex with focus on talin 1. The talin 1 rod domain (blue) is separated into domains (blue boxes), and the head domain is shown in green. The dimerisation domain (DD) is shown in magenta connecting two talin proteins in a dimer as is found in cells. Integrins are shown in light blue, attached to both talin and the ECM (blue line) and spanning the cell membrane shown in red. Actin fibres are shown bound to the C-terminal end of the protein (beige). The forces exerted on the FA by the ECM and cytoskeleton are shown as black arrows and labelled appropriately. (Diagram adapted from Klapholz and Brown 2017)

1.1.3 The focal adhesion

For cells to 'interact' with the surrounding ECM in a mechanical fashion (as opposed to chemical interaction such as cytokines) they do so via focal adhesions (FAs). These are a specialised type of integrin adhesion complex (IAC), composed of many specialised adhesion proteins which anchor the cell to the ECM as well as forming a mechanically sensitive link between the ECM and the cell. This linkage allows cells to 'sense' mechanical forces applied

via the matrix and allows them to apply force to the matrix in return. FAs span the cell membrane via integrins, which act as transmembrane cellular receptors for mechanically exposed IBS present on fibronectins (Wu, 2007). The transmembrane and cytoskeletal regions of a FA linkage are modulated by the adhesion protein talin (among other proteins such as focal adhesion kinase (FAK)) which in turn is anchored to the cytoskeletal F-actin fibres as shown in fig 1. Because of its modulating role, activation of talin is central to the formation of a FA. Prior to FA formation, talin begins in an autoinhibited state where its long rod domain (fig. 2A, blue/red region) is folded tightly around its head region (fig. 2A green) (R9 domain bound to the F3 domain in the head). The interaction between these regions keeps talin in a tightly closed cytosolic form (Wegener *et al.*, 2007 fig. 2A). However, the presence of PIP2 (which is correlated with high interaction between talin and integrin (Martel *et al.*, 2001)) causes talin to unfold (fig. 2B) and expose an integrin binding site 1 (IBS1), which is located within the head region (F3). The exposed IBS allows binding of the talin head to the β-integrin cytoplasmic tail, this 'couples' the ECM to internal cellular proteins (talin) via the integrins.

A mature and functional FA requires connection to the F-actin cytoskeleton of the cell. Once talin is anchored to the ECM via integrins, F-actin fibres, the cytoskeletal portion of the FA complex, bind actin binding sites (ABS) within talin. (fig. 2C red) ABS3, which is located on the far C-terminal end of the talin protein (R13-dimerisation domain), is the initial binding site for F-actin fibres as shown in fig. 2C (Klapholz and Brown, 2017). ABS3 and ABS2 are initially kept autoinhibited by R2,R3 and R9 (Atherton *et al.*, 2015a), as seen in fig. 2D, mechanical force applied by the actin cytoskeleton (after adhering to ABS3) unfolds domains within the rod which allow the binding of another key FA component, vinculin (fig. 2 orange). Vinculin binding activates the rod domains that make up ABS2, activation of ABS2

recruits F-actin fibres, as well as stabilising the bound actin with vinculin. Vinculin is able to bind to both the actin filament and the vinculin binding sites (VBSs) within the unfolded rod region. The unfolded domains, which have exposed VBS, can also recruit further F-actin filaments via talin-vinculin-actin interactions. This second stage of actin fibre binding is required for maximal traction force and stability of complete FAs (Atherton *et al.*, 2015a).

FAs allow a cell to alter their environment by 'pulling' on these talin mediated linkages, but they also allow the ECM to impact the cell, as mechanical force applied by the ECM is transmitted via FAs and can affect mechanical sensitive proteins such as talin, altering talins ability to bind ligands (such as vinculin), and therefore altering chemical processes within the cell (Iskratsch, Wolfenson and Sheetz, 2014).

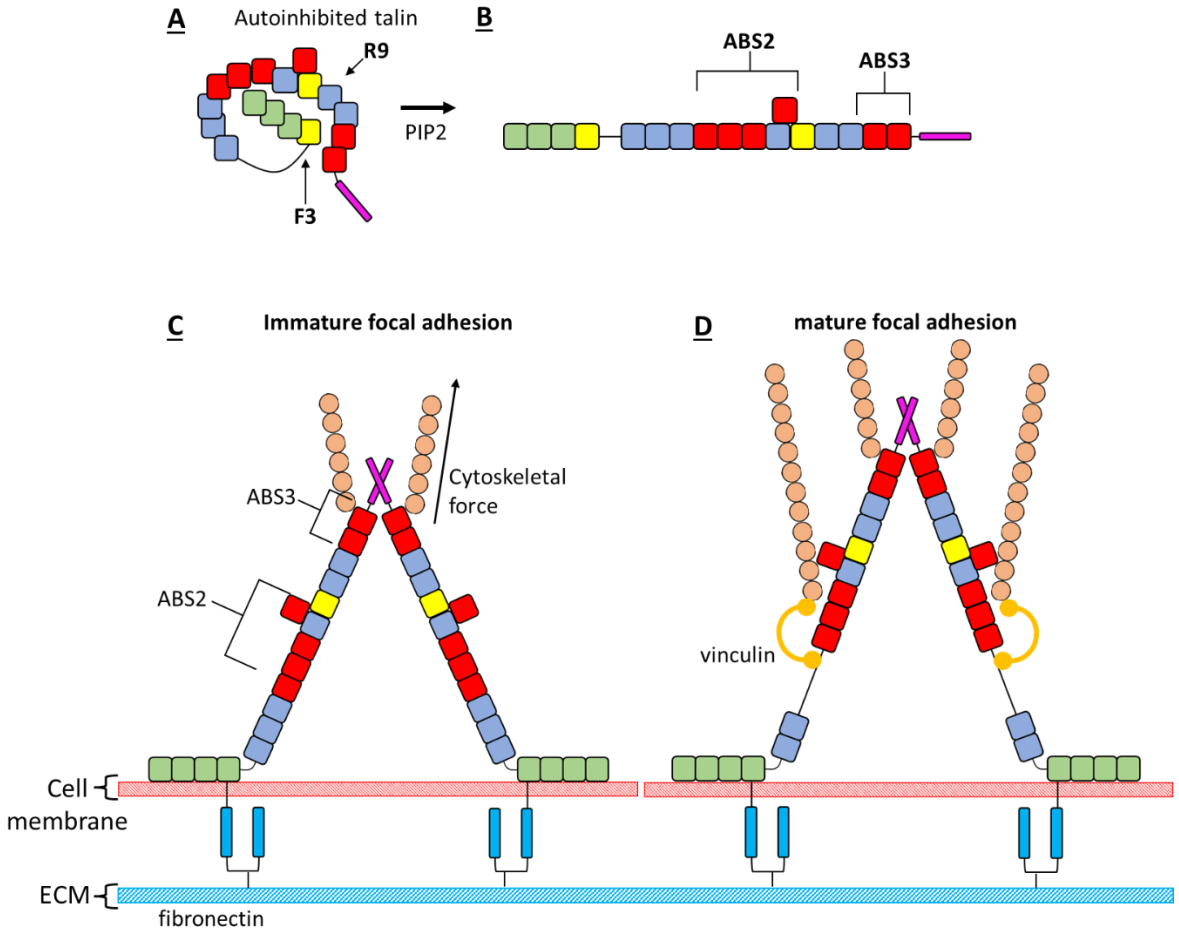


Figure 2: Process of the formation of a mature focal adhesion

Images showing the formation of a focal adhesion, highlighting the roles of the actin, integrins, vinculin and talin. A). Cartoon showing the autoinhibited state of talin, (rod domains in blue, head domains in green, DD domain in magenta, ABS highlighted in red) with the self-binding F3 and R9 regions in yellow. B). the uninhibited from of talin 1 after activation via PIP2. C). Initial formation of focal adhesion, with head domains interacting with integrins (light blue), which have bound activated fibronectin (light blue line) in the ECM, interacting across the cell membrane (red line). F-actin binds ABS3 in R13-DD, anchoring the FAs to the cytoskeleton which applies mechanical force. D). Mature focal adhesion. Mechanical force from the cytoskeleton unfolds the vinculin binding sites within ABS2. Activated vinculin (yellow), binds the VBS on ABS2, recruiting actin to ABS2 where it can bind and stabilise the focal adhesion. (A. and B. adapted from Rosemarie E Gough, Benjamin T Goult FEBS Lett. 2018)(C. and D. adapted from Klapholz and Brown 2017)

1.1.4 The structure of talin 1 head and rod domains

Talin is the mediator between the eukaryotic cell and its environment. Talin is perhaps the most important protein within cell-matrix adhesions. Talin is a 270 kDa, mechanically sensitive, cytosolic protein found within cells, encoded by a gene found on human chromosome 9 and encoded by 57 exons under standard conditions. The protein contains 18 domains which are separated into two segments, the talin 'head' and the talin 'rod', shown in fig. 3

The head region is a made up of 4 globular proteins: 4.1 protein, ezrin, radixin and moesin (FERM), referred to as F0-F3 (fig. 3 green). The standard FERM domain, found within the majority of FERM-containing proteins, forms a 3 lobed structure made up of 3 globular proteins. However, the FERM domain found within both talin 1 and 2 contains an extra region, called F0. This extra domain is similar in homology (40% similarity) to F1 and sits at the far N-terminal end of the head region of talin 1 and 2 (Goult *et al.*, 2010). The FERM domain is a structure found in many cellular proteins that localise to the cell membrane (Pearson *et al.*, 2000). The FERM family of proteins have been shown to be involved in many protein-protein interactions. In talin 1 the FERM domain anchors the talin protein to transmembrane region of the adhesion complex, the β-integrins. These integrins are bound via the F3 region(fig. 3 blue box) of the FERM head (Elliott *et al.*, 2010).

β-integrins, are transmembrane proteins whose cytoplasmic domain extends into the cell and bind the F3 domain via an NPXY motif found on the integrin cytoplasmic tail, interacting with the IBS located on F3, IBS1 (Klapholz and Brown, 2017)(Calderwood *et al.*, 2002). The F1-F2 domains, along with the extra F0 domain have been shown to play a role in activation of the integrins which bind the FERM head. Furthermore, the F2-F3 domains serve an alternate purpose as an F-actin binding site, ABS1 (fig. 2 red box ABS1) (Lee *et al.*, 2004).

However, due to the overlapping integrin and actin binding domains within the FERM head, the binding of actin is mutually exclusive with the binding of integrins.

The FERM 'head' domain is attached via an unstructured linker to the rod domain. This rod domain of talin is a collection of 62 α-helices (McLachlan *et al.*, 1994) which are organised into 13 helical bundles of 4-5 amphipathic helices (fig.3 blue regions), and ending in a single C-terminal helix which allows the dimerisation of two talin proteins, the dimerisation domain (DD)(Smith and McCann, 2007) (fig.3 pink region). The helical bundles, referred to as R1-R13, are further subdivided into a compact region towards the N-terminus made up of domains R1-R5, and a linear region made up of the remaining R6-R13 bundles. These helical bundles form the bulk of the talin protein and contain many of the binding sites found within talin 1. The talin rod also contains a secondary β-integrin binding site (IBS2) (fig.3 blue box IBS2) which is found on the R11-R12 domains. IBS1 is the binding site of integrins in mature talin adhesions, while IBS2 is able to bind integrins while talin is in a semi-unfolded form. This intermediate form of the R11 and R12 helical bundles, has been shown to only be bound by pre-activated β-integrins (Rodius *et al.*, 2008)(Gough and Goult, 2018), implying its role as an intermediate binding site in the development of mature adhesion complexes.

Also contained within the rod domain are 2 further actin binding sites (ABS2 and ABS3), and each of these binding sites serves separate functions within talin as part of the FA. ABS2 is located on domains R4-R8 (fig.3 Red box ABS2), and ABS3 is located on the R13 and DD domains (fig.3 Red box ABS3) at the far C-terminal end of the protein (Hemmings *et al.*, 1996)(Gingras *et al.*, 2008). Similar to IBS2, (which serves as a secondary integrin binding site) ABS2 is not the main anchorage point for actin, as ABS3 serves as the main force bearing binding site for actin(Yao *et al.*, 2014)(fig.2). Instead ABS2 provides talin with the

ability to sense the stiffness of the surrounding ECM, as knockout of ABS2 prevents cell stiffness sensing (Kumar *et al.*, 2016). ABS2 is only active once force, exerted from actin via ABS3, opens up the helical bundles of R2R3 (R3 is the least stable of the bundles due to its high, destabilising threonine content in its hydrophobic core (Yao *et al.*, 2016)). When R2R3 are unfolded, exposed vinculin binding site within R2R3 recruit vinculin, which in turn activates the R4-R8 region which makes up ABS2, allowing vinculin mediated actin binding of ABS2 (Sun, Guo and Fässler, 2016) . ABS2 and ABS3 act as the main F-actin actin links. These linkages integrate talin into the cytoskeleton which provides cellular motility, via lamellipodia (a form of actin network) among many other roles (Stricker, Falzone and Gardel, 2010).

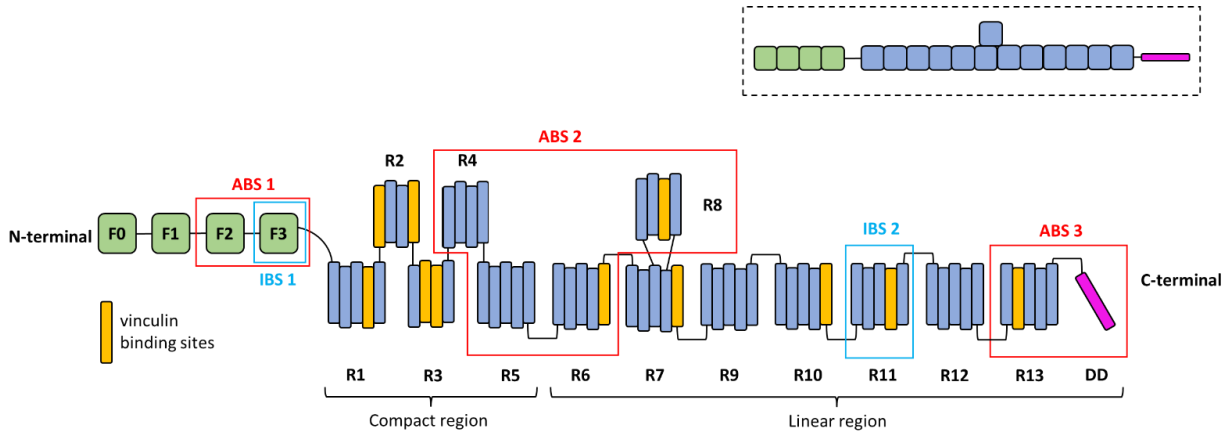


Figure 3:Simplified depiction of talin 1 with annotated binding sites

A cartoon representation of the talin 1 protein. With the FERM head domain (F0-F3) shown in green, and the helical bundles in the rod domain (R1-R13) shown in blue, with either 4-5 helices. Dimerisation domain (DD) is shown in pink. The talin 1 vinculin binding sites are shown by colouring the helices in which they are found orange. Actin binding sites are shown in red boxes and integrin binding sites are shown in blue boxes (adapted from Gough and Goult FEBS Lett. 2018)

1.1.2 The effect of mechanical force on talin

The mechanical force exerted on talin by either the cytoskeleton (via actin) or the ECM (via integrins) causes the helical bundles within talin to fold/unfold. It is this process that provides or prevents access of ligands to their cryptic binding sites within the helical bundles of talin.

Talin has binding sites for a large variety of intracellular proteins, many of which interact with the folded rod domains by binding the outwardly available surface of the closed helical bundles, such as CDK1 (Gough, 2018)(fig.4 left). However, talin also has ligands that can only bind once the helical bundles are unfolded.

Crucial to the function of talin is the protein vinculin. Vinculin is one of many talin ligands but it has the important property that vinculin binds an exposed VBS (such as the one within R2R3), and stabilises the bundles in their extended, unfolded state (their resting form is compact not extended). This prevents the bundles from relaxing back to their folded form (Atherton *et al.*, 2015b). The binding of vinculin keeps the bundles unfolded (fig.4 right) exposing cryptic binding sites to allow the binding of other ligands besides vinculin. Vinculin is of major importance to the function of FAs, once force has exposed VBS within the rod domain, vinculin maintains this unfolded configuration, allowing different focal adhesions to adopt a wide diversity signalling capabilities, despite all FAs containing the same protein components (Goult, Yan and Schwartz, 2018).

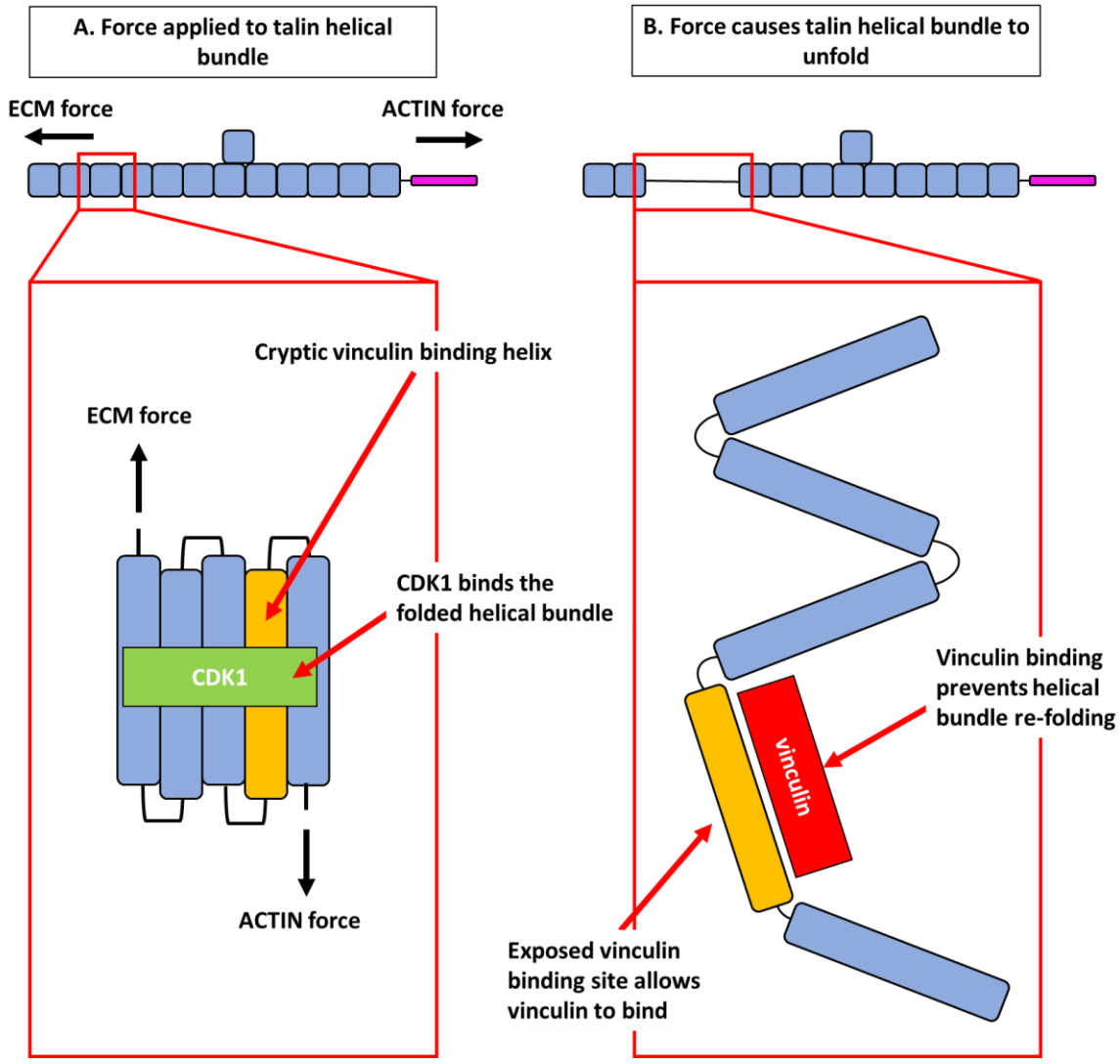


Figure 4: Mechanical unfolding of talin and exposed/cryptic binding sites

Simplified image showing the unfolding of a helical bundle within the talin rod region, and how mechanical force allows interaction of different ligands. The vinculin binding helix within the bundle is shown in orange and remaining helices are shown in blue. CDK1, a binder of folded bundles is shown in green, bound to folded conformation. On the right, vinculin, a binder of talins' unfolded conformation, is in red, shown binding the unfolded helical bundle via the orange VBS in the right of the image.

1.1.5 Vinculin

Vinculin is a 116 kDa talin interacting protein, that is important in the formation and function of many types of cellular junction, being crucial to both cell-cell adherens junctions and cell-matrix focal adhesions (Maddugoda *et al.*, 2007). The structure of vinculin is similar to talin, being made up of 7 helical bundles of 4 α-helices and a single 5 helix bundle. The bundles are subdivided into 5 domains, which comprise the vinculin head and vinculin tail regions. The head is made up of the first 4 domains (D1-D4)(fig.5 top, red box) and is a multi-lobed structure that interacts with α-catenin as well as talin (Carisey and Ballestrem, 2011).The tail region contains just D5 (fig.5 top, blue box), which is made up of 5 amphipathic helices (Critchley, 2000) and contains the vinculin binding site for F-actin.

Also, similarly to talin, vinculin is kept in an autoinhibited state where the head region binds the tail region in a 'pinching' motion (Fig.5, bottom left). This autoinhibition is mediated by the interactions of the vinculin tail (Vd5) with both Vd1 and Vd4(Cohen *et al.*, 2005). The presence of PIP2 (phosphatidylinositol 4,5-bisphosphate ($PI4,5P_2$)), a type of acidic phospholipid has been shown to prevent the interaction between the vinculin head region and the vinculin tail region (Gilmore and Burridge, 1996) this phospholipid converts vinculin to its constitutively active state (fig.5 bottom right). Once in its active form, the vinculin head is able to bind VBS located on talin and other adhesion proteins and the tail is able to recruit F-actin. This allows vinculin to reinforce cellular junctions (such as FAs) by additional integration of the F-actin cytoskeleton.

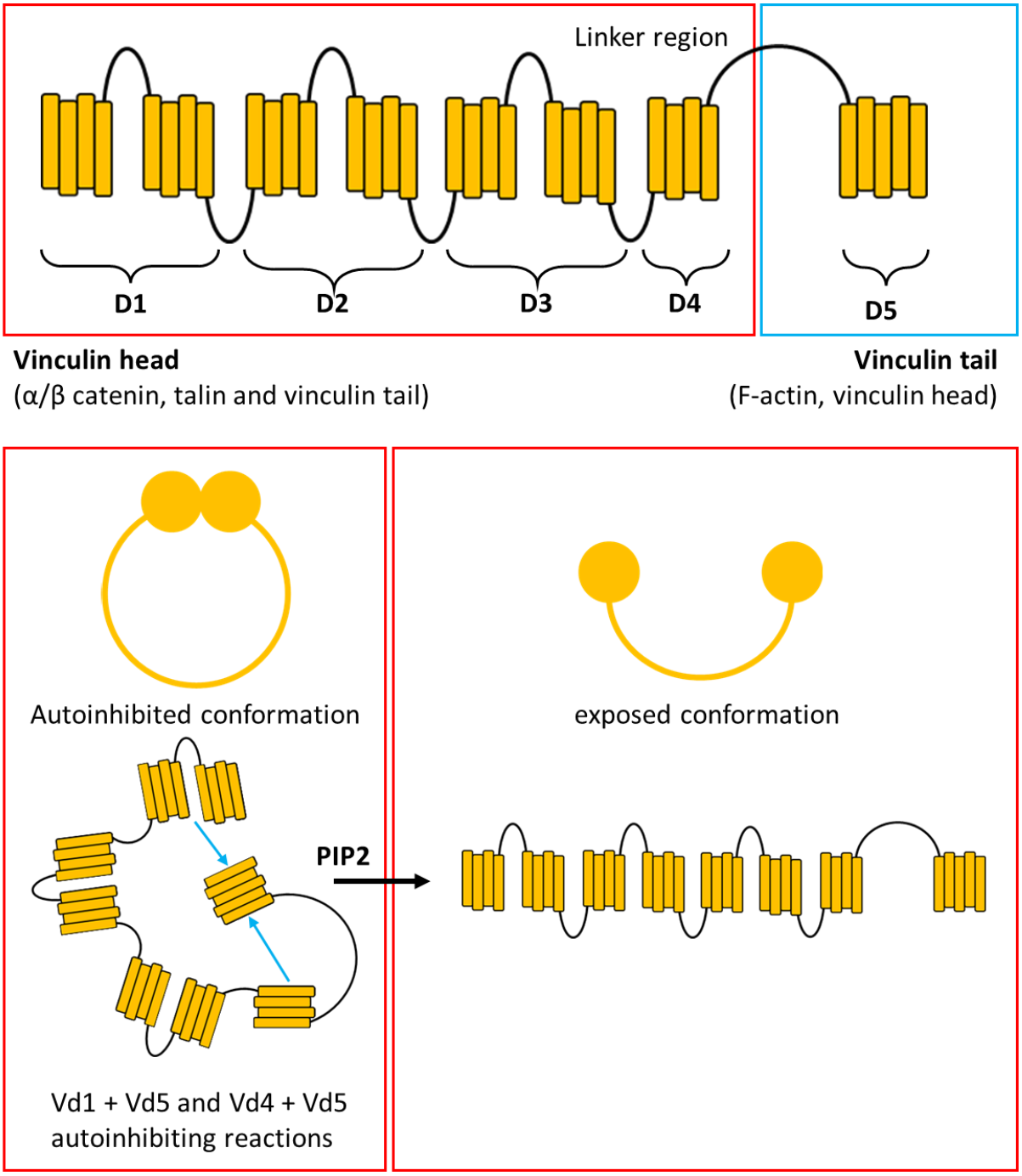


Figure 5: Structure and activation of vinculin

(top) Cartoon showing a simplified depiction of the vinculin protein as a collection of helical bundles (D1-D5), with the location of the vinculin head and tail regions marked in red and blue accordingly (with binding locations found within head and tail included). (Lower left-right) Further simplified image to show the unfolding of autoinhibited vinculin, with the tail 'pinched' between the head regions. (Adapted from Jennifer L Bays, Kris A DeMali Cell Mol Life Sci 2017).

1.1.6 Vinculin allows continued survival of adhesions

FAs are highly dynamic cellular machines, which can change size and assemble or disassemble rapidly in response to tension or signalling by the ECM or the cytoskeleton (Wolfenson *et al.*, 2009). Because of this dynamic nature, regulation of FA formation is important, and vinculin is a key component in the formation and survival of a focal adhesion. Once vinculin enters its active form and exposes its binding sites, vinculin migrates to FAs (Carisey and Ballestrem, 2011). The presence of activated vinculin has been shown to prevent the deconstruction of immature focal adhesions and promotes the continued existence of vinculin-bound FAs (Humphries *et al.*, 2007). Without vinculin present at these adhesions complexes, the complexes are quickly 'turned over' and disassembled.

1.1.7 The pathogenic effects of faulty talin

Due to its role as a major mechano-sensing and cellular movement hub, the dysfunction of talin would potentially be catastrophic to a cell (and therefore patient). Even slight changes in the structure of the protein could result in talin no longer binding or cause the over-interaction with, actin or other crucial adhesion proteins, both of which could have pathogenic consequences. For example it has been shown that the mutation of just 4 residues in the R3 rod sequence (mutation of T809/T833/T867/T901 to IVVI) prevents the function of the R3 domain VBS under normal levels of tension (Goult *et al.*, 2013a). This mutation drastically impacts the function of R3 and therefore altering the function of the adhesion complex as a whole.

1.1.8 Altered ECM in cancer

Cancer is commonly associated with disorganisation of tissues and cancer cells possess distinctly different characteristics from their WT cells. Often this is the failure of cellular components, including proteins involved in FAs. As mentioned earlier, mechanical forces

from the ECM are sensed by the cell through focal adhesions which respond with cytoskeleton-induced force in return. This back and forth tension is termed mechanoreciprocity, which is important for maintaining a normal homeostatic environment. Alterations to ECM component concentrations results in a different elastic modulus (such as softer/stiffer extracellular matrices)(Butcher, Alliston and Weaver, 2009). Matrices with different elastic moduli exert different mechanical forces on their surrounding cells, which results in increased or decreased survival of those cells. This is because all cell types have specific optimum matrix stiffnesses which promotes their existence (Guck *et al.*, 2005), but softer substrates are shown to generally decrease cell proliferation/survival (Lin *et al.*, 2015). The ability of the ECM stiffness to affect cellular processes relies on cells being able to sense the stiffness of their surroundings via focal adhesions and other integrin adhesion complexes. Tumours usually display abnormal stiffness sensing, and with the stiffness of their substrate no longer affecting them, cancer cell proliferation is unaffected by softer substrates(Lin *et al.*, 2015). As well as reduced stiffness sensing, cancer cells such as pancreatic cancer cells also display a threefold decrease in cellular elastic stiffness (Suresh, 2007). This loss of stiffness sensing 'unshackles' cancer cells from regulation via the ECM, which enables them to divide and replicate out of control.

As talin is a major component of FAs which are involved in stiffness sensing, it could be suggested that mutations/changes in talin expressed by cancer cells could render their FAs unable to 'sense' their environment correctly, misinterpreting extracellular signals, or preventing sensing all together. Both may aid in the pathogenicity of the cancer.

1.2 Genetic splicing and errors therein

1.2.1 Overview of the concept of gene splicing

During the production of protein within a cell, the DNA sequence is copied into mRNA (messenger RNA) which is then translated into the protein, this is referred to as the 'central dogma of molecular biology' and was proposed by Francis Crick in 1970 (Crick, 1970), and has been important for molecular biology since. In more recent times, it has become apparent that the production of a protein is far more complicated, with the translation of DNA into RNA being affected by numerous features such as alternative splice sites, transcription factors and even the structure of DNA chromatin itself (WANG *et al.*, 2015). All of these factors, but especially alternative splicing (which is central to this project) allows for a single region of DNA to be translated into many different proteins. Humans only have around 25,000 protein encoding genes, but express over 90,000 different proteins (Abdellah *et al.*, 2004) and it is alternative splicing that makes this possible. A protein encoding gene (DNA) contains regions called exons and introns, with exons providing the actual sequence of the desired protein and introns serving other non-coding purposes. Specialised restriction enzymes cut introns (and occasionally exons) out of the mRNA transcribed from the gene's DNA. After splicing out desired components, the remaining exons undergo ligation to form a completed protein encoding mRNA transcript. This process is initiated by a large ribonucleoprotein (RNA + protein) called the spliceosome (Will and Lührmann, 2011)

Normally exons are ligated in the order they appear in the original gene, however, alternative splicing allows exons to be shuffled, retained when they are normally cut out, or removed entirely, as shown in fig.6, there are several generally accepted alternative methods of splicing. Cassette exon skipping/retaining, fig.6F is both the most common and the most relevant to this project.

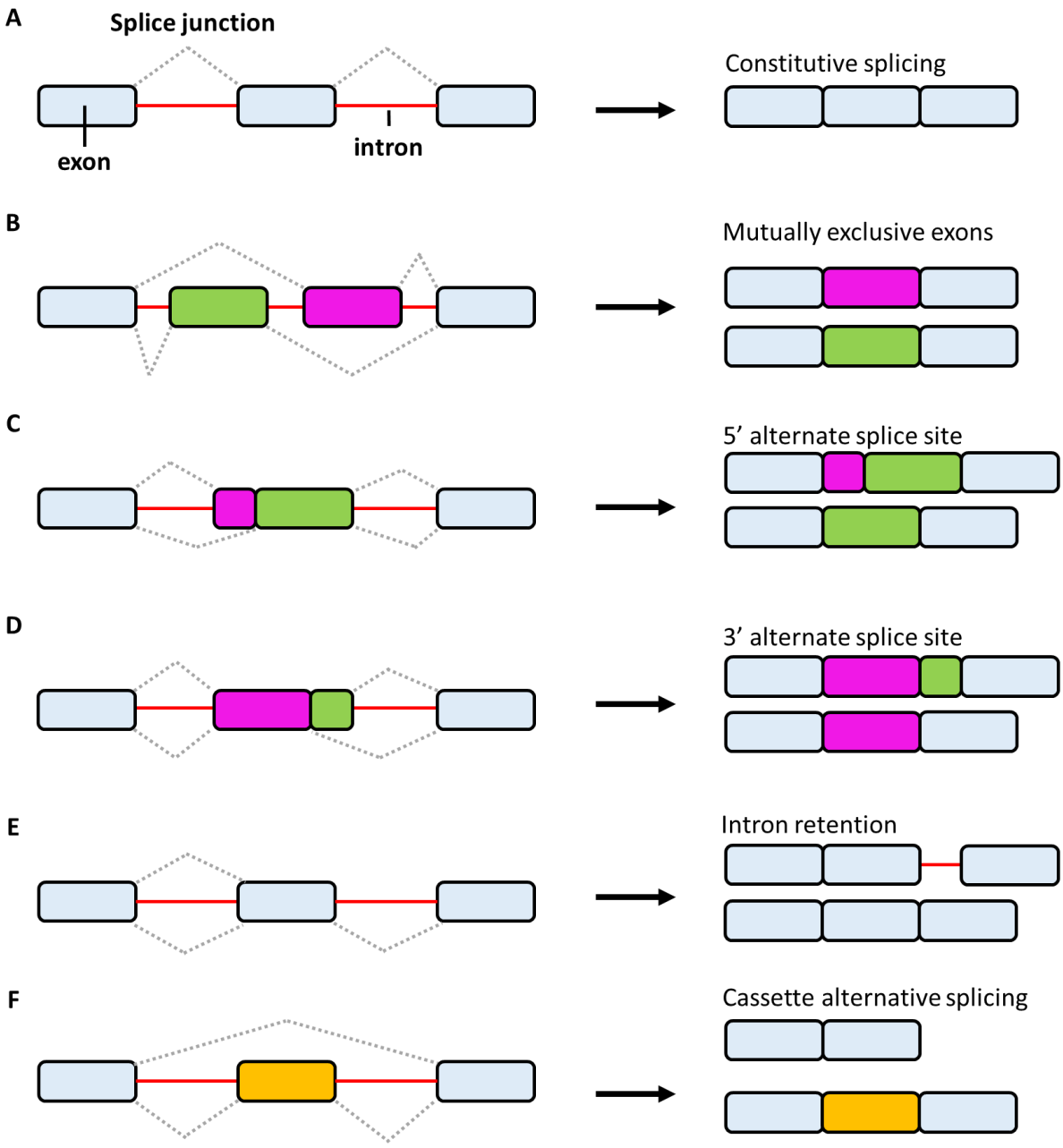


Figure 6: Alternate mRNA splicing overview

A cartoon depicting the main, generally accepted forms of alternate mRNA splicing. Exons are shown in blue and the introns in red, with exons affected by alternate splicing shown in purple and green. Alternative cassette exon splicing shown in orange. (Adapted from YAN WANG et al Biomed Rep. 2015)

1.2.2 splicing regulation

Strict regulation of the spliceosome is crucial, this includes regulation of both cis and trans elements within the cell. 'Cis' elements are regions within the mRNA code which allow recruitment of the spliceosome, while 'trans' elements are factors that bind to the mRNA. Regulation of the spliceosome subunits and its related splice factors is also very important to correct splicing. Errors in the regulation of one or all of these elements can lead to irregular splicing, in particular mutations in different 'cis' elements generally cause pathogenic splicing errors. If the mutation occurs within the 3' or 5' splice sites, the affinity of the spliceosome for the mRNA can be lowered, leading to incorrectly skipped exons or retained introns (Singh and Cooper, 2012). If the mutations affect enhancer or silencer regions which alter spliceosome affinity, then the expression of exons becomes irregular or the alternate exons can be overexpressed.

These kinds of splicing errors differ from other pathogenic errors such as point mutations, as the exons that are added in/taken out all encode for in frame changes, meaning rather than being stunted due to insertion of stop codons, large chunks of a protein can be missing or added while the rest of the protein remains intact and often semi functional.

1.2.3 Diseases caused by splicing errors

An example of the pathogenic results of splicing errors is spinal muscular atrophy, which is one of the major genetic causes of infant mortality. In 90% of cases this is caused by splicing errors or mutations in the survival motor neuron 1 gene (SMN1) which leads to the skipping of exon 7 (Singh and Singh, 2018). In the case of spinal muscular atrophy this skipping causes a lack of the SMN protein which is important in maintaining the survival of all animal cells. The resulting lack of this protein is also linked to other diseases such as amyotrophic

lateral sclerosis (ALS). These severe diseases both result from a single exon skipping event, demonstrating the damage even minor splicing errors can cause.

1.3 The discovery and importance of exon 17b

Recently, an alternately spliced variant of talin 1 has been discovered by a group in Germany, with whom I have collaborated on this project (Gallego-Paez *et al.*, 2022). Their discovery was of a talin exon which was unannotated in human exon databases, and which was highly expressed in certain types of cancer compared to its expression in WT cells. The new splice junctions, junction 18 and 20, indicated the presence of a new exon in the TLN1 gene, between exons 17 and 18. This novel exon was termed exon 17b. The presence of this exon in cancer suggests that this variant form of talin may have a role in the pathogenicity of 17b expressing cancers.

1.4 Systemic capillary leak syndrome (SCLS)

1.4.1 Background of disease

Systemic capillary leak syndrome (SCLS), sometimes referred to as Clarkson's disease (Clarkson *et al.*, 1960a), is an uncommon and life threatening disease (with a mortality rate between 20-35% (Clarkson *et al.*, 1960b)). It is defined by the rapid movement of intravascular fluids into their surrounding tissues (Marks and Shuster, 1973), this loss of fluid results in high blood cell concentrations (haemoconcentration) and low protein concentrations (hypoproteinaemia). The lowered blood pressure can cause acute kidney injury and result in the heart stopping due to hypovolemic shock (dangerously low blood pressure). Other symptoms can include syncope (loss of consciousness), myalgia (muscle pain) and polydipsia (extreme thirst) along with pulmonary, gastric and extremity edema (Siddall, Khatri and Radhakrishnan, 2017). These symptoms while significant, are not permanent but occur in 'episodes' (lasting between 48 to 96 hours) and are sporadic but

generally recurrent, with symptoms dissipating without targeted treatment other than dealing with resultant hemodynamic issues. This is done by introducing intravenous fluids, which increases blood pressure and reduces hypovolemia. However, this treatment comes with its own issues, as the capillaries of the patient are still 'leaky' so this fluid is often 'third spaced' and also enters the surrounding tissue (Hahn, 2021), worsening various edemas while improving heart and kidney function (fig.7). Onset of episodes can begin anywhere between child and adulthood. Fortunately, this condition is only known to affect a few hundred people worldwide.

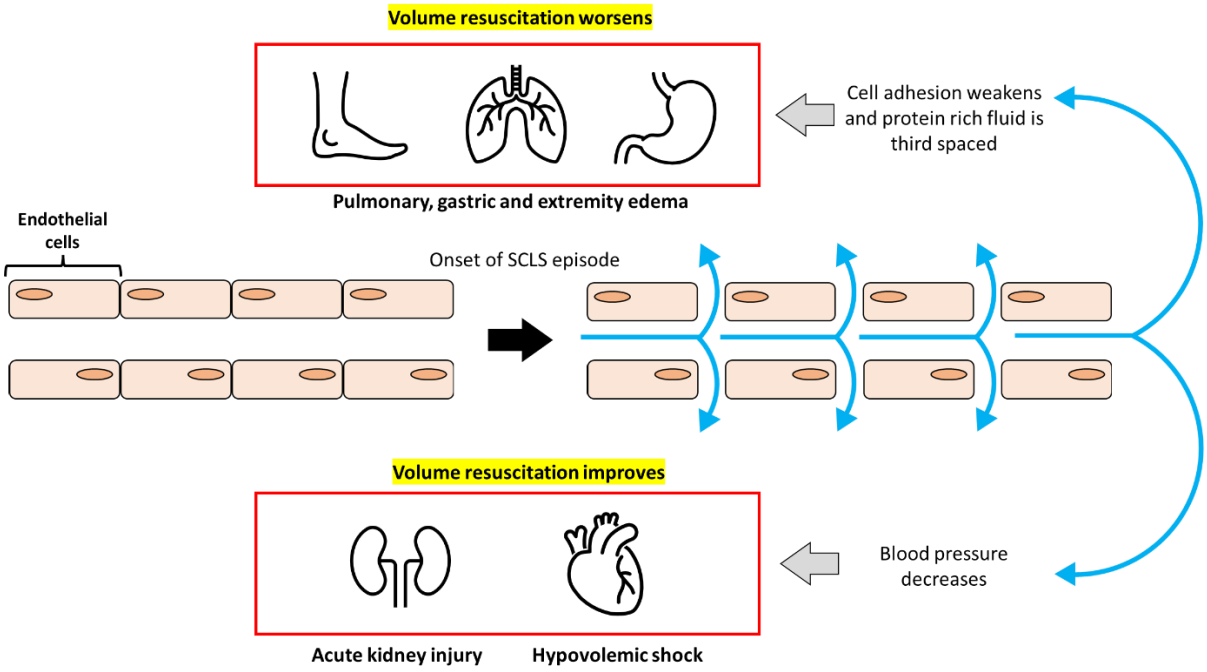


Figure 7: Overview of systemic capillary leak syndrome

Cartoon showing the effects of the onset of an episode caused by SCLS, and the effects of resulting treatment. Fluid loss due to leaky capillaries is shown in blue, as well as the symptoms resulting from the fluid loss which are highlighted in red boxes. (Adapted from Eric Siddall, Minesh Khatri, Jai Radhakrishnan Kidney International 2017)

1.4.2 The lack of a documented SCLS inheritance pattern

Until recently the cause of SCLS was not known, with no known pattern of inheritance. One similarity in unrelated patients that was known was that the majority (~80%) of adults suffering from SCLS also suffered from a type of asymptomatic plasma cell disorder, monoclonal gammopathy of undetermined significance (MGUS) (Hsu *et al.*, 2015). Confusingly this similarity between adult patients was not observed in any of the confirmed child patients. The role and correlation of MGUS in causing/impacting the disease has not yet been identified, and due to it lacking in child patients it is not considered a key identifier in diagnosing the disease.

Other avenues have been explored to try and find a cause of the disease. A genetic study of both adult and child patients (and their immediate relatives who did not exhibit the disease) was performed using whole exon sequencing (WES), with scans for hetero and homozygous variations as well as checking for enrichment of other genes/gene groups. The study found no consistent variations or a pattern of inheritance across its participants (Pierce *et al.*, 2019). However, recently an exon skipping event in the splicing of the TLN1 gene has been identified in family with a history of the disease, this splicing error will be one of the topics discussed in this project.

1.5 Intrabodies and antibodies

1.5.1 Overview of antibodies and their production

Antibodies are a type of specialised glycoprotein (glycosylated protein) produced by differentiated B immune cells that have become plasma cells (Lipman *et al.*, 2005). The glycoproteins are composed of 4 polypeptide chains (two light and two heavy), which are disulphide bonded together to form a 'Y' shape. The ends of each arm of the Y form specialised binding sites for a target epitope. Artificially produced antibodies have become a

useful tool in medicine since the first licensed antibody, muromonab-CD3, was used to prevent the rejection of kidney transplants in 1986 (Liu, 2014) (Zahavi and Weiner, 2020). antibodies have since been developed for use as treatments for a wide range of diseases, including cancers.

Most therapeutic antibodies are monoclonal, and rabbits are a common source for monoclonal antibodies (mABs). The rabbits are first immunized with the desired antigen for the antibody, once the rabbits immune system has mounted a response and has begun producing B-cells against the antigen, the B-cells are extracted and fused with a chosen rabbit myeloma cancer cell line with the addition of polyethylene glycol, these fusion cells are referred to as hybridomas which are functionally immortal cells that secrete the desired antibodies (Weber, Peng and Rader, 2017)

1.5.2 Phage display

Filtering out viable mABs from fusions that did not produce antibodies with desirable characteristics was originally crucial due to the difficulty in the production of rabbit hybridomas, and the difficulty in examining the many cultured cells.

To filter out viable antibodies a method was devised where the many antibody-encoding sequences produced by the B-cells could be inserted into a type of bacteriophage (bacteria-targeting-viruses) called a filamentous bacteriophage. The filamentous phage acquired the target gene by insertion of the DNA encoding for a fragment of an antibody/whole antibody into their filamentous phage gene III. Gene III encodes PIII, a minor viral coat protein. Insertion of a target protein sequence into gene III creates a fusion protein of both the target and PIII, which is displayed on the surface of the virus. The expression of the fusion protein does not affect the infectivity of the virus (Alfaleh *et al.*, 2020)(Stacey *et al.*, 1985).

These mABs produced by the hybridomas had the sequence encoding for their specific antibody inserted into the phage. These modified phage are compiled to produce a 'phage library', which contains phage that display all the different antibodies produced by the fusion cells. The viruses, displaying the antibody on their surface, could be filtered out by measuring their binding affinity to the intended antigen of the antibody. This was done using antigens immobilised on magnetic agarose beads, in a process not dissimilar to chromatography done to purify proteins.(McCafferty *et al.*, 1990). After displaying the target antigen to many different phage libraries, possible candidates (binders) are then filtered out by increasingly unfavourable conditions such as very low pHs (2.2), or trypsin and other proteases(Smith and Petrenko, 1997). After multiple repeats through the phage library, only phage displaying antibodies (or antibody fragments) with very high affinity for their target antigen remain.

1.5.3 The origin of the intrabody

Recently, a research group from Bordeaux have developed artificial proteins that could function as highly specific binding proteins that involve no animal vector in their production. These proteins termed 'intrabodies' are a bacterial protein containing a fibronectin type III domain (Fn3) that functions similarly to an antibody. Fibronectin type III domains are a type of module (usually ~100 residues) within a modular protein (Campbell and Spitzfaden, 1994), these domains are found in bacterial and animal proteins such as fibronectin and titin (Bork and Doolittle, 1992). The intrabodies developed by the group in Bordeaux produce Fn3 proteins with a 7 strand β-barrel structure (fig.8, blue). These barrels have a hypervariable region on their concave surface (fig.8, red region). This region is allowed to

mutate until it 'evolves' into a functional binding site for its intended target, resulting in a small, antibody like protein.

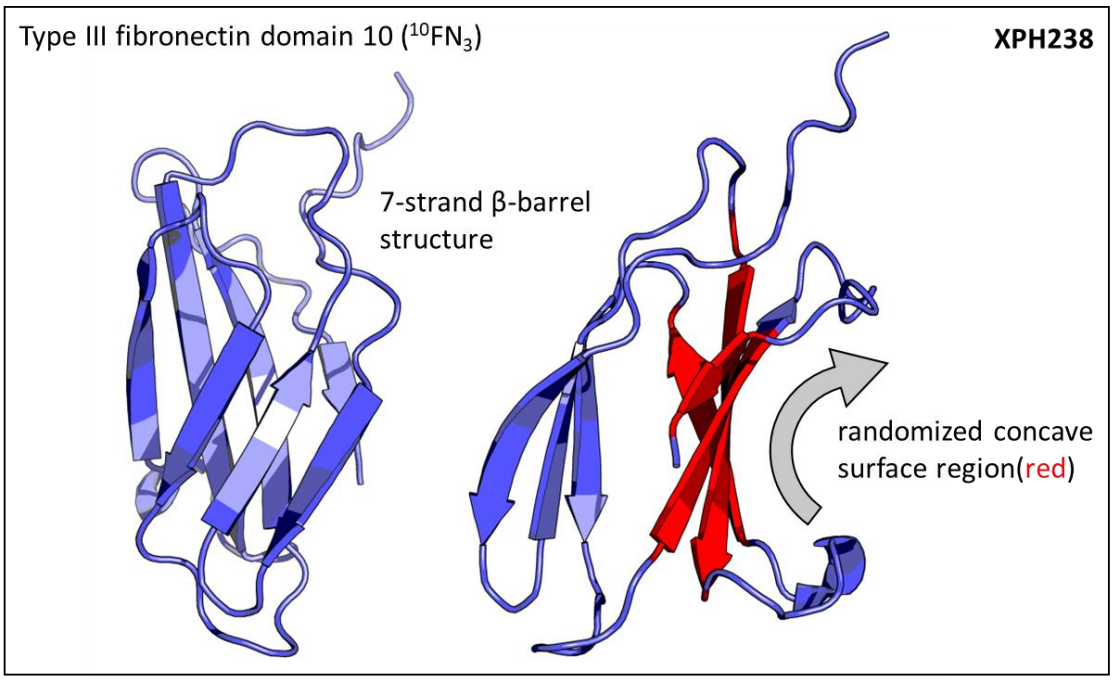


Figure 8: Intrabody structure and hypervariable region

PyMol image showing a basic overview of intrabody structure (using intrabody XPH238 as example). The fibronectin type III domain is shown in blue. Hypervariable randomised concave region is shown in red. This region is the intended binding site for produced intrabodies.

1.6 Project Aims

The first 2 sections of this project aim to look at 2 recently identified splice variants of talin 1 which have been correlated to certain cancers and a rare fluid disorder. The project will aim to show the effect to which the addition or removal of a single exon in each case can affect not just the proteins they encode but the function of the cell as a whole.

Identification and characterisation of a cancer associated novel exon in the TLN1 gene

The newly discovered exon 17b will be the topic of the first chapter, my first aim was to produce and purify samples of the alternately spliced protein, then to characterise its structure and physical differences from the WT. Once these have been established, I aim to

discover if the alternately spliced protein is functionally different to the WT and what these differences, if they are present, may cause in vitro and what role they may play in the cancers correlated with the exon's transcription.

Investigation into the effects of an exon skipping event correlated with SCLS

Discussed in second results chapter will be the exon skipping event in talin 1 seen in sufferers of SCLS. I aim to follow a similar pattern of characterisation and analysis as that of exon 17b. I will attempt to produce and purify the protein, examine its physical characteristics, and then attempt to compare its functionality to that of the WT. If a significant difference is seen between the two, I aim to suggest what this exon skipping event might mean in the pathogenicity of the disease.

Testing the viability of a new, ethical antibody alternative

As well as working on the splice variants and their potential pathogenic effects, in the third results chapter I will discuss my work with one of our collaborators and my project to characterise/evaluate the functionality of their intrabodies. These intrabodies are artificial bacterial proteins that can be evolved to bind an intended target. I aim to support the data they have already gathered as well as confirming the viability of the intrabodies as a possible ethical replacement for standard polyclonal antibodies.

Chapter 2: Materials and methodologies

2.1 Media and buffers

*All buffers are made with ultrapure milliQ water

Table 1: Media and buffers

Table of buffers and media used throughout the project, contents of buffers/medias are shown in right hand column. Majority of components were sourced from Thermo Fisher Scientific, with exceptions: N15H4Cl (Cambridge Isotope Laboratories, Inc.), BME vitamins (Sigma-Aldrich).

Buffer	Contents
Nickel column affinity buffer A	20 mM Tris (hydroxymethyl), 500 mM NaCl, 20 mM imidazole, pH 8
Nickel column elution buffer B	20 mM Tris (hydroxymethyl), 500 mM NaCl, 500 mM imidazole, PH 8
Q column affinity buffer A	20 mM Tris (hydroxymethyl), 50 mM NaCl, pH 8
Q column elution buffer B	20 mM Tris (hydroxymethyl), 1 M NaCl, pH 8
S column affinity buffer A	20 mM $NaH_2PO_4 \bullet 2H_2O$, 50 mM NaCl pH 6.5
S column elution buffer B	20 mM $NaH_2PO_4 \bullet 2H_2O$, 1 M NaCl pH 6.5
GST affinity buffer A	20 mM Tris (hydroxymethyl), 250 mM NaCl, pH 8
GST elution buffer B	20 mM Tris (hydroxymethyl), 250 mM NaCl, 10 mM L-glutathione, pH 8
Gel filtration buffer	20 mM Tris (hydroxymethyl), 150 mM NaCl, 2 mM DTT, pH 8
NMR buffer	50 mM NaCl, 15 mM NaH2PO4, 6 mM Na2HPO4 ,pH 6.5 ,2 mM DTT

10x SDS-PAGE running buffer	50 mM MOPS, 50 mM Tris (hydroxymethyl), 0.1% SDS, 1 mM EDTA, pH 7.7
Stacking gel buffer	0.375 M Tris (hydroxymethyl), pH 8.8
Separating gel buffer	1 M Tris (hydroxymethyl), pH 6.8
Lysogeny broth	10 g/L Tryptone, 10 g/L NaCl, 5 g/L Yeast extract
Minimal media broth solution A	12.5 g/L Na_2HPO_4, 7.5 g/L KH_2PO_4
Minimal media broth solution B (per 24mL)	10 mL H_2O, 10 mL BME vitamins, 2 mL 1 M $MgSO_4$, 1 mL, 0.1 M CaCl2, 1 mL 1 M Ampicillin stock, 4 g glucose, 1 g $N^{15}H_4Cl$,

2.2 Plasmid constructs (SCLS 17b and intrabodies) and primers

Synthetic genes in the pET151/pMX expression vector were purchased from GeneArt (Thermo Fisher Scientific) and Intrabody vectors (XPH238,241,242) were kindly sent from the lab of Dr Olivier Rossier from the Bordeaux Neurocampus.

Table 2: Synthetic gene plasmids:

Table of synthetic plasmids used during the project; table shows the expression vector the synthetic gene was produced in as well as the resistance offered by the vector.

Vector	Construct	Resistance
pET-151	hTLN1-R1R2	Ampicillin
pET-151	hTLN1-R1R2-17b	Ampicillin
pET-151	hTLN1-R12DD	Ampicillin
pET-151	hTLN1-R12DD-SCLS	Ampicillin
pET-151	mVinc-D1	Ampicillin

pET-151	TEV	Ampicillin
pET-24	Xb-1282 pIGc-Xph238	Kanamycin
pET-24	Xb-1285_pIGc-minXph241	Kanamycin
pET-24	Xb-1286_pIGc-minXph242	Kanamycin
pET-151	hTLN1-R11	Ampicillin
pET-151	hTLN1-R11-V2078K	Ampicillin
pET-151	mTLN1-R1R2	Ampicillin
pET-151	mTLN1-R1	Ampicillin
pET-151	mTLN1-R2	Ampicillin
pMX	hTLN1-R1R3-stretch	N/A
pMX	hTLN1-R1R3-17b-stretch	N/A
pGEX-stretch	Empty vector	Ampicillin

Table 3: Primers:

Table showing primers used to check the correct insertion of the R1R3-WT/17b insert into the pGEX-stretch vector.

Primer	Sequence
pGEX_GST_rod_fw	GAAGACTAAGTCGACTCGAGCGGCCG
HALO-R	CGAATGGAAAGCCAGTACCG
pGEX-Tln1-R1R3-17b_R3-FWD	CGGATAGCGAAGAACAGC

2.3 Production of bacterially expressed protein

2.3.1 Transformation of plasmids into *E. coli*

pET-151 plasmids were transformed into aliquots of BL21 competent cells. The cells and 1.5µL of plasmid stock at >100 ng/µL were incubated on ice for 25 minutes. The cells were then heat shocked for 60 seconds at 42°C. After heat shock cells were allowed to recover in 250 µL of fresh LB media at 37°C in a shaking incubator for 45 minutes. 100µL of the transformed proteins were plated onto ampicillin containing agar plates.

2.3.2 Protein expression in *E.coli*

To express protein, colonies of the transformed BL21 *E.coli* cells were picked from agar plates, and transferred to 50 mL falcon tubes containing 10 mL of fresh LB media. The falcons were left overnight at 37°C in a shaking incubator. 7.5 mL of the inoculated media was transferred to 750 mL of pre-autoclaved LB containing 100 µg/ml ampicillin. The flasks were incubated at 37°C in a shaking incubator until optical density (OD) had reached between 0.7-0.8. The flasks were then induced with 750 µL of 1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG). The flasks were left overnight at 18°C in a shaking incubator.

2.3.3 Expression of ^{15}N labelled protein

Both 1D and 2D nuclear magnetic resonance (NMR) experiments were performed throughout the project. NMR experiments detect the energy emitted by the change in 'spin' of atomic nuclei that have spin of ½. Hydrogen (^{1}H) has spin of ½ and is present in all proteins making all proteins viable for 1D NMR. However, for 2D NMR experiments, there must be a second nuclei connected via a single bond to the proton (^{1}H). Therefore, for 2D experiments protein is expressed in a controlled environment in which the sole nitrogen source is isotopically labelled ^{15}N ammonium chloride, which has spin of ½.

Minimal media broth (shown in Table 1) containing the isotopically labelled ^{15}N ammonium chloride is prepared in two steps. Solution A is autoclaved, then solution B is prepared, and filter sterilised into solution A. To inoculate flask, a small amount of a glycerol stock of transformed cells is added to 10 mL of solution A and B and left overnight. 5 mL of this overnight cell culture is used to inoculate 500 mL of minimal media and grown at 37°C in a shaking incubator. Once OD reached 0.7-0.8, 500 µL of 1M IPTG stock was added. Temperature was reduced to 18°C and flasks were left overnight.

2.3.4 Subcloning (digestion ligation)

To examine protein force loading capability, the R1R3-WT and 17b plasmids were subcloned into the specialised pGEX_Stretch vector, which contained HALO and AVI tags for anchoring the protein to the stretching apparatus.

The empty pGEX_stretch and R1R3 WT/17b vectors first underwent restriction digestion, and the R1R3 plasmids (WT/17b) contained cut sites either side of the inserted R1R3 sequence. The insert was cut out using with BamHI and Xho1 restriction enzymes (New England Biolabs (NEB)) using 2 µL of each enzyme in a total volume of 100 µL. The enzyme and plasmids were left for 1 hour at 37°C. Both digestions were then run on a 1% agarose gel. The band containing the insert from the pET151 vector were gel extracted using the QIAquick Gel Extraction Kit from Qiagen following manufactures protocol.

The extracted insert was then combined with the now open pGEX_stretch vector and 1 µL of ligase enzyme (to a total volume of 10 µL) was added and ligated at 4°Cover 48 hours.

2.3.5 Agarose / DNA gels

Agarose gels were made using 1% agarose powder from (Melford) and made up to volume with 0.5% TAE buffer.

2.3.6 SDS-PAGE gels

To perform SDS-PAGE gel electrophoresis, acrylamide gels which contained a 10% acrylamide separating gel and a 4% acrylamide stacking gel, prepared in a 1mm Invitrogen Novex empty cassettes following the Invitrogen protocol.

Table 4: *Components used to produce 1 SDS-PAGE gel, showing components of both the separating and stacking gels required in each gel.*

Component	Separating gel (10% acrylamide)	Stacking Gel (4% acrylamide)
Ultrapure MilliQ water	2.35 mL	1.45 mL
Separating gel buffer	2.35 mL	
Stacking gel buffer		1.05 mL
SDS (10%)	31.25 µL	62.5 µL
Sucrose (50%)	1 mL	
Acrylamide (40%)	1.5 mL	0.375 mL
TEMED	1.56 µL	1.25 µL
Ammonium persulfate catalyst (0.05 g/L)	156.25 µL	250 µL

2.4 Protein extraction and purification

2.4.1 Cell harvesting

Induced cell cultures (having been allowed to grow overnight) were spun down using a centrifuge (Beckman JA 10 rotor) at 4025 G for 10 minutes at 4°C, and the pellet was resuspended using Nickel column affinity buffer A.

2.4.2 Cell lysis

Resuspended pellets were lysed using a MSE Soniprep 150 sonicator set to 55% amplitude for 5 minutes with a 30 second on 30 second off pulse method. The lysed cells were then spun down at 31304 G (Beckman JA 25.50 rotor) for 40 minutes at 4°C. The supernatant removed and filtered through a 0.2 µm pore filter into a 50 mL falcon tube.

2.4.3 His tagged protein purification

Proteins were purified using nickel affinity chromatography, followed by anion exchange as described previously in (Khan *et al.*, 2021).

The plasmids used throughout this project all encoded a 6xHis or 10xHis (intrabodies) histidine tag attached to the protein of interest. This tag was used to purify proteins using Ni^{2+} ion resin containing columns, also known as nickel ion chromatography. The high affinity of the histidine tag for nickel ions tightly binds the protein to the resin, leaving non-binders to be washed out.

For the purification process an AKTA pure (GE healthcare) was used, UV absorbance at 280 nm and 315 nm was used to track the purification process. Sonicated cell supernatant was loaded onto the 5 mL HisTrap HP column. The bound protein was then eluted from the column using an increasing gradient of imidazole in the Ni-NTA/nickel column elution buffer B.

Once protein was eluted, the 280 nm absorbance reading was used to select samples for SDS-PAGE gels. To these samples was added 5x sample running buffer to make a total volume of 20 uL, 5 uL of each was loaded into the acrylamide gel. samples containing protein were then dialysed into appropriate buffer (Q/S column affinity buffer A).

2.4.4 Sample Dialysis

Dialysis into ion chromatography buffers was done using 4-5 L of either Q or S affinity buffer and SnakeSkin™ (Fisher scientific) dialysis tubing. Aliquots of purified Tobacco Etch Virus (TEV) protease were added to cleave the His-tag from the sample, samples dialysed overnight at 4°C.

2.4.5 Ion chromatography (Q and S)

After nickel chromatography and dialysis, ion exchange chromatography was used to clean up the sample. Depending on the isoelectric point (PI) of the protein, which was determined using ExPASy ProtParam(Gasteiger *et al.*, no date). Either a HiTrap Q (for anion exchange) or S (for cation exchange) exchange chromatography column (GE healthcare) was used.

2.4.6 GST chromatography (gravity flow)

The pGEX_stretch vector encoded a TEV-cleavable glutathione S-transferase (GST) tag, which allowed the pGEX_stretch proteins to be purified using the batch method. 1 mL of glutathione agarose beads (fisher scientific) was equilibrated and washed with 10mL (20 column volumes) of GST affinity buffer A. 10 ml of the filtered supernatant was added to equilibrated beads. The beads were washed with another 10mL of GST affinity buffer A to wash out unbound protein. 500 µL of TEV protease was added to a total volume of 750 µL. The TEV protease was allowed to incubate with the beads for 3 hours at room temperature. The contents of the falcon were filtered through a gravity flow column and washed with 500 µL GST affinity buffer A 7 times. With the his-tag cleaved by the TEV protease this step washed out only cleaved pGEX_stretch protein and left contaminants and un-cleaved protein bound to the GST beads. The eluted fractions were collected and run on a SDS page gel.

2.4.7 Measuring protein concentration (Bradford/nanodrop)

Protein purity was checked using an INPLEN NanoPhotometer® N60/N50. The Nanophotometer requires proteins with tryptophan or tyrosine residues that absorb at 280 nm, in the case of proteins like talin R2, which lacked tryptophan or tyrosine, a Bradford assay was used. Bradford reagent was combined with known concentrations of bovine serum albumin (BSA) to create a standard curve of protein concentration to absorbance (at 595nm). The unknown concentration protein sample was added to the same volume of Bradford reagent and its absorbance plotted against the standard curve to give an accurate protein concentration. Once protein purity was calculated, proteins were aliquoted and flash frozen with liquid nitrogen.

2.5 Biochemical characterisation of protein

2.5.1 Bioinformatics (UniProt alignment etc.)

The Alignment of sequences was done throughout the project to analyse similarity between WT and variants of the protein of interest. This process was performed using protein and nucleotide Basic Local Alignment Search Tool (BLAST) (Madden, 2003).

2.5.2 Structural modelling (AlphaFold)

Structural modelling of the target proteins that did not already have a known crystal structure (e.g., R1R2 WT) was done using ColabFold, an open-source software running AlphaFold2 (Jumper *et al.*, 2021)(Mirdita *et al.*, 2022). This generated novel, predictive structures for the proteins of interest. The resulting structures were processed and visualised in the PyMOL Molecular Graphics System (Version 2.0 Schrödinger, LLC).

2.5.3 Circular dichroism

To perform circular dichroism experiments a JASCO J-715 spectropolarimeter was used, as well as a Jasco PTC-423S/15 Temperature Controller. Protein samples were run at 0.5 mg/mL, in Q column affinity buffer A.

Far UV spectra were collected between wavelengths of 200-260 nm, with 6 scans taken at 50nm/min, a 1 nm band width and a 1nm step resolution.

Melting curves data was collected at a fixed wavelength of 222 nm, the temperature was then varied from 20-90°C with a 1 mm band width and 1 nm step resolution.

2.5.4 Size exclusion chromatography

Size exclusion chromatography (SEC) was used to analyse protein-protein interaction. An ÄKTA pure protein purification system (Cytiva) and a Superdex 200 increase 10/300 GL size exclusion column (Sigma Aldrich) was used. The protein samples were run in either S column affinity buffer A or Q column affinity buffer A depending on the PI of the protein in question.

Talin proteins were run at 100 μM, with the binding protein (e.g., Vinculin) at a range of concentrations in a 1:2, 1:3 or 1:4 ratio with talin.

2.6 Structural biochemistry

2.6.1 Nuclear magnetic resonance (NMR)

Many NMR experiments were run throughout this project and to perform them a Bruker Advance III 600 MHz NMR spectrometer was used equipped with a CryoProbe. Two types of experiments were conducted. 1D proton experiments (using 128 scans) were run to establish the presence of the protein in a given sample, and 2D experiments were run to examine binding between two proteins. The 2D experiment samples were isotopically labelled using ^{15}N ammonium chloride. Both proteins in 2D samples (labelled and unlabelled) were buffer exchanged into standard NMR buffer and 5% vol/vol D_2O was added to allow spectrometer locking. The isotopically labelled samples were used for ^{15}N TROSY HSQC experiments. The resulting 2D spectra was processed using Topspin and analysed using CCPN Analysis 3.04 (Mureddu and Vuister, 2019).

Chapter 3 Results

3.1 A novel exon in the TLN1 gene and its role in cancer pathogenicity

3.1.1 The discovery and initial work on exon 17b

As discussed in the introduction, the novel exon 17b lies between exons 17 and 18 and was shown to encode a region within the talin 1 gene. The exon was discovered by fellow researchers at BioMedX in Germany, Nina Grexa, Lina Gallego-Paez, and Jan Mauer. During bioinformatic analysis of cancer sample databases using the DJExpress alternate splicing visualisation method (Gallego-Paez *et al.*, 2022) they uncovered that in COADREAD (Colorectal adenocarcinoma), BRCA (breast cancer) and NSCLC (non-small cell lung cancer), there were 2 splice junctions that were unannotated until this discovery. These junctions, junctions 18 and 20 (fig.9A), were significantly upregulated in these cancer types as opposed to healthy WT cells. The presence of these two splice junctions implied the existence of a new exon between the exons 17 and 18. This novel exon was named 17b. The new exon was found to encode mRNA for a region within the talin rod domains R1R2.

Data from the cancer cell line encyclopaedia (CCLE) showed that the expression of 17b was correlated with the increased dependency of the cancers on certain gene clusters. To examine which gene clusters were utilised by the cancers, the process of gene set enrichment analysis (GSEA) was used. GSEA measures if a predetermined set of genes shows significantly changes in regulation between two cellular phenotypes (cancer and non-cancer). The results (fig.9B)showed gene clusters involved in cytoskeletal organisation, cell adhesion and the ErbB (erythroblastic oncogene B) signalling pathways were all enriched in cells which expressed 17b. ErbB signalling is involved in cell differentiation, proliferation, apoptosis, migration, and cell motility (Yarden and Sliwkowski, 2001). These enriched gene

clusters support the involvement of the 17b variant of talin in cancer, as talin is especially important in adhesion, motility, and the cytoskeletal reorganisation, all of which are properties involved in the cancer-dependant gene clusters.

The survival rate of cancer cells expressing (and not expressing) exon 17b post drug treatment was also characterised by my collaborators, they examined the effects of a broad range of drugs that target different known cancer involved pathways. Shown in fig.9C, cells expressing 17b (in red and dark red) were sensitive to drugs targeting the epidermal growth factor receptor (EGFR) such as Ibrutinib. EGFR helps to regulate the homeostasis and development of epithelial tissue in non-cancerous cells, but has also been shown to be a strong tumorigenic factor in breast and lung cancers (Sigismund, Avanzato and Lanzetti, 2018).

Further analysis of the 17b exon in cancer cells showed that it was correlated with epithelial cancer biomarkers as opposed to mesenchymal cancer biomarkers, this means that 17b-expressing cancer cells were not as likely to transition from epithelial (specialised, generally non-motile) cancer cells into mesenchymal (undifferentiated) cancer cells. Mesenchymal cancers are cells that are able to metastasise and move from their place of origin, and indicates a more pathogenic cancer (Kalluri and Weinberg, 2009).

A possible regulation pathway for exon 17b splicing was also established, fig.9D shows the results of the addition of epidermal growth factor (EGF) (ligand of the EGFR) and transforming growth factor beta (TGF-β) (a common cancer-associated cytokine). The addition of these growth factors resulted in HeLa cells that originally expressed 17b, to switch expression and skip exon 17b. Downstream transcription factors of the EGF and TGF-β pathway, SMAD3 and PCBP1 (in order of downstream presence) were knocked down. The

knockdown resulted in the return of 17b expression. This implies the presence of an EGF/ TGF-β -> SMAD3 -> PCPB1 regulatory pathway of exon 17b.

Some of the data shown in this section was from (Gallego-Paez *et al.*, 2022), a paper which I contributed to, providing the predictive modelling, and biochemical analysis, which will be discussed in the results. I was not involved in the bioinformatics or the DJExpress pipeline work shown in figure 9.

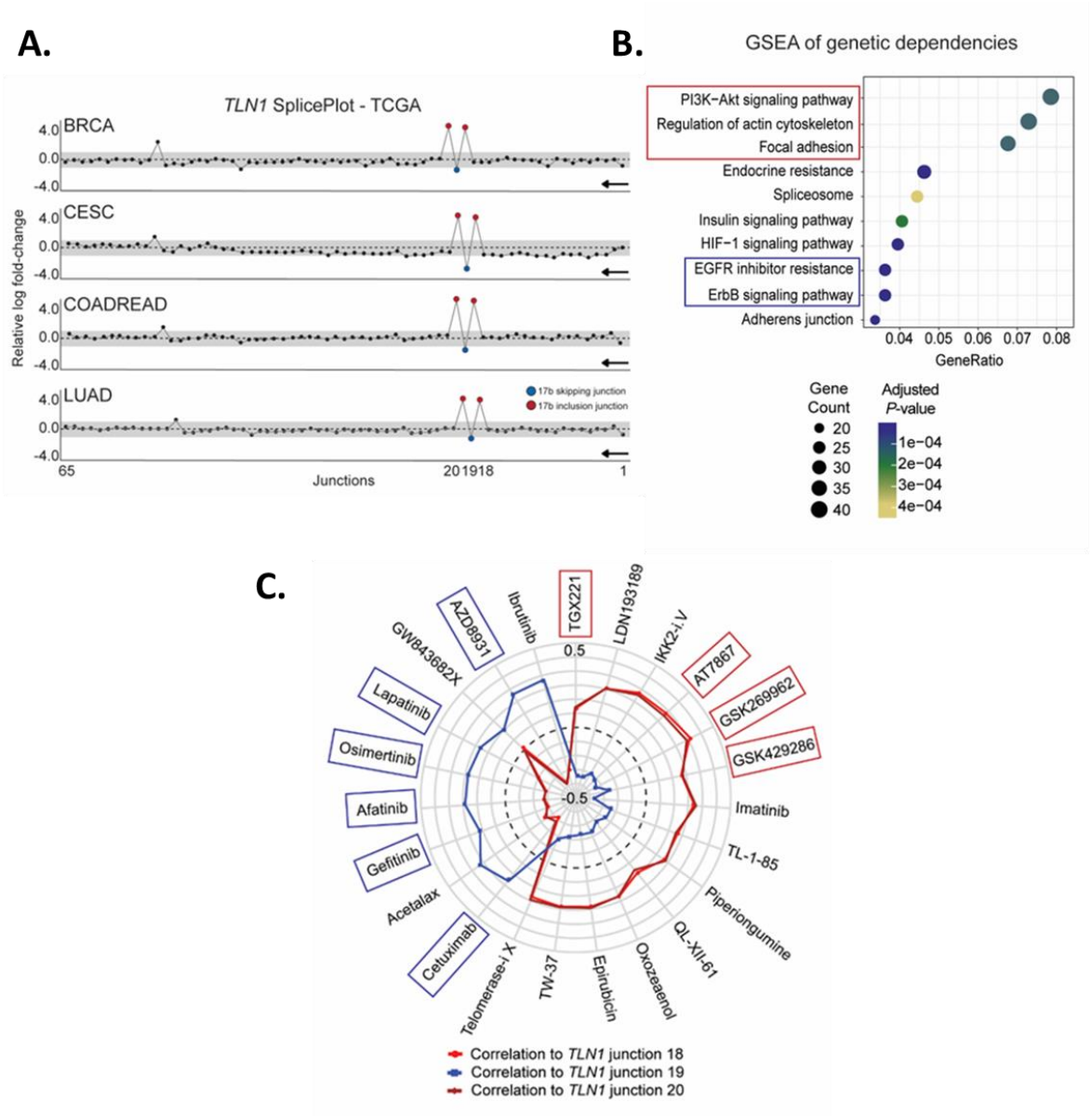

A.
TLN1 SplicePlot - TCGA
BRCA
CESC
COADREAD
LUAD
Relative log fold-change
17b skipping junction
17b inclusion junction
65
Junctions
1
B.
GSEA of genetic dependencies
PI3K−Akt signaling pathway
Regulation of actin cytoskeleton
Focal adhesion
Endocrine resistance
Spliceosome
Insulin signaling pathway
HIF−1 signaling pathway
EGFR inhibitor resistance
ErbB signaling pathway
Adherens junction
GeneRatio
Gene Count
Adjusted P-value
C.
TGX221
LDN193189
IKK2-I-V
AT7867
GSK269962
GSK429286
Imatinib
TL-1-85
Piperlongumine
QL-XII-61
Oxozeaenol
Epirubicin
TW-37
Telomerase-I X
Cetuximab
Acetalax
Gefitinib
Afatinib
Osimertinib
Lapatinib
GW843682X
AZD8931
Ibrutinib
Correlation to TLN1 junction 18
Correlation to TLN1 junction 19
Correlation to TLN1 junction 20
D.
TGF-β/EGF vs ctrl
TGF-β/EGF + shSMAD3 vs ctrl
TGF-β/EGF + shPCBP1 vs ctrl
Junction log fold-change
TLN1 junctions
17b skipping junction
17b inclusion junction

Figure 9: BioMedX data on the novel exon 17b

A). identification via the DJExpress pipeline of the inclusion of the 17b novel exon in cancer Gene-wise SlicePlots (breast, cervical, colorectal and lung cancers (BRCA, CESC, COADREAD, LUAD). All splice junctions in talin 1 protein are represented as circles Shown in red are splice junctions upregulated in the cancers (junctions 18 and 20) and blue, downregulated (junction 17)(>0.5 relative log fold change is considered significant). The two un-annotated junctions 18 and 20 which are not expressed in WT talin indicate the presence of exon 17b and its expression in cancers. B). Gene set enrichment analysis (GSEA) of cancers expressing exon 17b. Gene groups highlighted in red are overexpressed/depended on by 17b expressing cancers, while blue are under expressed. Plot shows gene groups involved in focal adhesions and cytoskeletal regulation as well as PI3K-Akt signalling are upregulated in exon 17b expressing cancers. C). DepMap drug sensitivity plot, showing the ability of cancers expressing WT (blue) and 17b (red) talin to survive exposure to certain drugs are shown. Low values indicated low cell survival post drug exposure, showing that drugs highlighted in blue were effective against 17b expressing cancers. D). Gene-wise SlicePlots showing the effect on 17b expressing HeLa cells of EGF and TGFβ factors (left), the knockdown of SMAD3, a downstream transcription factor of EGF and TGFβ (middle) and knockdown of PCBP1, another downstream transcription factor (right). Addition of EGF and TGFβ result in suppression of 17b, knockdown of transcription factors of EGF and TGFβ cause the return of 17b expression suggesting a regulatory pathway of 17b via EGF and TGFβ.

3.1.2 The R1R2-17b insertion

The cancer data strongly suggested that exon 17b was a cancer biomarker, and next the physical differences of the insert were examined. The 17b exon region was sent for sequencing which revealed it encoded an in-frame 17 amino acid insert occurring within the R2 domain sequence of talin 1. With the exon mostly spanning the flexible linker region between R1 and R2 it was predicted that this might extend the linker, further separating the R1 and R2 helical bundles. This separation was predicted to affect the functionality of the region (reducing the ability of the helical bundle to re-fold) which may play a role in the pathogenicity of cancers that express exon 17b.

3.1.3 The implications of 17b going undetected until now

Exon 17b had gone un-annotated in humans until its recent discovery but had been identified in the primate transcriptome. As this exon has only just been identified in humans it is not present in any databases of cancer genomes, meaning even if many human cancers are actually expressing the 17b talin variant, database sweeps would not have recognised its presence. This means that if exon 17b is found to be a cancer biomarker in many other

cancers, its addition to databases such as ensemble, could be used by other researchers/physicians in their future examinations of cancers.

3.1.4 Predictive modelling of the exon 17b insertion

The first step in analysis of exon 17b was a sequence alignment of R1R2-WT and R1R2-17b. Fig.10 shows that the insertion does not occur as a linker between R1 and R2 but rather the 17 amino acid insertion occurs within the first few residues of R2, between Q665 and D666.

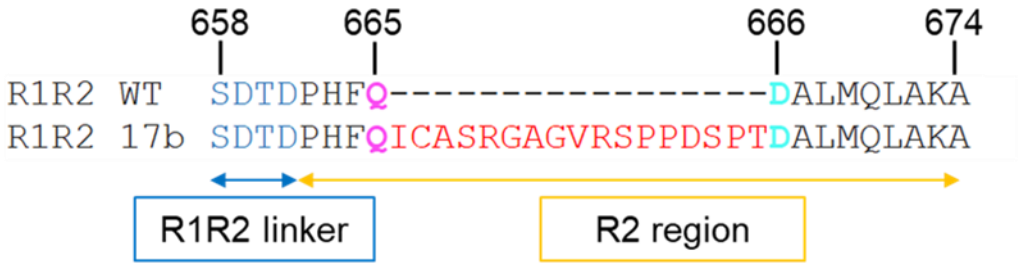


Figure 10: R1R2-WT/17b sequence alignment

Sequence alignment of talin 1 WT (top) and 17b (bottom). The R1R2 linker is shown in blue and the exon 17b insertion is shown in red. The residues the insertion occurs between Q665, D666 are coloured magenta and cyan accordingly.

Next, using the protein structure prediction tool AlphaFold (Mirdita *et al.*, 2022) a model of the R1R2 domains was generated containing the 17 amino acid (AA) insertion caused by the splicing in of exon 17b. The accuracy of the AlphaFold2 generation is shown in figure 11. The predicted Local Distance Difference Test (plDDT) is a measure of how accurately the generated model reflects the environment of the reference structures used in its generation, this score is calculated between all pairs of atoms that fall within specified distances; 0.5 Å, 1 Å, 2 Å and 4 Å (Mariani *et al.*, 2013). A high pLDDT score (70-90) is considered to be accurate, while <70 pLDDT scores are less likely to be accurate (Varadi *et al.*, 2022). Therefore, as shown in fig. 11 A and B the pLDDT scores for the majority of the

R1R2-17b structure are very accurate (>70 pLDDT)(blue) while the 17AA linker region showed considerably lower pLDDT score (30-45)(yellow) so the confidence that this region has been modelled correctly is low. These scores from AlphaFold2 (fig. 11 A) are shown mapped onto the rank 4 model in fig. 11 B.

The Predicted Aligned Error (PAE) score shows the level of positional error at a known residue X, by aligning the reference and predicted structure along Cα, N, C atoms (Varadi *et al.*, 2022), when generating a structure using AlphaFold2 this score is displayed as a heat map. The PAE score for the generated R1R2-17b structure is shown in fig. 11 C, and it can be seen that of the 5 structural variants generated, 1-4 (ranks 1-4) have a similar PAE score, with low error over most of the structure with the exception of the 17 AA inserted linker region, rank 5 showed slightly more predicted error than the other 4.

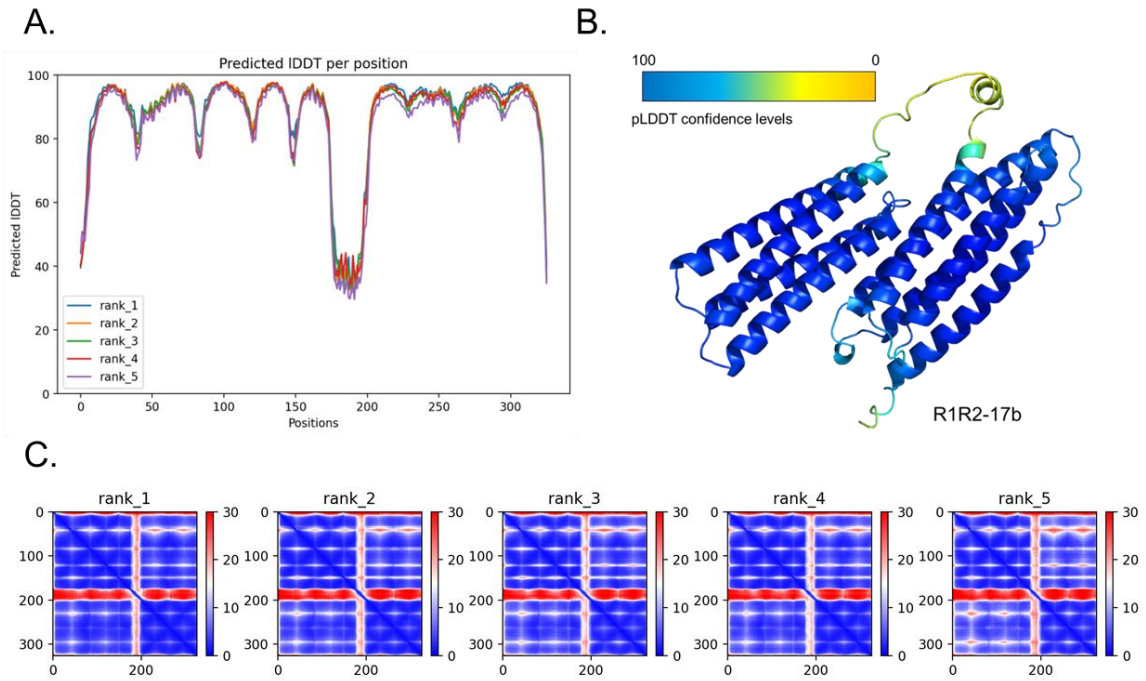


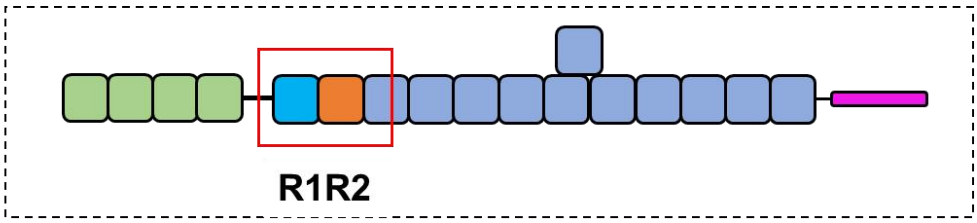
Figure 11: Reliability of AlphaFold2 R1R2-17b prediction

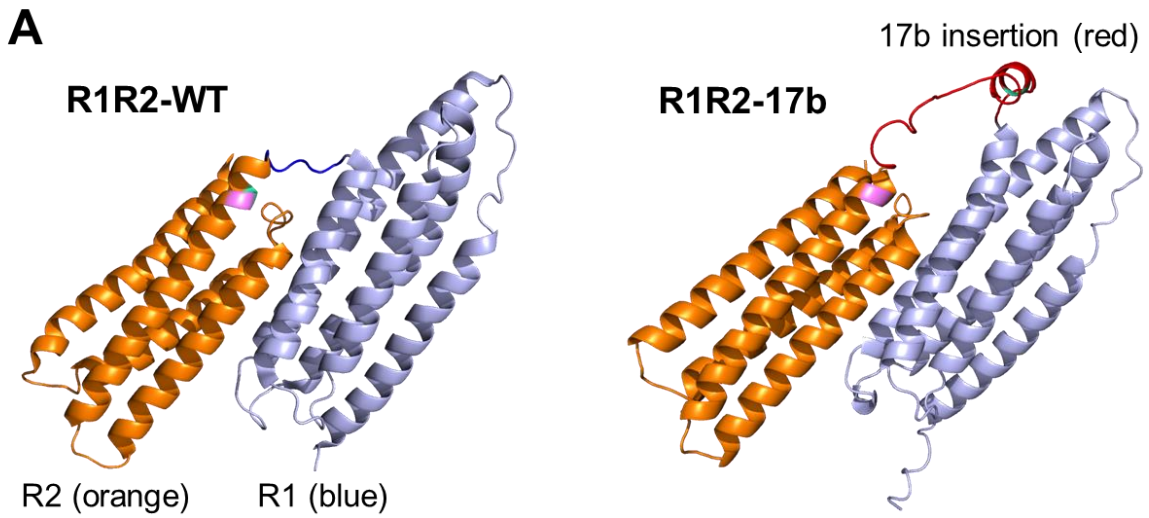
A,C). predicted Local Distance Difference Test (plDDT) data and Predicted Aligned Error (PAE) data provided by AlphaFold2 on the accuracy of the R1R2-17b generated structure. B). The pLDDT scores are shown mapped onto their respective residues on the model for R1R2-17b, with higher pLDDT scoring residues coloured blue and lower scoring residues coloured yellow/orange.

The AlphaFold model and the crystal structure of the R1R2 WT (pdb 1sj8 (Papagrigoriou *et al.*, 2004)) are shown in fig.12A and the 17 amino acids inserted via 17b are highlighted in red, the crystal structure of R1R2 was shown to agree strongly with the regions of the predicted R1R2-17b model unaffected by the insertion of the linker.This suggests the predictive modelling was accurate for the majority of the R1R2-17b model. Seen by comparing the two proteins, the addition of the 17AA insertion produces an extended linker region between the two helical bundles.

Shown in fig.12B, a close-up view of the area surrounding the insertion shows the 17b insertion has formed a long, unstructured linker with a small helical region towards R1, the helical region being made of the few residues of R2 before the insertion between residues

665-666. Also shown in the close up is that the original linker between R1R2 has been partially incorporated into R1, with 3 residues now forming part of the R1's 5th helix (E657, S658, D659). This is likely because the short and therefore, highly taut linker between R1R2-WT, is now relaxed due to the linker extension and it is interesting to see it has added to the R1 helix. This possibly suggests this addition to the R1 helix occurs during the reduced tension in the unfolded form of R1R2-WT. It should of course be considered that these predictions are based on a region of the AlphaFold2 structure that has a low pLDDT score and so further analysis is needed to validate these predictions.





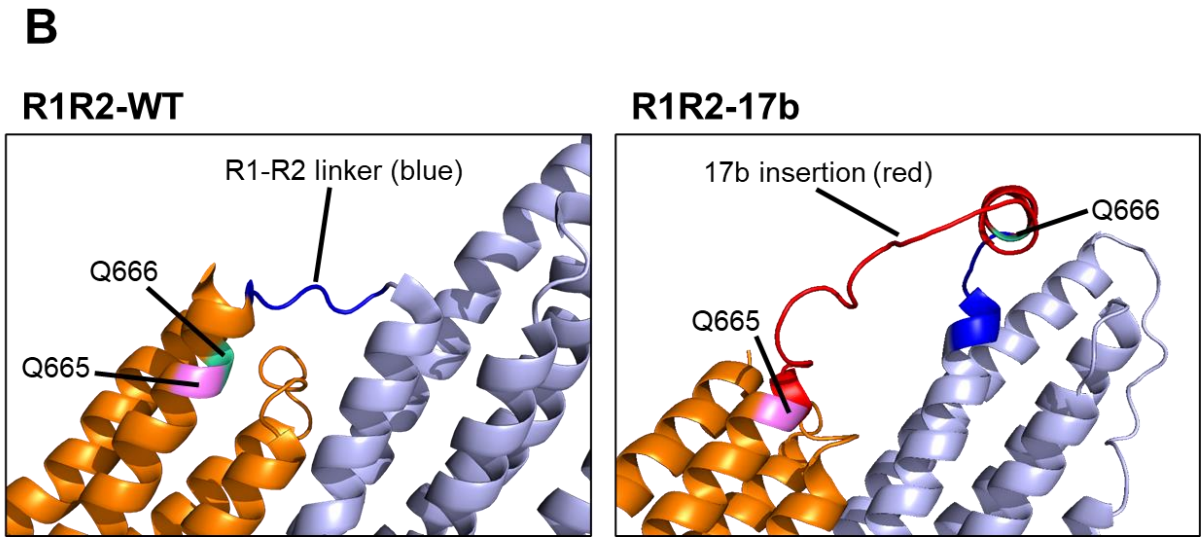


Figure 12: Predictive modelling of talin 1 WT/17b

Predictive modelling results of the effects of exon 17b expression. A). PyMol images showing the crystal structure of R1R2-WT (pdb 1sj8) (left) and AlphaFold 2 generated PyMol structure of R1R2-17b (right). R1 shown in light blue and R2 in orange, the linker, 17b insertion and Q665, D666 are coloured in accordance with sequence alignment in fig.10. Seen from the R1R2-17b model, the 17b insertion occurs within R2 and greatly increases the length of the linker between R1 and the remainder of R2. B). Close up of the PyMol images shown in C. R1R2-WT is shown left and R1R2-17b right.

The initial predictive modelling showed the location of the insertion. The next set of digital analysis done was overlaying the AlphaFold model of R1R2-17b with the WT crystal structure (fig.13A), which revealed that there were significant changes to the R1 and R2 helices due to the larger linker. As mentioned earlier and shown in detail in (fig.13B) the now slackened linker between R1R2 has allowed for the 5th helix of R1 to extend, incorporating residues of the R1R2-WT linker into the helix (residues G656-S658). The linker extension has also affected the 1st helix of R2 at the c terminal end. As the 17 amino acid insertion occurs within this helix, and the extension is unstructured, the 1st R2 helix has been shortened by 4 residues (D661-F665)(fig.13C). These changes to the structure of the helices were predicted to alter the stability of the helical bundle they make up, as a shorter helix would have less area to make the standard interactions within the bundle.

Alterations to the functionality of the R2 helical bundle were predicted as shown in (fig.13C). The 17b insertion occurs within the 1st R2 helix, which is also one of 2 VBS within R2. The separation/shortening of the VBS helix was predicted to alter its binding capability as the VBS may now lack crucial residues required for vinculin binding. The shortened helices within the bundles were also predicted to result in destabilised R2 domain which would unfold with less required force, increasing exposure of the VBS, which would alter the functionality of the R2 VBS.

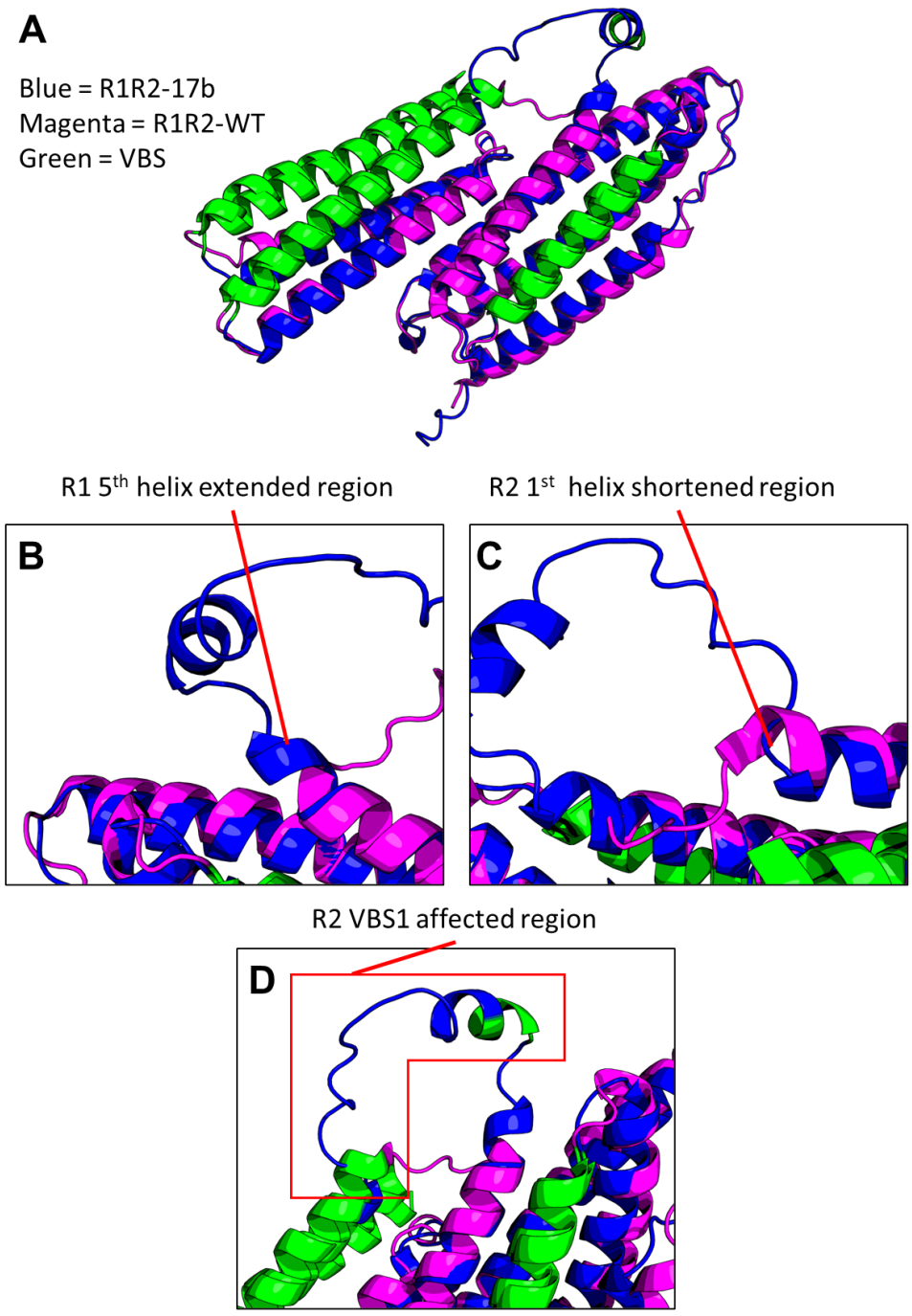


Figure 13: Alterations to talin 1 R1R2 helices due to exon 17b

PyMol models overlayed to highlight the effect of the 17b insertion on the length of helices within R1R2. A). R1R2-WT (crystal structure 6R9T) is shown in magenta and R1R2-17b (AlphaFold 2 generation) is shown in blue, with VBS located within both in green. The two models are overlayed. B). close-up of the area affected by the 17b insertion. The 5th helix of R2 (left) in which the insertion occurs is seen to be shortened (when compared to WT). C). close-up of the 1st helix of R1 which is shown to be extended (compared to the WT) as a portion of the linker region is incorporated into the helix. D). close-up of the first VBS located in R2 (green), which the 17 AA insertion (caused by exon 17b splicing) occurs within, a portion of the VBS can be seen incorporated into the linker.

3.1.5 17b Biophysical characterisation

Having analysed the structural differences caused by the 17b insertion and made predictions about the effects of it, I next aimed to characterise the biochemical and biophysical properties of the 17b variant of R1R2 and to compare it to that of the R1R2-WT. I also aimed to test my hypotheses about the insertion's effect on R1R2 stability and functionality, with the insertion either resulting in a reduction in VBS activity due to shortened binding region, or a possible increase in VBS activity due to overall reduced bundle stability.

To allow for biophysical characterisation of the R1R2-17b protein, the R1R2-WT and 17b proteins were transformed and expressed following standard protocols. Both WT and 17b proteins were expressed and purified well, resulting in high concentrations of very pure talin protein.. The first experiment conducted with the produced R1R2 proteins was circular dichroism, which uses the absorbance of circularly polarised light to detect changes in secondary protein structure. This was chosen as it would show if the protein had been produced correctly, by confirming if the proteins had the correct alpha helical structure. Secondly it would provide stability data as a melting curve for the proteins, which shows at what temperature the structure of the protein becomes unfolded. A less stable protein will unfold at lower temperatures. (The unfolding is measured by the reduction in absorbance of circularly polarised light at 222nm wavelength)

The circular dichroism far UV experiments were conducted at 20°C, the data, shown in fig.14A shows that both R1R2 proteins were alpha helical in nature, indicated by both proteins displaying the characteristic 'double dip', as well as a similar minimum absorbance (minima) of -120 mdeg (R1R2-17b, fig.14A red) and -140 mdeg (R1R2-WT, fig.14A blue) in

the far UV spectra. The far UV absorbance reading at 90°C was also consistent between both proteins with a CD at -40mdeg suggesting both proteins had fully unfolded at 90°C.

The melting curve data for both proteins (fig.14B) were shown to have a stark difference between R1R2 WT and 17b. The melting curve spectra for R1R2 WT (fig.14B blue) showed a singular distinct unfolding event and melting temperature (TM) of 77°C (characterised by a sharp reduction in absorbance, shown visually as a rising curve in CD over a short increase of temperature). In comparison, R1R2 17b (fig.14B red) also had a singular unfolding event, however it occurred at a lower TM of 64°C. The difference between the two suggests that the 17b variant is significantly less stable than that of the WT, with its helical bundles unfolding at a much lower energy level. This aligned with the predictions from the AlphaFold structure as any insertion that disrupts the helices has the potential to reduce the stability of the helical bundles.

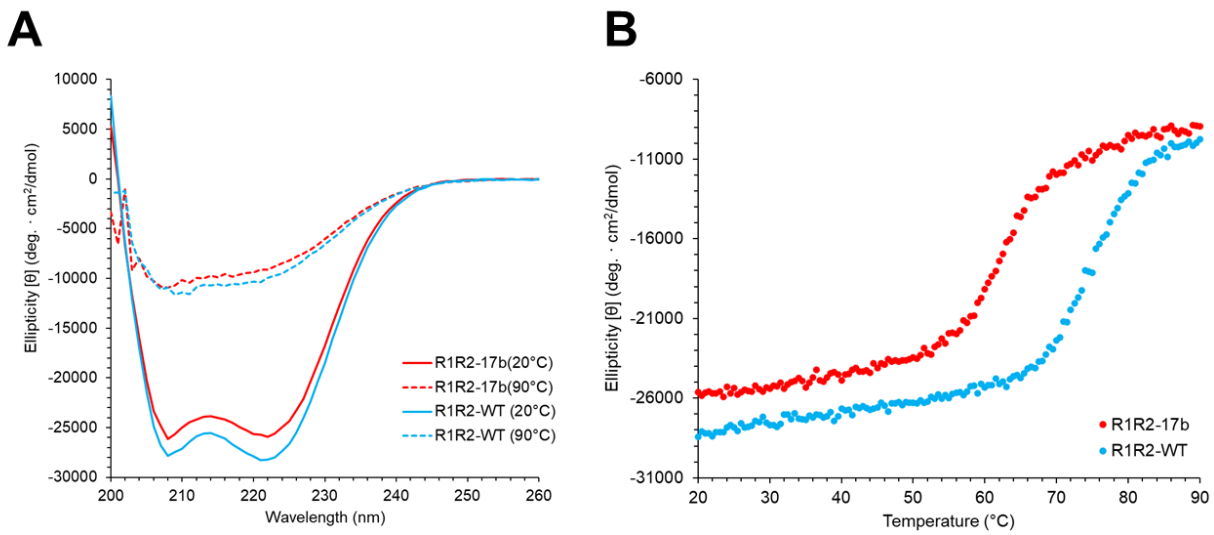


Figure 14: Circular dichroism comparison of R1R2-WT/17b

Comparison of circular dichroism data of R1R2-WT and R1R2-17b, all samples were run in Q column affinity buffer A (20 mM Tris (hydroxymethyl), 50 mM NaCl, pH 8) and at a protein concentration of 0.5 mg/mL. A). Far-UV spectra data (between 200-260 nm wavelengths) showing R1R2-WT at 20°C (blue line) and 90°C (blue dashed line). This data was compared to R1R2-17b at 20°C (red line) and at 90°C (red dashed line). The 'double dip' seen between 205-225 nm indicates α-helical structured protein. Both proteins are α-helical in nature and are fully unfolded (ellipticity minima of ~-10000) at 90°C. B). Melting curve data ran at 220 nm wavelength and between 20-90°C, comparing the unfolding of R1R2-WT (blue) and R1R2-17b (red). A steep increase in circular dichroism (CD, ellipticity) across a short temperature range indicates an unfolding event within the protein. Both proteins are shown again to be fully unfolded by 90°C, with R1R2-WT showing a single unfolding step at 77°C and R1R2-17b displaying a singular unfolding event at a lower temperature (64°C), this suggests lower stability of the R1R2-17b protein.

The CD confirmed that the secondary structure of both proteins was alpha-helical as expected, as well as confirming the proteins had folded correctly after expression. The next biophysical analysis was done using size exclusion chromatography (SEC), which uses resin beads to filters proteins based on size. The larger the protein the earlier it will elute on the column and produce a peak on the elution spectra. To examine protein-protein interaction between R1R2 and its ligands SEC would be used, for those binding experiments required a baseline to know where the proteins would elute when run on their own. This experiment

was also useful as if one protein was incorrectly folded it would elute earlier than it was expected to.

To establish the baseline for SEC, both R1R2-WT and R1R2-17b protein samples were run at a 100 μM concentration (100 μM was used as all future talin samples would be run at this concentration during later biochemical analysis). Initially both proteins ran with a dimer peak, indicated by a peak eluting earlier than the monomer. The dimer peaks seen are an artefact of the synthetic genes resulting exposed cysteines on the outward protein surfaces that are able to form disulphide bonds, the resulting dimer is shown in fig.15 orange elution profile, the dimer is indicated by the first larger peak at 13.5 mL. To eliminate this artefact, the reducing agent dithiothreitol (DTT) was added to the buffer used for SEC at a 2 mM final concentration. DTT reduced the dimer peak to a non-interfering level, shown in fig.15 by the blue elution profile, note the loss of the larger dimer peak at 13.5 mL. A small peak at 13mL elution volume (fig.15 Blue) is also visible, however, this is likely contamination and not the greatly reduced R1R2-WT dimer peak. Interestingly the dimer peak of R1R2 17b was much smaller to start with than the WT and was unaffected by the addition DTT, suggesting the structure changes due to the 17b insertion resulted in exposed cystines in R1R2 WT being less accessible for interaction.

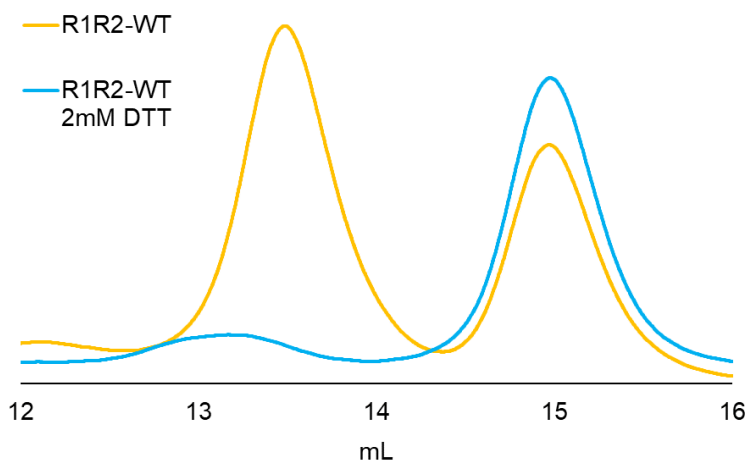


Figure 15: The effect of DTT on reducing R1R2-WT dimerisation

Showing size exclusion chromatography data, both samples run at 100 µM concentration in gel filtration buffer. R1R2-WT is shown in orange (with two peaks), and R1R2-WT containing 2 mM DTT is shown in blue (single peak).

Fig.16B shows the combined elution profiles of both R1R2-WT and 17b ran at 100 uM in the presence of 2mM DTT, the result is a single peak for each (R1R2 WT in blue and 17b in red) with R1R2 17b eluting at 14.75 mL and the WT eluting later at 15 mL. This elution profile was the expected result as R1R2 17b is larger so it should elute earlier.

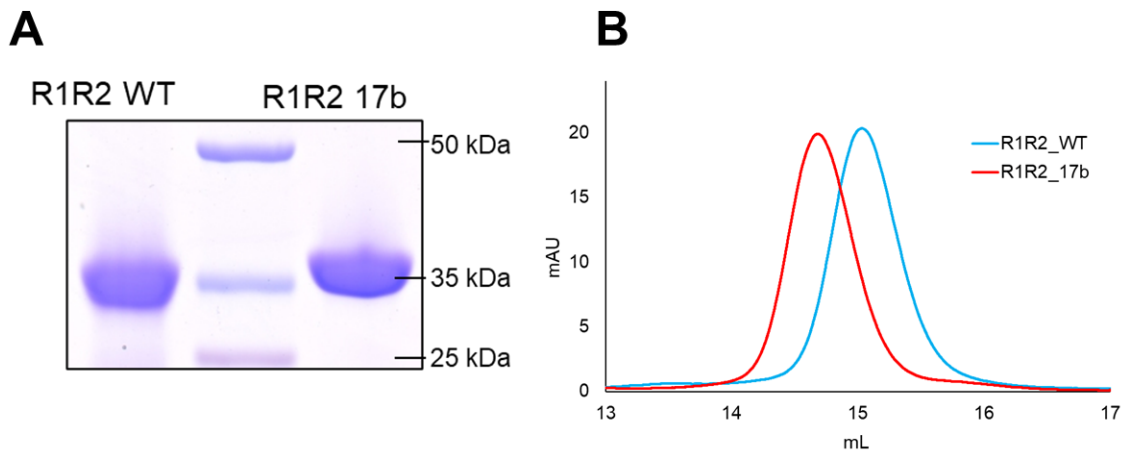


Figure 16: Biophysical characterisation of R1R2-17b

Biophysical characterisation of R1R2-17b in comparison to R1R2-17b. A). SDS-PAGE gel letterbox showing 100 µM R1R2-WT and R1R2-17b, left of the ladder is R1R2-WT at 31,207 Da, while R1R2-17b, right of the ladder, is 33185 Da in weight. B). Initial gel filtration data for R1R2-WT/17b. samples were both run at 100 µM concentration in Q column affinity buffer A (20 mM Tris (hydroxymethyl), 50 mM NaCl, pH 8) + 3 mM DTT. R1R2-WT (blue) and R1R2-17b (red) both show a single monomer peak, R1R2-17b elutes earlier as it is larger and travels faster through the column resin.

The final biophysical characterisation method was to use NMR to acquire and compare 2D ^{15}N TROSY HSQC spectra of the R1R2-WT/17b, the spectra will also serve as a baseline for further biochemical analysis. The data in fig.17 shows the 2D ^{15}N TROSY HSQC spectra of R1R2 17b (red) and R1R2 WT (blue) overlayed, both were run at a 100 uM concentration. The spectra of both R1R2-WT and R1R2-17b are similar but with many peaks being lost or added in due to the additional 17 AA in the 17b mutant. Overall the majority of peaks present in R1R2-WT can be seen in the spectra for R1R2-17b, showing that there is structural similarity between the WT and variant protein.

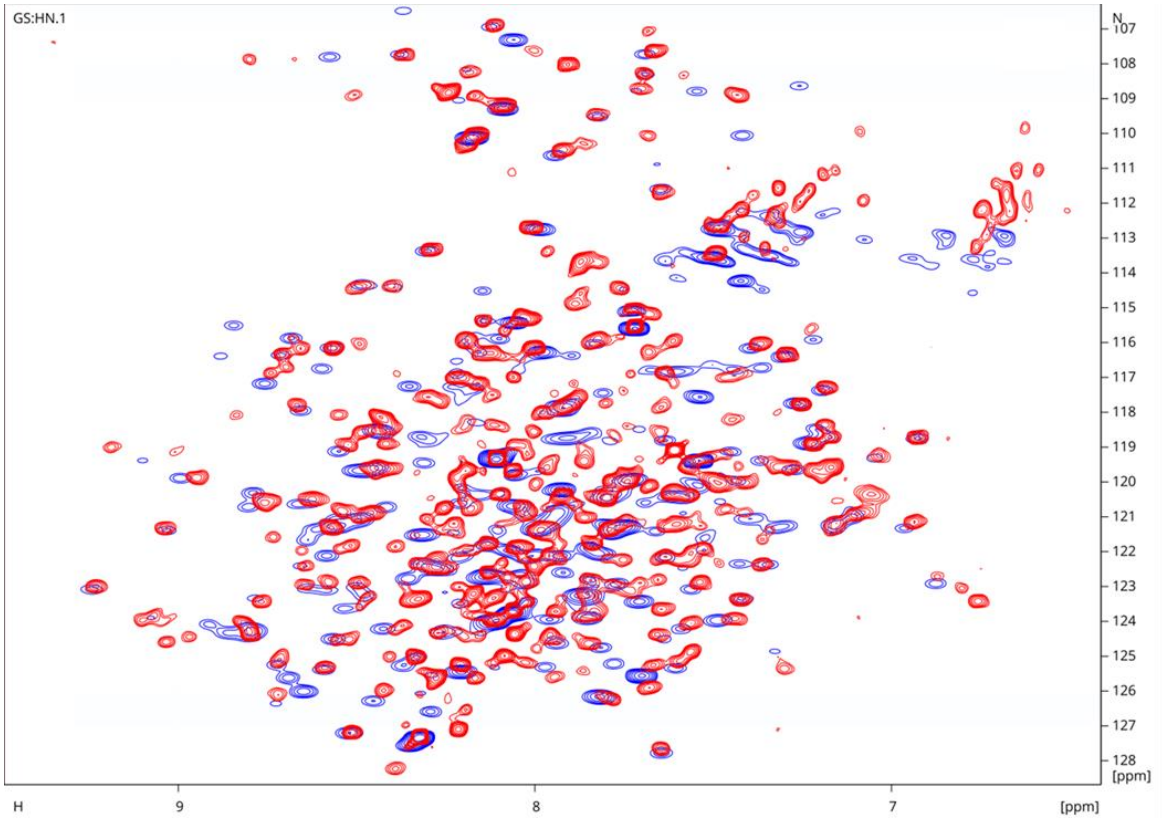


Figure 17: R1R2-WT/17b NMR spectra overlay

Nuclear Magnetic Resonance (NMR) spectra comparison between R1R2-WT(blue) and R1R2-17b (red). Samples were run at 100 µM concentration in S column affinity buffer A with 1 mM DTT and 5% V/V D_2O to allow locking-on of the sample. A ^{15}N TROSY HSQC experiment was run with each sample (at 300K), the resulting spectra are overlayed, significant differences can be seen between WT and 17b spectra due to chemical shift caused by the presence of the 17 AA insertion. However, many peaks are seen to overlap indicating the majority of the R1R2-17b protein is identical to the WT.

3.1.7 17b Biochemical characterisation/functionality

Having analysed the biophysical properties of both R1R2-WT and 17b variant, the data shows that there are potentially physiologically relevant differences between the R1R2-17b and the WT. Therefore, the next step was to investigate my predictions on the altered functionality R1R2-17b, based on the previous experimental data.

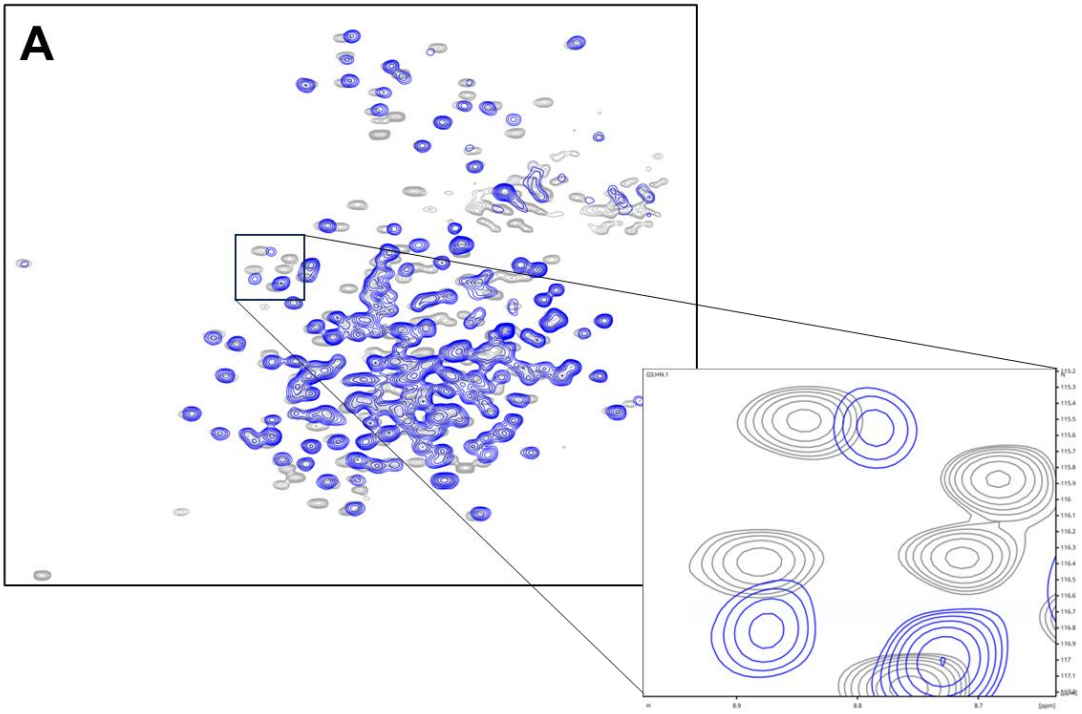
R1R2, like all regions of talin, has ligand binding sites, some of which are exposed on the outside facing regions of the helical bundle such as RIAM, while others such as the VBS are cryptic and cannot be accessed while the helical bundle is folded. There are 11 vinculin

binding sites throughout talin 1, 3 of which are in R1R2 (one in R1 and 2 in R2, as shown in fig.3). The 17 amino acids inserted by the splicing in of exon 17b occur within the second of the three vinculin binding sites and shortens the helix it is found within, so it would be predicted that the insertion might alter or prevent vinculin binding. On the other hand, due to the previously shown stability differences between 17b and WT, it is also possible that the lower stability of the R1R2-17b variant cause increased exposure of the 2 binding sites within R2. To validate these hypotheses, a series of biochemical experiments were performed.

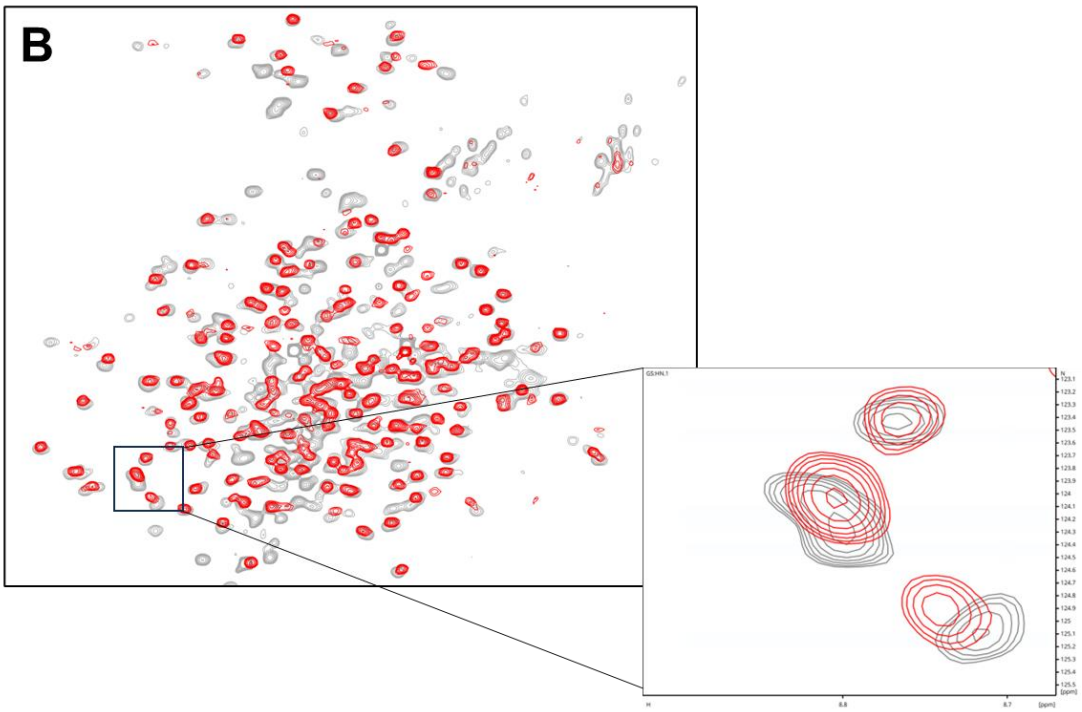
3.1.8 The RIAM binding site in R1R2 remains functional

The first experiment was to confirm if rap1-GTP-interacting adaptor molecule (RIAM) (RIAM peptide residues 6-30), a well-established binder of talin 1 was still able to bind the altered R2 within R1R2-17b (RIAM binds via the unfolded state of R2 (Goult *et al.*, 2013b)). Although the CD suggested the protein is folded, differences within the structure of the helices (due to the 17b insertion) could alter the secondary structure and perturb the binding sites. RIAM was chosen as its binding is mutually exclusive to that of vinculin (Goult *et al.*, 2013a) and was a good comparison for the functionality of internal/external R2 binding sites. To examine RIAM binding, NMR was used, utilising the spectra of R1R2-17b and WT gathered earlier as a baseline. Both R1R2-WT and 17b were run in 2D 15N TROSY HSQC experiment in the presence of 300 µM RIAM peptide. Fig.18A shows the binding of RIAM to R1R2-WT, with 100 µM R1R2-WT alone (grey) and R1R2-WT + 300µM RIAM (blue). Significant chemical shift can be seen between the two spectra indicating interaction between talin and RIAM, as expected. Fig.18B shows 100 µM R1R2-17b (grey) and R1R-17b in the presence of 300 µM RIAM (red). Comparison of the two spectra shows there is chemical shift perturbations, which confirms interaction between RIAM and R1R2-17b, and that the RIAM binding site

within R2 is functional. The NMR data also supports the CD data that the unaffected portion of the R2 helical bundle within R1R2-17b is folded correctly despite the 17 amino acid insertion.



R1R2-WT + RIAM



R1R2-17b + RIAM

Figure 18: R1R2-17b RIAM binding

NMR data showing interaction of R1R2-WT and 17b with RIAM, a known binder of R1R2-WT. Both samples contained 100 µM talin protein and 300uM RIAM peptide in S column affinity buffer A with 1 mM DTT and 5% V/V D_2O. For both R1R2-WT and 17b, experiments conducted were ^{15}N labelled TROSY HSQC (2D) runs. A). Overlay of spectra gathered on the interaction between R1R2-WT and RIAM, considerable shift is seen between R1R2-WT alone (grey) and R1R2-WT +300 µM RIAM (blue), indicating interaction and binding of RIAM as expected. B). Overlay of spectra of R1R2-17b alone (grey) and R1R2-17b +300 µM RIAM peptide (red), chemical shift was seen to occur indicating RIAM also bound R1R2-17b.

3.1.9 R1R2-17b retains vinculin binding capability

Having shown that the RIAM binding site was intact and functional, and therefore showing the outward facing structure was not altered significantly, I then examined the functionality of the vinculin binding sites present on the unfolded, internal faces of the protein.

To examine vinculin binding, size exclusion chromatography was used, as full-length vinculin is a large protein (124KDa) I used only vinculin domain 1 (VD1), as this is the talin binding region of the protein. Early data shows a vinculin dimer peak, but not in the presence of 2 mM DTT. The first set of SEC data was test if the VBS in R1R2-17b was functional, the second experiment was to compare the VBS of R1R2-17b to WT. A 3:1 ratio of VD1 to talin was used when comparing R1R2-17 to the WT to ensure that all 3 binding sites on R1R2 WT/17b could be satisfied.

For the first set of experiments into the vinculin binding capability of R1R2-17b, R1R2-17b (fig.19 dark grey dashed line) and VD1 (fig.19 light grey dotted line) were run by themselves at 100 µM. With a baseline established, a 1:1 ratio of vinculin to talin was run (fig.19 green line). The presence of vinculin immediately caused new set of peaks to be observed (all eluting at an earlier position indicating higher MW), as well a stark drop in the intensity of the free vinculin peak and complete disappearance of the free R1R2-17b peak. The new peaks suggest the formation of complexes between the two proteins. To further confirm that the new peaks seen were in fact complexes, an increasing concentration of VD1 was

titrated with 100 µM talin, a 1:2 (fig.19 blue line) and 1:4 (fig.19 red line). The close up of the complex region in fig.19, bottom panel, reveals the suspected complex peaks change in height in response to VD1 addition with the second larger complex peak at 10.7mL increasing with intensity as the concentration of VD1 increases. To confirm that these peaks are complexes between the R1R2-17b and VD1, an SDS gel containing samples from the compound peak elutions was run, and confirmed the presence of both R1R2-17b and VD1 in all these peaks. This data confirms that the VBS within R1R2-17b are functional.

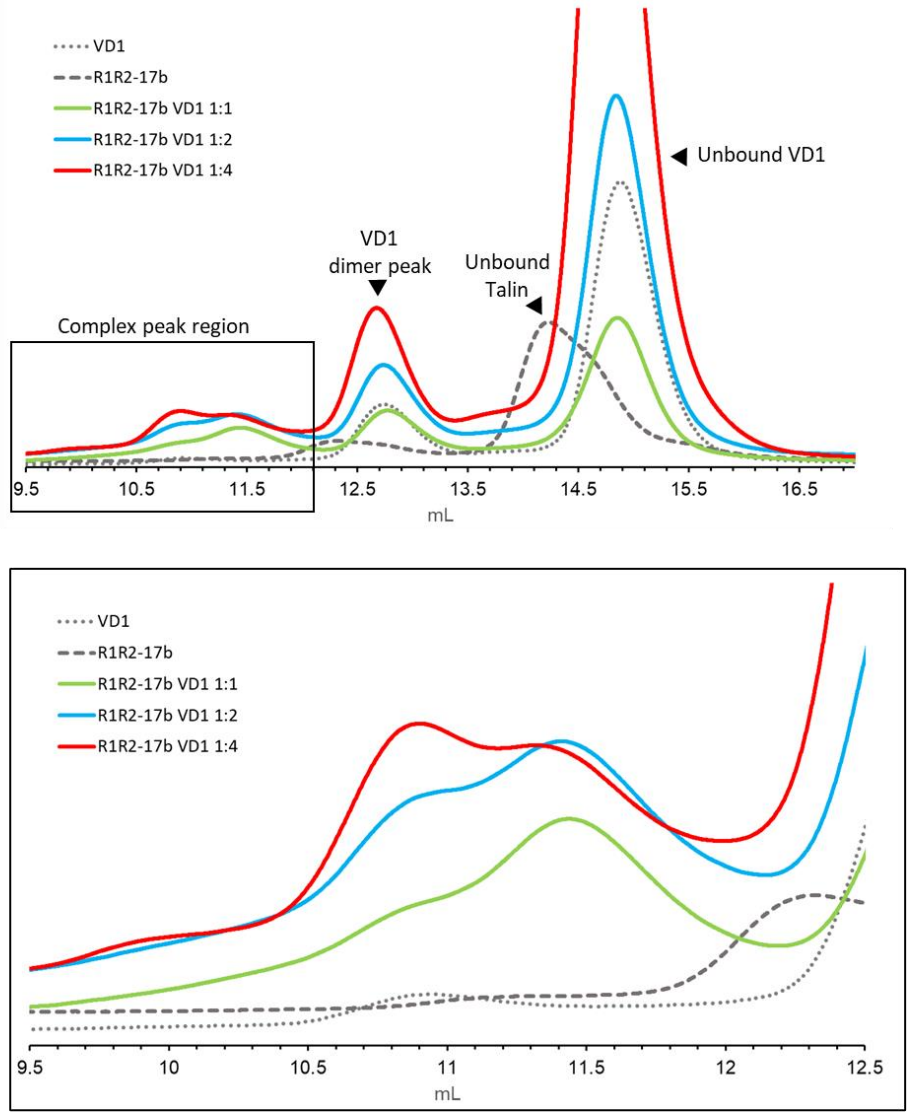


Figure 19: R1R2-17b initial vinculin binding gel filtration data

Gel filtration data showing interaction between R1R2-17b and VD1. Gel filtration experiments were conducted using samples containing 100 μM talin 1 R1R2-17b and increasing concentrations of VD1, samples were run in Gel filtration buffer: 20 mM Tris (hydroxymethyl), 150 mM NaCl, 2 mM DTT, pH 8. Top image shows gel filtration elution spectra for a variety of samples. 100 μM VD1 alone (grey dotted line), 100 μM R1R2-17b alone (grey dashed line), both showed a monomer and dimer peak. Once baseline spectra had been established, combinations of talin and VD1 were ran. R1R2-17b + 100 μM VD1 (green), R1R2-17b + 200 μM VD1 (blue) and R1R2-17b + 400 μM VD1 (red). All combination samples produced complex peaks. A close up of this complex peak region is shown in lower image.

3.1.10 R1R2-17b shows increased affinity for vinculin

Having shown that R1R2-17b retains functional vinculin binding sites. I next compared the vinculin binding ability of R1R2-17b to R1R2-WT. The hypothesis, based on the CD data, was that the reduced stability of the R1R2-17b resulted in helical bundles that were more susceptible to unfolding and increase the availability of R1R2 VBS. The comparison data is shown in fig.20A-B. 300 μM VD1 in grey, produces a single peak at 15.2 mL. 100μM of Talin R1R2-WT was then incubated with 300 μM VD1 at room temperature for 1 hour, then run on the SEC column. The elution profile (fig.20, A blue line) showed no significant decrease in the intensity of the free VD1 peak, and a collection of possible complex peaks are visible between 10-12 mL. The close-up of the WT complex peak region is shown right of fig.20C, as the peaks seen are larger than the VD1 dimer or talin dimer peak, they are assumed to be complexes between R1R2-WT and VD1, a 1:2 and 1:3 complex respectively (complexes between VD1 and the 2 VBS in R2, and a complex between VD1 and the 3 VBSs in R1R2). To confirm this, a gel was again run containing samples from each supposed complex peak elution. Due to the dilution caused by SEC, the stoichiometries of the complexes was difficult to confirm, however, as R1R2 contains 3 VBS and would increase in mass with the addition of each VD1 ligand it is logical to assume that the 1:1, 1:2 and 1:3 stoichiometries are correct. An increase in temperature was expected to increase complex formation and

peak intensity. A 1:3 ratio of R1R2-WT and VD1 was incubated at 37°C for an hour. When this sample was run through the SEC column (fig.20A blue dashed line), the 1:2 and 1:3 complexes had increased in intensity although the overall amount of free VD1 does not decrease suggesting the VD1 binding sites in R1R2 WT remained mostly cryptic even at 37°C.

In comparison to the WT, when 100 µM R1R2-17b was incubated with 300 µM VD1 at room temperature for 1 hour then ran on the SEC column (fig.20B red line), 3 complex peaks were visible in roughly the same elution positions as those seen from the WT, with peaks at 10.8, 11.8 and 13.2 mL (factoring in the size difference between R1R2 17b and WT). However, all 3 peaks were observed with a much higher intensity than that of R1R2 WT, as well as a visible drop in the free VD1 peak intensity. To confirm these were talin vinculin complex peaks, the 1:3 R1R2-17b to VD1 ratio was incubated at 37°C. When the 37°C incubated sample was ran on the SEC (fig.20B red dashed line), the complex peaks were significantly more intense with the 1:2 complex now exceeding the free VD1 peak in intensity.

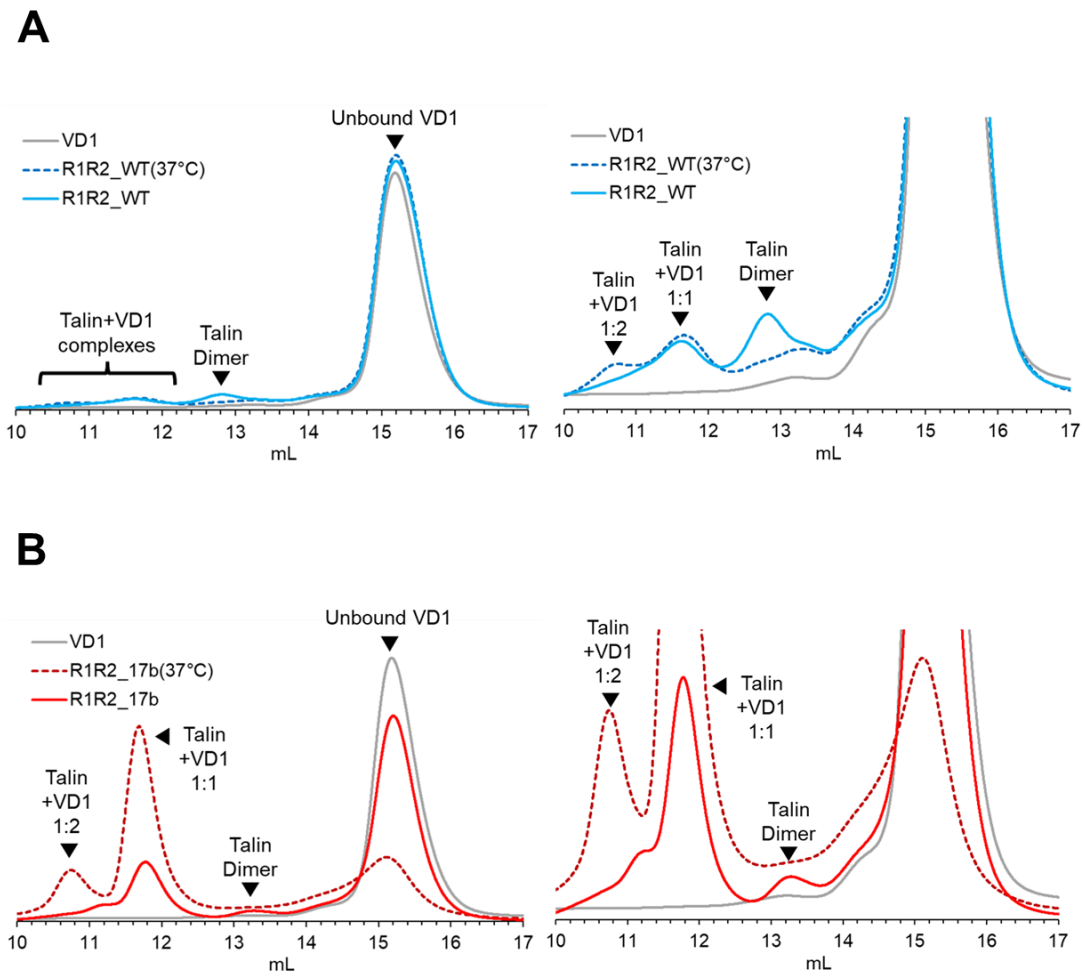


Figure 20: R1R2-17b vinculin binding functionality

Gel filtration data comparing the vinculin binding ability of R1R2-WT and R1R2-17b. Samples were run in Gel filtration buffer: 20 mM Tris (hydroxymethyl), 150 mM NaCl, 2 mM DTT, pH 8, and contained 100 µM talin 1 R1R2-WT/17b, titrated with 300 µM freshly to satisfy all 3 VBS within R1R2.A). Left. overlayed gel filtration elution profiles of R1R2-WT interaction with VD1. To establish a baseline VD1 was run at 300 µM alone (grey)(this was the baseline for both WT and 17b), showing a single monomer peak. A 1:3 titration of talin R1R2-WT to VD1 (room temp incubation) was run (light blue), the same 1:3 titration was run again at 37°C incubation (dark blue dashed). Both showed complex peaks, close-up of complex peak region shown (right). B). (left) overlayed elution profiles of R1R2-17b titrated with VD1. VD1 alone (grey), 1:3 titration R1R2-17b to VD1 (room temp)(red), and 1:3 R1R2-17b to VD1 (37°C) (dark red dashed). Significantly more intense complex peaks were seen, close-up of complex region shown (right).

This dataset confirms these are talin/VD1 complexes and further confirms that the VD1 binding sites on R1R2 17b are functional. The data also reveals that the VBS within R1R2-17b are more accessible than R1R2-WT, both at room temperature and at 37°C. This confirms R1R2-17b is unfolding more easily than the WT at 37°C, agreeing with the results of the CD data and confirming my hypothesis that the 17b insertion reduces stability of the helical bundle.

3.1.11 Production of 'stretchable' R1R2-WT/17b

While the circular dichroism data was conclusive in showing that R1R2-17b is less stable than the WT, it did not confirm if one or both helical bundles had reduced stability as both R1 and R2 were shown to unfold in a single step. Furthermore, the melting temperature of the proteins may show changes in stability (in regard to structural integrity) but it does not show the effect of the 17b insertion on the mechanical stability/force loading of the helices. As one of talins key structural features is the ability for the helical bundles to unfold in response to force, I wanted to examine if the load bearing capability of the proteins had been altered in R1R2-17b.

To do this I partnered with a group in Singapore who have pioneered a technique that allows 'stretching' of a single molecule using two specialised protein tags on either side of the talin protein. A C-terminal AVI tag and an N-terminal Halo tag allow the protein to be anchored at either end. These experiments were conducted using the R1-R3 domains as previous talin stretching experiments had been performed on mouse R1R3, providing a useful baseline. To perform the force-loading experiment, a specialised coverslip is anchored to one end of the R1R3 region (either WT or 17b) while the other end is anchored to a paramagnetic bead (fig.21). Once tethered between the two, a set of 2 permanent

magnets are used to apply force to the magnetic bead (shown in fig.21 black arrows). As the protein deforms and unfolds due to the mechanical force, the distance/gap between the bead and the 2 magnets decreases, the diffraction pattern of light off the bead will alter, and the differences in light diffraction can be measured to allow nanometre resolution on changes in the tethered protein (Yan *et al.*, 2015).

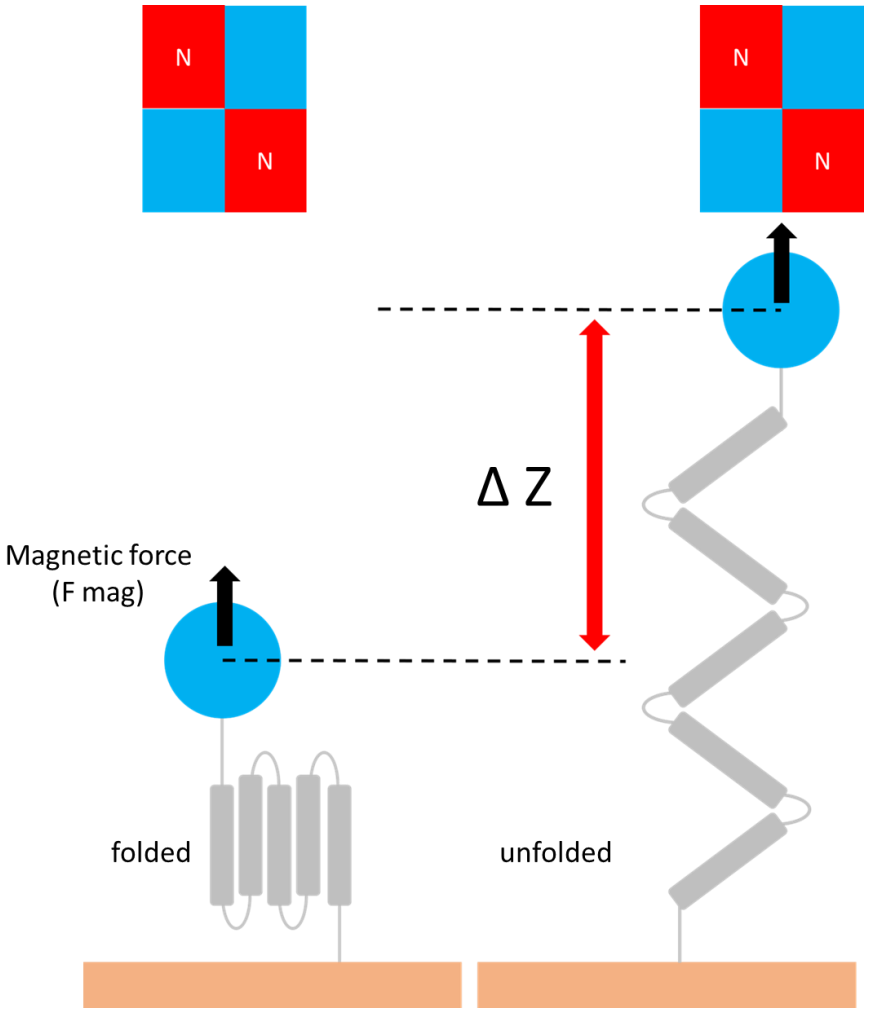


Figure 21: pGEX-Stretch experimental method

Cartoon depicting the process used to detect force loading capability of talin regions. The desired talin protein (grey) is subcloned into the pGEX-stretch vector, this contains an AVI and HALO tag, these two tags, at either end of the protein, are used to anchor the talin region. One end is anchored to a glass slide (Beige), while the other end is tethered to a paramagnetic bead. Once tethering is complete, force can be applied to the protein by using 2 magnets (red and blue), the magnetic force (black arrow) stretches and unfolds the talin protein. As the protein unfolds, the bead moves (Δ Z), light diffraction changes, allowing detection of protein unfolding. Adapted from (Yan et al., 2015)

For the R1R3-17b protein to be used in single-molecule experiments it first had to be subcloned into a vector containing the Halo and AVI tags, the process of which is shown in fig.22. The vector used was a modified version of the pGEX vector group, which are a commonly used type of E.coli expression vector for GST purification (Borgi and Gargouri, 2008). This modified version, called pGEX-stretch, contained both tags needed for the single molecule work (AVI/HALO) as well as a GST tag to allow glutathione affinity chromatography purification (as shown in fig.22). A TEV site was included post GST tag to allow the GST tag to be cleaved away after initial protein purification. Between the AVI and Halo tags, a cut site for restriction enzymes was used, the two enzymes chosen were BamH1 and Xhol1, which were chosen as they both had high activity in the CutSmart® Buffer (from New England Biolabs). A buffer in which both enzymes exhibited low STAR activity (the cutting of similar, but not the intended sequences).

Before the R1R3 insert could be subcloned into the pGEX-stretch vector, it had to be removed from a standard (non expressing) cloning vector. Synthetic genes of R1R3-WT and 17b were ordered which contained the desired sequences with BamH1 and Xhol1 cut sites either side of the R1R3 insert. The insert was cleaved from the cloning vector using BamH1 and Xhol1 and then extracted from an agarose DNA gel using a DNA extraction kit (QIAgen). The same enzymes were used to open the blank pGEX-stretch vector. Once cleaved, the blank vector and the R1R3 fragment were ligated together to complete the stretchable talin vector. This completed vector was then transformed into DH10β DNA producing competent cells, then miniprepped and transformed into protein expressing BL21 competent *E.coli* cells, following standard protocol.

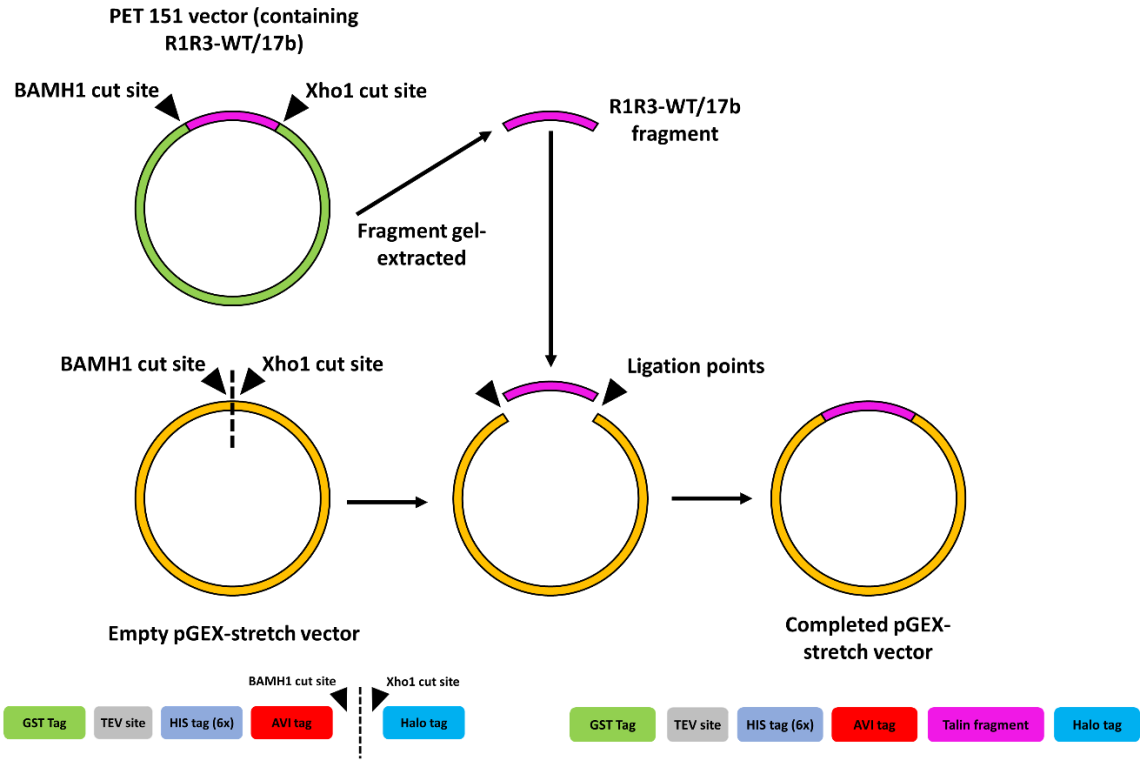


Figure 22: pGEX-Stretch subcloning process

Cartoon showing the process of subcloning a desired protein region (R1R3-WT/17b) into the pGEX-stretch vector. A synthetic gene containing R1R2-WT/17b within PET 151 vector (green) is ordered, containing restriction enzyme cut sites for BamH1 and Xho1). The desired talin fragment is cut from the vector using these enzymes in a restriction digest. Fragment is separated via a DNA agarose gel, which is gel extracted. The pGEX-stretch empty vector (orange), is opened via restriction digest (using BamH1, Xho1 cut sites). The gel extracted fragment is then ligated into the opened empty pGEX-stretch vector (orange + purple). The completed plasmid is run through a check digest to confirm ligation. Contents of completed vector are shown bottom right.

Once the transformation had been completed, the BL21 cells were grown in LB (Luria broth) media and were induced and harvested following standard protocols. Glutathione-agarose (Sigma Aldrich) was used to batch purify the expressed protein. The supernatant was loaded into a 50mL falcon tube containing 1mL of Glutathione-Agarose beads (ThermoFisher). Unbound contents were discarded, and the protein was then cleaved from the beads using TEV protease, this way in principle, only the protein (now lacking its GST tag) should elute.

As shown in fig.23, the R1R3-WT and R1R3-17b stretchable proteins expressed well

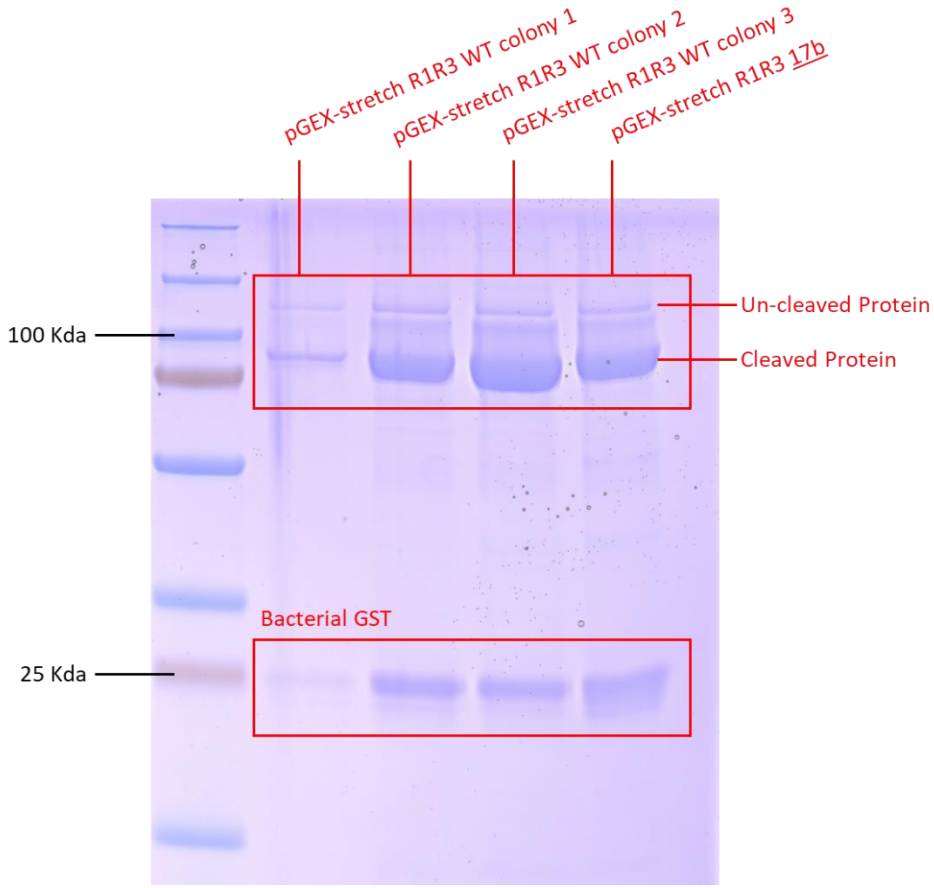


Figure 23: pGEX-Stretch subcloning results

Gel showing purified R1R3-17b/WT pGEX-stretch vector protein post cleavage from GST resin. Higher (heavier) 2 bands at ~120kDa indicates un-cleaved pGEX-stretch protein (containing GST tag) and ~90kDa indicates protein that has had its GST tag cleaved. The lower band at 25kDa, is the bacterial GST produced by the E.coli, as well as cleaved GST tags from the pGEX-stretch protein.

To confirm that the subcloning had worked correctly, both the R1R3-WT and 17b variants were sent for sequencing. For this sequencing, custom primers were designed: a pGEX 5' forward primer (pGEX_GST_rod_fw) which would confirm the insertion of the GST and AVI tags and the R1R3 insert, as well as a second forward primer (pGEX-Tln1-R1R3-17b_R3-FWD), with a binding site in R3. This second primer would confirm the insertion of the latter portion of the R1R3 insert as well as the insertion of the HALO tag. The sequencing came

back confirming the sequences had been assembled correctly and the protein was sent to our collaborators in Singapore.

3.1.12 Altered force loading capability of R1R2-17b

The data from the Singapore group (shown in fig.24) compares the force required to unfold a single molecule of either R1R3-WT or 17b. fig.24A-B shows the physical height of each bead anchored to the talin protein measured in nm. 3 separate unfolding steps (marked by a sharp increase in bead height over a short timespan) can be seen in R1R3-WT (fig.24A) and R1R3-17b (fig.24B). The force required for the 3 unfolding steps is represented as a histogram in fig.24Ai-Bi. The first unfolding step occurs around 5 pN. The fact this unfolding step is identical in both proteins strongly implies it is the unfolding step for R3 (which is identical in both constructs). However, despite the similarity of the R3 unfolding event, there is a visible difference between the unfolding steps of R1 and R2 in the two variants. In R1R3-WT the second and third unfolding steps are seen at 15 and 25 pN, while in R1R3-17b these unfolding events occur at 13 and 21 pN. This is a significant decrease in the force required to unfold the helical bundles within R1R3-17b. This data, combined with results of the CD conclusively show that the splicing in of exon 17b, reduces the force loading and stability of the helical bundles within R1R2.

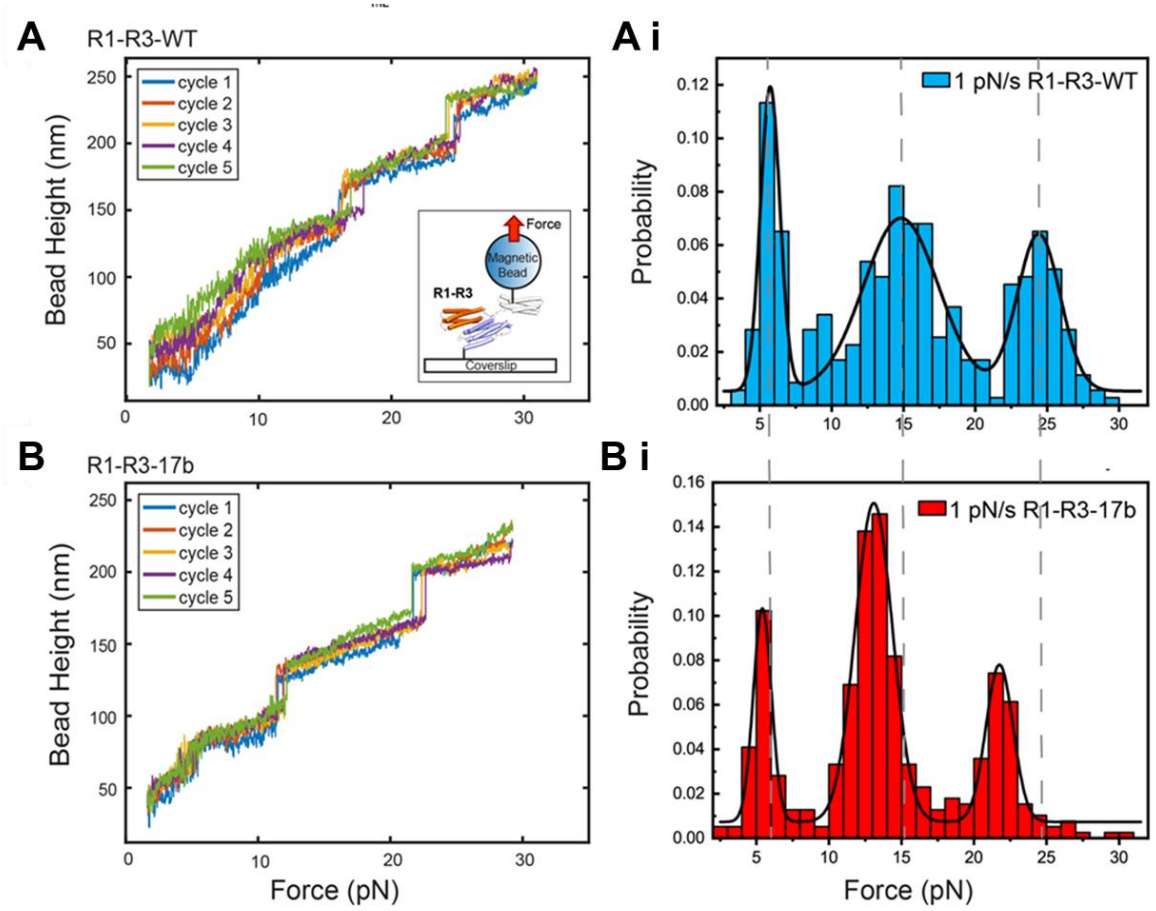


Figure 24: comparison of force loading capacity of R1R2-WT and 17b via single molecule stretching

Data from (Gallego-Paez et al., 2022) showing single molecule data from 7 molecules of R1R3-WT and 9 molecules of R1R3-17b. A-Ai). Timewise graph displaying the height of the talin-anchored bead (A, R1R3-WT Ai, R1R3-17b) under increasing force loading of 1 pN S^{-1} per scan. B-Bi). The average unfolding forces of R1R3-WT (B) and R1R2-17b (Bi) represented as a histogram. The grey dotted lines denote the unfolding of each helical bundle in R1R3-WT

17b discussion

3.1.13 17b is functionally and structurally different from the WT

The aim of the R1R2 17b project was to establish if the novel exon 17b, which when spliced into talin, would alter the proteins' structure and its functionality. Furthermore, the project aimed to establish a link between exon 17b and the cancers that were shown to upregulate exon 17b splicing.

The data presented here shows that R1R2-17b is a distinct protein from the WT, both structurally and functionally. The insertion of the 17 amino acids, caused by the splicing in of exon 17b, results in a variant form of talin that exhibits different properties from the WT. The melting curves and single molecule work (provided by the Singapore group) shows that the stability of the R1R2-17b helical bundles is significantly reduced resulting in a protein that unfolds at both a lower temperature and a lower mechanical force. Despite the altered physical properties of the 17b variant, the NMR and far-UV data supports that the majority of the structure of R1 and R2 are kept unperturbed by the 17 AA insertion, and still form helical bundle structures.

NMR and SEC data revealed that the external faces of the R1R2-17b helical bundles were shown to still allow binding of established ligands of folded R1R2 such as RIAM, and R1R2-17b also still showed affinity for vinculin, a binder of the unfolded form of the protein. The binding of both RIAM and vinculin confirmed that despite the insertion, R1R2-17b maintained functional binding sites. Furthermore, the insert did not reduce the functionality of the internal binding sites and 1:1, 1:2 and 1:3 complexes between talin and vinculin were observed in both the wildtype and in R1R2-17b, meaning the stoichiometry of the protein-protein complexes was maintained between both variants.

However, although the R1R2-17b protein remained functional, its biochemical properties were altered. When vinculin's ability to bind R1R2-WT and 17b was examined using SEC at both room and body temperature (~20°C and 37°C), it was found that the R1R2-17b variant protein bound more vinculin in both cases, but significantly more at body temperature. The high vinculin binding seen at 37°C implies that the R1R2 region is at least partly unfolded at a temperature similar to that found *in vitro* (37°C). This suggests the VBS within R1R2-17b may be active regardless of physical strain applied to talin as normal body temperature alone is enough to cause the helical bundles to separate and allow vinculin to bind. This is in comparison to the wildtype which showed little binding of VD1 even at 37°C, as it likely requires mechanical force to deform and expose the internal regions of the protein.

3.1.14 Predicted effect of R1R2-17b on focal adhesions

The lower force loading capability (the ability to mitigate and control cytoskeletal and extracellular strain) seen within R1R2-17b may prevent the protein and the FA it composes from functioning as intended. As discussed in the introduction, vinculin allows the continued survival of a focal adhesion by recruitment to talin within immature FA (Humphries *et al.*, 2007). If R1R2-17b recruited vinculin at a higher rate than the wildtype due to the increased accessibility of R1R2-17b VBSs (which was observed from the SEC data), in a cell this may stabilise and allow the maturation of more focal adhesions than a cell not expressing the exon 17b. Fig.25 predicts how an overly-active talin could result in more FAs by stabilising an abnormal number of immature FAs.

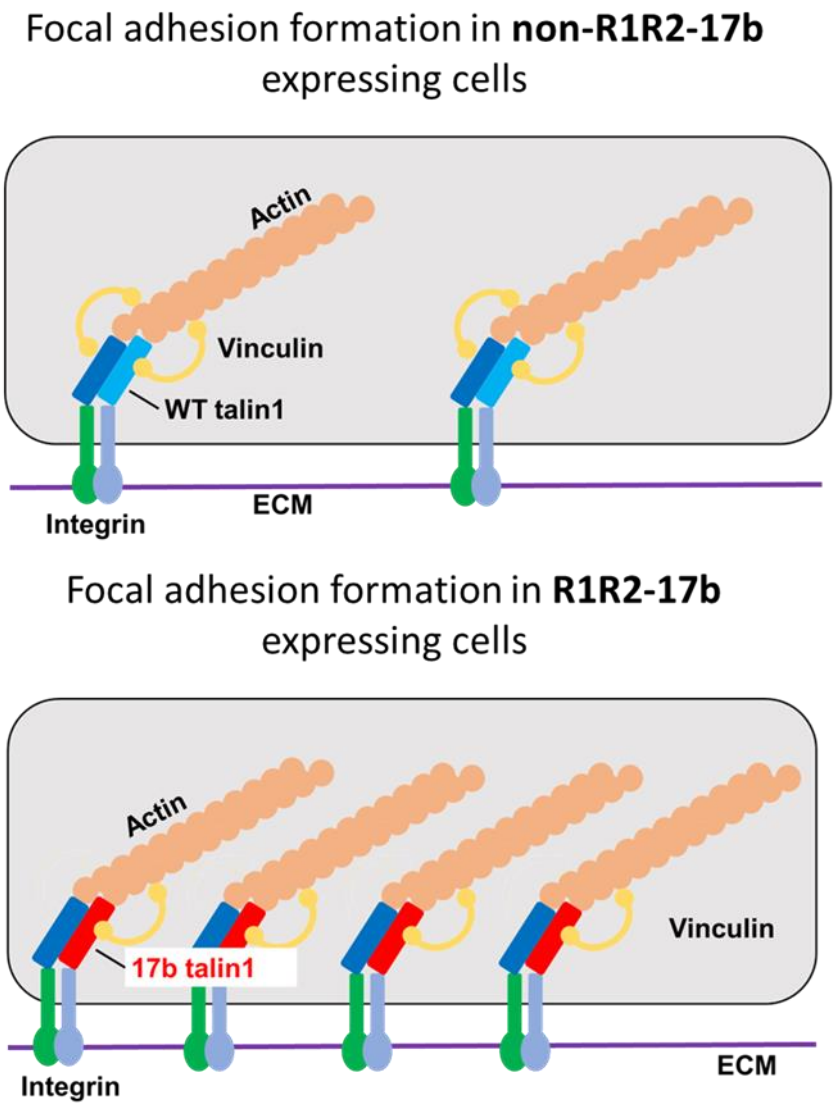


Figure 25: Predicted effect of exon 17b expression in focal adhesion formation

Cartoon showing the predicted effect of exon 17b splicing on focal adhesion formation. (top) R1R2-WT expressing cells forming standard focal adhesions, anchored to cytoskeletal F-actin (beige) and β-integrins (blue/green). WT Talin (blue) has 2 active vinculins (orange) bound, indicating ABS2 is active and the focal adhesion is mature and stable. In 17b expressing cells (bottom), the alternate talin (blue/red), displays incorrect force loading and increased vinculin binding, spreading available vinculin across an increased volume of immature focal adhesions.

3.1.15 The role of R1R2-17b in cancer motility and pathogenicity

Having showed conclusively that exon 17b expression resulted in a physically and chemically different protein, my collaborators on the R1R2-17b project (Gallego-Paez *et al.*, 2022) examined cancer cell lines (mouse embryonic fibroblasts (MEF)) expressing the R1R2-17b and WT talin to test the effects of the variant protein on the larger scale within the cell as a whole. Shown in fig.26D they found that the cells expressing R1R2-17b did indeed produce more focal adhesions than the WT agreeing with the predictions based on the biochemical data. 17b expressing cells also produced FAs that were significantly smaller than the WT (fig.26B) and those FAs were shown to recruit similar levels of vinculin as the WT (fig.26E). The properties of the 17b FAs agrees with my hypotheses that the variant protein should stabilise/produce more FAs by recruiting more vinculin to immature FAs, and as the 17b insertion does not affect the accessibility of the VBS, the same level of vinculin should be able to bind to the increased number of FAs.

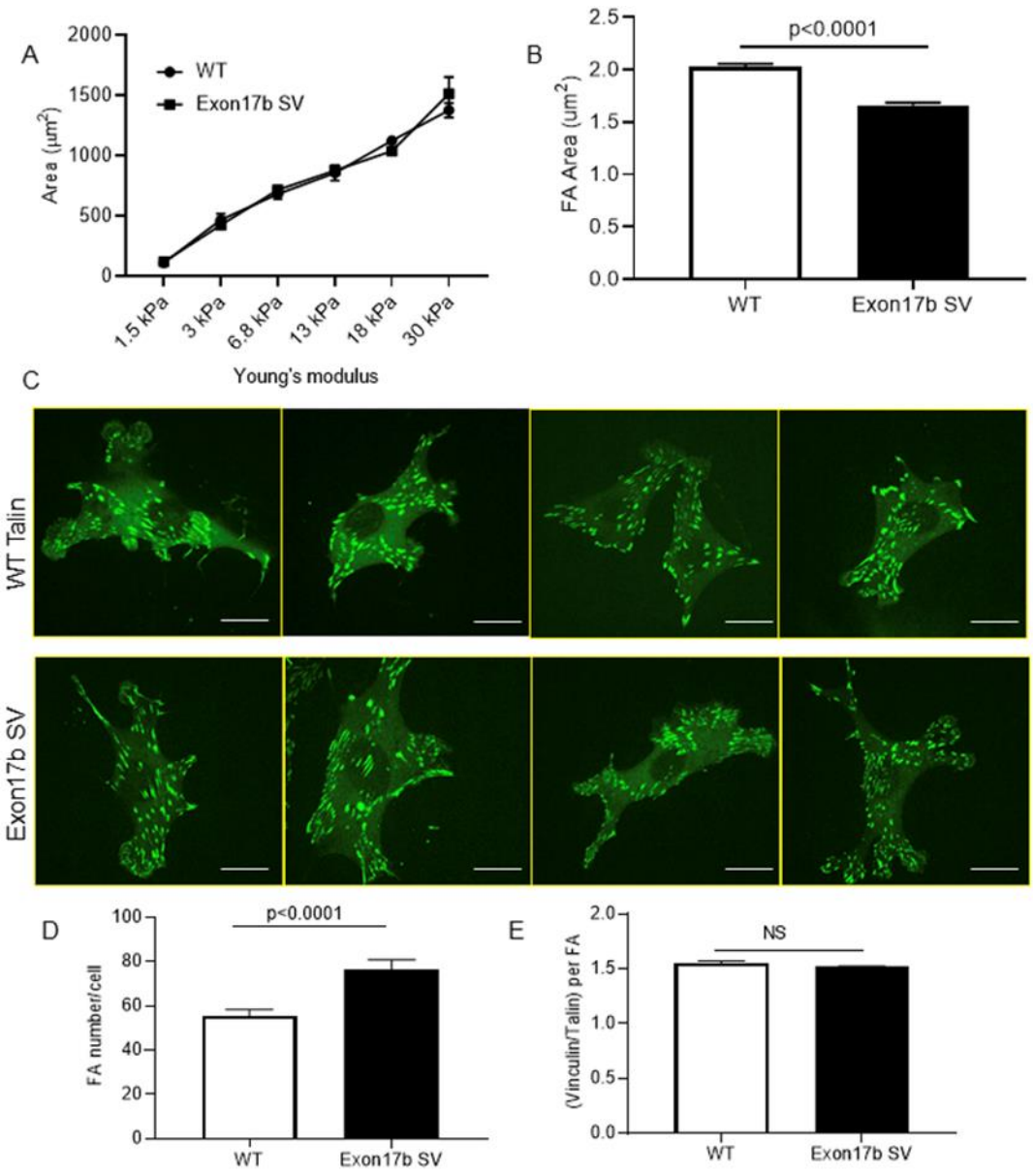


Figure 26: R1R2-17b in vitro data

Selection of cell culture work from ***(Gallego-Paez et al., 2022)*** *examining the effect of exon 17b in cancer mouse embryonic fibroblasts (MEFs). A). analysis of cell spreading in WT and 17b expressing MEFs. Cells were plated onto polyacrylamide gels of increasing stiffness (1.5 kPa-30 kPa) and area (μM^2) of cells was measured. Both WT and 17b expressing cells showed similar stiffness sensing and area increased with stiffness. B). comparison of area (μM^2) of focal adhesions between 17b and WT expressing cells. 17b expressing MEFs display a FAs significantly smaller in size than WT. C). microscopy images of MEFs (17b/WT) tagged with GFP talin, showing differences in focal adhesion localisation. D). comparison of number of FAs per cell (WT/17b), 17b expressing MEFs display significantly higher (~80) numbers of FAs per cell than WT (~60). E). comparison of the vinculin to talin ratio (volume of recruited vinculin) between WT and 17b. The ratio of vinculin to talin is very similar between the two.*

Talin in WT cells is unfolded by cytoskeletal strain. The resulting open talin configuration recruits vinculin which stabilises the immature FAs. Mechanical force therefore regulates FA formation and in doing so regulates many other cellular properties involving FAs, such as motility and propagation(Wozniak *et al.*, 2004). However, in cancers, that force dependant stiffness sensing ability of cells is commonly shown to be lost. If a component of the focal adhesions in cancer underwent changes in its force loading ability (as exhibited by R1R2-17B data), then it may produce FAs that can no longer 'feel' and interpret applied mechanical forces properly. These improperly functioning FAs (caused by the splicing in of exon 17b) may interpret their substrate as soft, which may be the reason why the FAs seen in 17b expressing cancer cells are smaller, as it has been shown that there is a direct correlation between soft substrates and the production of smaller FAs in cells. These smaller FAs are also correlated to the speed at which those cells are able to move, showing a gaussian relation between cell motility/speed and focal adhesion size (Kim and Wirtz, 2013). This supports the observed data as exon 17b expressing cancer cells not only produced smaller FAs, but 17b cancers were also negatively correlated with extra-mesenchymal transition (EMT), meaning 17b expressing cells are less prone to motility and metastasis (Gallego-Paez *et al.*, 2022). Overall, this suggests that the expression of exon 17b in cancer cells reduces their ability to correctly sense mechanical strain, this in turn reduces the size of their focal adhesions, which prevents from undergoing EMT and becoming more pathogenic.

This may be advantageous to the cancer as becoming mobile too early may bring it into contact with the immune system before it has gained the ability to evade it, meaning exon 17b could act as a 'brake' on cancer malignancy until it is ready to become metastatic.

3.1.16 R1R2-17b may allow cells to change the mechanosensitive ability of their focal adhesions in response to injury related matrix stiffening

Exon 17b has been shown here to alter the mechano-sensing ability of cancers that express the alternate exon due to their creation of 'faulty' focal adhesions which can no longer detect the differences between hard and soft substrates. However, exon 17b is not the result of a mutation like that seen in SCLS. Instead, exon 17b is a natural exon within the talin 1 gene. Studies done by our collaborators into the levels of exon 17b expression in non-cancerous tissue showed that most tissues types such as muscle tissue displayed negligible levels of 17b. However, a significant group of other tissue types did display exon 17b expression. The liver, spleen, skin, pancreas and the pituitary gland, as well as the sex organs (testis and endocervical tissue) all expressed exon 17b to some extent. This data suggests the exon must serve a functional purpose within non-cancerous cells.

After tissue damage has occurred within the body, be it pathogenic (such as bacterial infection) or physical (such as cuts or burns), it results in changes within the mechanical homeostasis of the effected tissue. In response, cells exert increased strain on the ECM which stiffens in response to the increased physical stress (fig.27) (Jansen *et al.*, 2013).

When plated on unstressed collagen gels, cells (fibroblasts) adhere to the gel and form mechanical linkages (focal adhesions) and when the stress/strain is increased or decreased on the gel, the cells increase or decrease their own strain to return to a homeostatic state of force, this is referred to as tensional homeostasis (Humphrey, Dufresne and Schwartz, 2014). Where the injury cannot be healed, fibrous scar tissue is created by myofibroblasts. This scarred tissue displays a permanent increase in mechanical force on the tissue. Exposure to this increased stress in turn upregulates myofibroblast activation(Aarabi *et al.*,

2007). I propose (shown in fig.27) that this positive feedback loop of scar tissue ECM increasing cell stiffening could be reduced, and myofibroblasts could self-regulate their activation by reducing their own ability to detect the increased stiffness of scar tissue. By reducing the detection of ECM strain on the myofibroblasts, they could deactivate themselves, breaking the feedback loop. The expression of exon 17b in myofibroblasts could produce the less sensitive focal adhesions needed for this to occur. The usage of exon 17b in this way is supported by data showing that in scarred liver tissue (a tissue shown to express 17b), as recovery occurs, myofibroblasts quiesce and reduce activation without any external input (Tschumperlin *et al.*, 2018), resulting in a reduction in their expression/production of fibrotic proteins such as collagen type 1 (Kisseleva *et al.*, 2012). Therefore, I suggest that the intended purpose of exon 17b is to prevent excessive production of scar tissue within damaged tissues. These properties produced by exon 17b splicing are simply hijacked by cancer cells to allow them time to increase pathogenicity prior to immune detection.

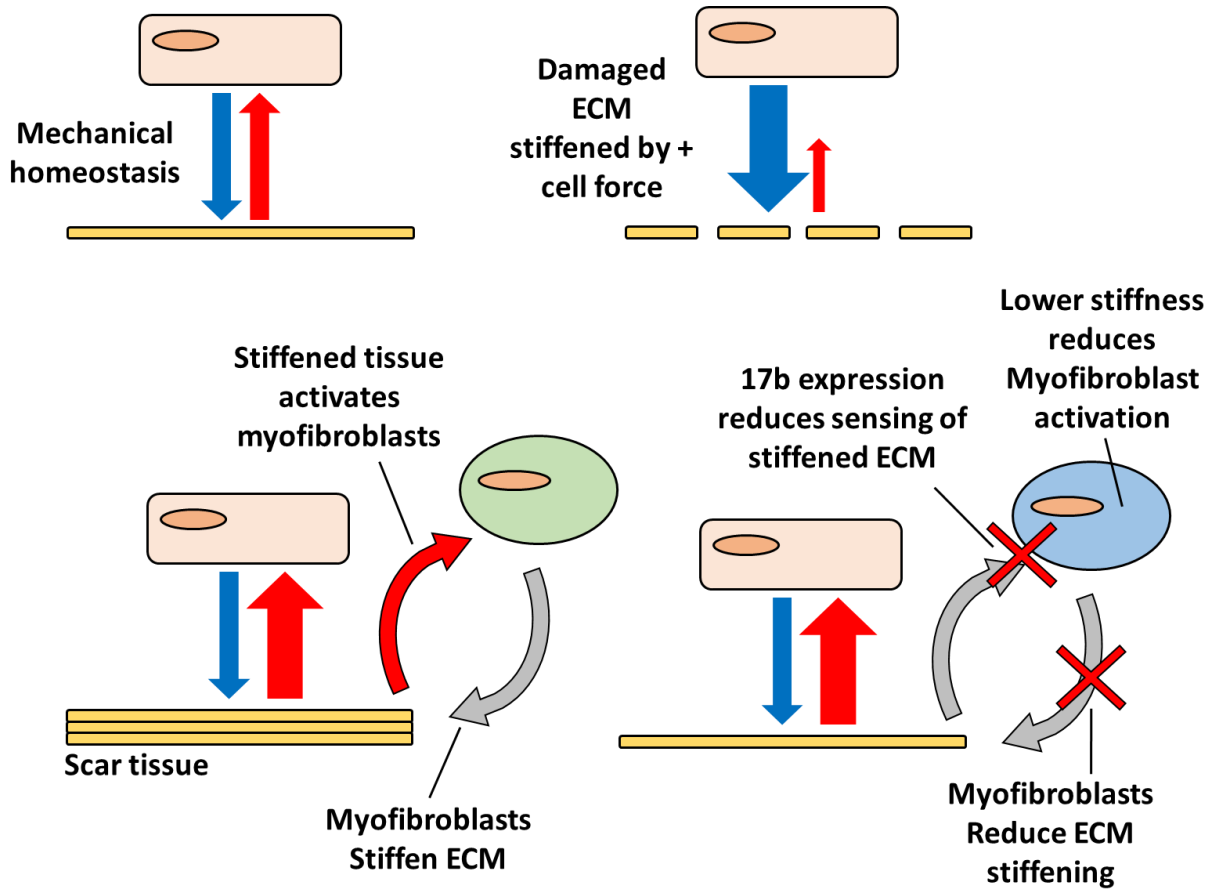


Figure 27: Expression of exon 17b to downregulate ECM force signals

Cartoon showing predicted role of exon 17b in WT cells to reduce force exerted on cells by the ECM. (Top left) shows mechanical homeostasis between the cell exerting force of the ECM (blue) and the ECM exerting force on the cell (red). (Top right) when ECM is damaged during injury (reducing force, red arrow), anchored cells, exert increased force (blue arrow) to stimulate stiffening of the ECM to restore mechanical homeostasis. (Bottom left) if healing does not occur, fibrous scar tissue forms, which exerts continuous increased force (red arrow) on cells. This increased ECM stiffness, recruits and activate myofibroblasts, which secrete fibrous proteins, spreading scar tissue. This creates a positive feedback loop, resulting in continued scarring. (Bottom right) predicted role of exon 17b. To allow myofibroblasts to escape the feedback loop, they switch from WT talin to 17b. The altered force loading of 17b reducing stiffness sensing of the myofibroblasts (grey arrow), which deactivates the cells (lower grey arrow).

3.1.17 Usage of R1R2-17b as a risk stratification method for breast cancers

It was also found by our collaborators (through reverse transcriptase PCR of cancer cell line DNA) that cancers expressing 17b had different resistances to certain drug types (fig.9). 4 different breast cancer cell lines were analyzed, 2 expressing low levels of exon 17b (BT549 and MDA231) and 2 expressing high levels of exon 17b (ZR751 and BT20). The cell lines expressing exon 17b in higher amounts showed increased resistance to phosphoinositide-3-kinase (PI3K) signaling and cytoskeletal regulation targeting drugs (fig.9C). This data shows that cancers splicing in exon 17b improves their survival against known anti-cancer treatments, which is important as, in the future, detection of exon 17b splicing could be used as a factor in stratifying the risk of cancers (breast cancers in particular) and provide insight on whether certain drug courses would be successful.

Chapter 3.2 The role of the systemic capillary leak syndrome deletion in the pathogenicity of the disease

As mentioned in the introduction, the condition known as systemic capillary leak syndrome (SCLS) was found to be inherited in an autosomal dominant manner in a recently discovered family lineage. Within the family many suffered from the disease, and all possessed a distinctive and inherited point mutation in the talin 1 gene. This mutation was believed to be a potential cause of the disease as the failure of cell adherens junctions is a key trait of the disease, and talin plays a major role in these adherens junctions. I aimed to examine the effects of the SCLS mutation in a similar fashion to that of 17b, with an initial bioinformatic analysis of the affected region (R13) in WT and SCLS variants. This would be followed by the analysis of both the physical and functional characteristics of the variant protein.

3.2.1 Discovery of a talin exon skipping event in SCLS patients

This section of the project centres around a recent discovery by N. Elephant and R. Sion-Sarid (Elefant *et al.*, 2022) who I have collaborated with, their discovery shows an inherited mutation which I (and my collaborators) belive leads to SCLS. 3 distantly related SCLS patients had their family pedigree examined which revealed a mutation in their talin 1 gene expressed by all 3 members of the family. The mutation was found to the talin 1 gene. The family pedigree is shown in fig.28, which shows that (assuming SCLS affected family members who died young before the pedigree was produced) the mutation showed an autosomal dominant inheritance pattern with incomplete penetrance.

Despite the lack of a known cause for SCLS (until recently) there has been promising treatments for the disease, although no cure. The use of a combination of terbutaline/various β_2 agonists (medications that prevent bronchospasm/relax muscles) alongside immunoglobulin therapy showed a significant decrease in episodes of the disease and an increase in survival over 5 years, from 20% (in patients not on the therapy) to 85% (in treated patients) (Gousseff *et al.*, 2011).

Although a useful preventative measure, without knowledge of the cause of the disease it is more difficult to devise an effective cure. This is why the discovery of an inherited mutation is so important, as it may reveal a therapeutic target for drug treatments and allow screening for high-risk patients before the onset of SCLS.

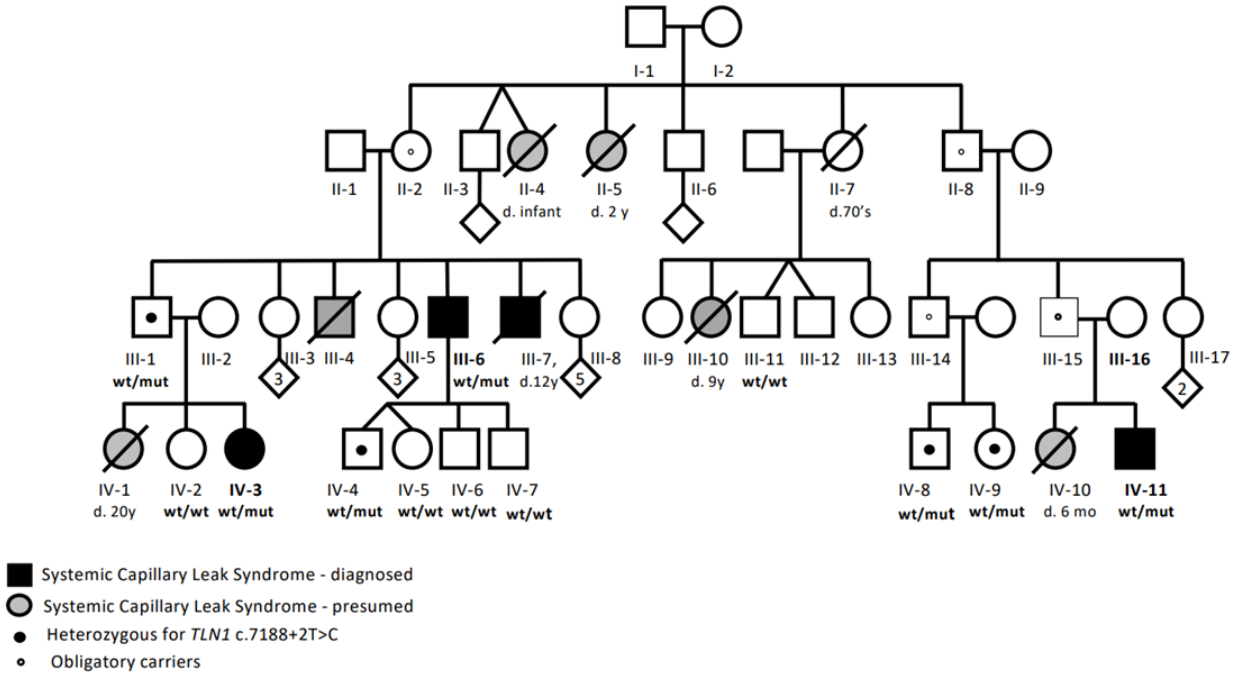


Figure 28: SCLS Exon 54 skipping mutation family pedigree

Data from (Elefant et al., 2022)

Family pedigree from (Elefant et al., 2022), showing recently discovered inherited mutation linked to SCLS. Black coloured subjects are/were affected by SCLS, due to the death of family members prior to discovery (diagonal line through subject) some are presumed to have had SCLS, these are shown in grey. Based on data shown, the inheritance pattern was predicted to be autosomal dominant.

3.2.2 SCLS predictive modelling

When the details of this point mutation were examined, it was found that the mutation resulted in the skipping of talin 1 exon 54. This alternative splicing event removed an exon that encoded for a region of the 13th rod domain of talin 1 (R13) as seen in fig.29A that resulted in a deletion that occurs between residues V2376 to A2396 and shortened the R13 domain by 21 amino acids.

This deletion is of interest as it significantly changes the length of a helical bundle within a key region of talin 1, the affected R13 region contains both a VBS (fig.29B green helix) as well as primary actin anchorage point in talin 1 ABS3 (fig.3). As with the splicing event seen

in R1R2-17b any changes to the structure of a helical domain such as R13 could potentially alter its stability and function.

The first step in evaluating the effect of the exon 54 skipping event was to produce an AlphaFold model of the affected R13 region to compare to the WT. The R12 domain situated before the affected R13 was included in the model as this would mimic the constructs used in the later biochemical analysis, as R13 alone was known to be insoluble. The proteins modelled were the WT R12 to DD region (R12-DD WT) and model of the same regions but missing the 21 residues lost in the exon skipping event. The modelled mutant is referred to as R12-DD SCLS.

Figure 29 shows these AlphaFold models alongside the data on the accuracy of the models. The pLDDT scores for both models are shown in fig. 29, A and D, which show the majority of both models have a pLDDT score of >80 which suggests these regions are accurate. The ends of the helices present in both the WT and SCLS variants have a lower score, in particular the N and C terminal regions of the R12R13 WT and SCLS models. These areas have a pLDDT score below 50 and should not be considered to be accurate. Fig. 29, B and E show these pLDDT scores mapped onto their respective AlphaFold models. The positional error of the AlphaFold generations are shown in the PAE graphs (fig. 29 C and F). The WT and SCLS variants show a similar pattern, with the helical bundles R12 and R13 showing low positional error (the blue regions), the low scores outside of the helical regions indicate that although the bundles themselves are positioned accurately, the interaction/position of the two bundles in respect to each other is likely modelled incorrectly.

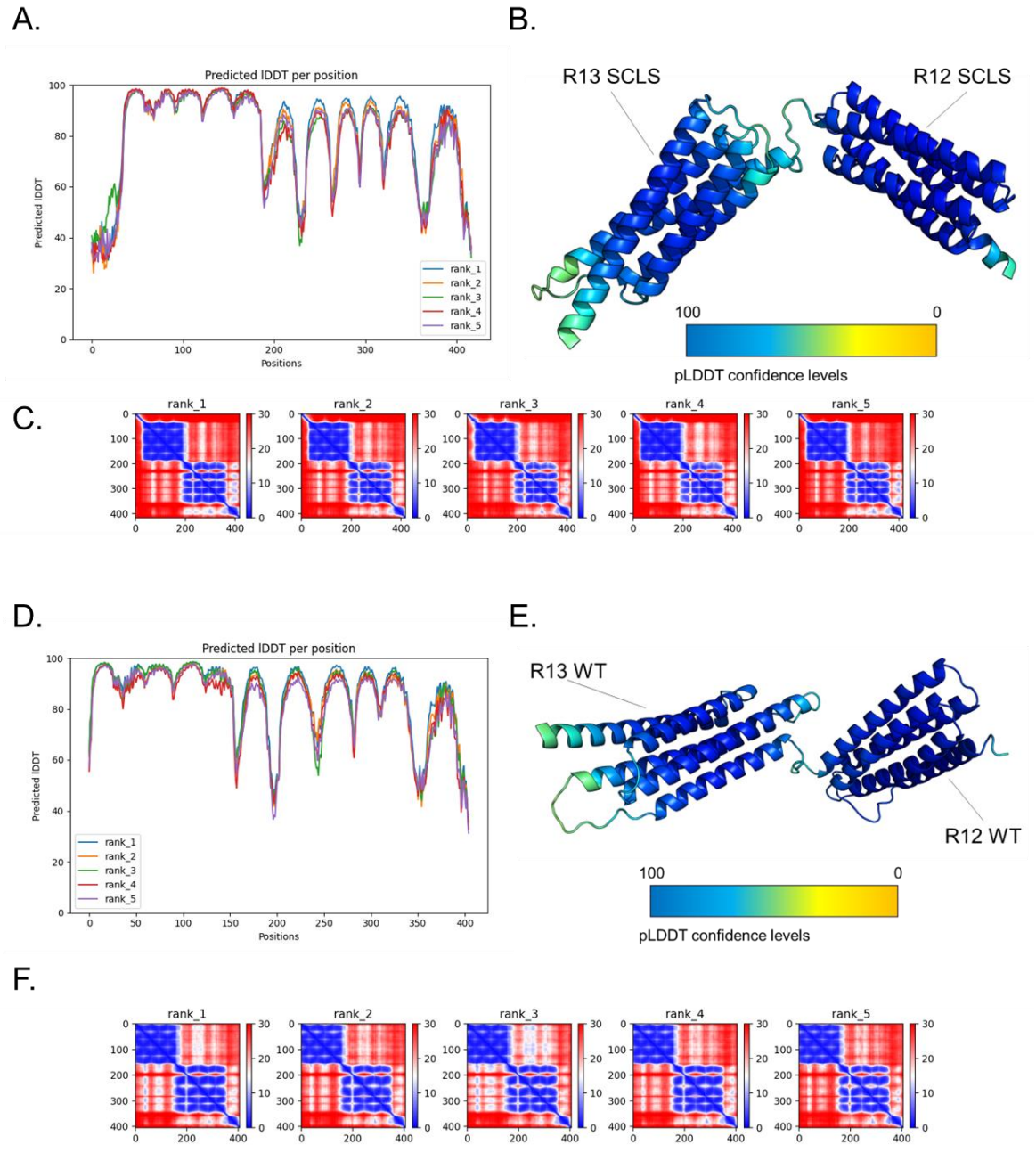


Figure 29: Reliability of AlphaFold2 R12R13 WT/SCLS prediction

A,D). predicted Local Distance Difference Test (plDDT) data and (C,F). Predicted Aligned Error (PAE) data provided by AlphaFold2 on the accuracy of the R12R13 Wt/SCLS generated structures. B,E). The pLDDT scores are shown mapped onto their respective residues on the models for R12R13 WT and SCLS, with higher pLDDT scoring residues coloured blue and lower scoring residues coloured yellow/orange.

The AlphaFold models of both can be seen in fig.30B, of which only R13 is shown. Both models predicted an alpha-helical bundle structure. R12-DD WT has the helix affected by the mutation marked in blue, with the region lost during the exon skipping event marked in magenta. Fig.30C shows a close-up of the third helix where the deletion was found to occur, and a resulting kink has been introduced into the R13-SCLS bundle structure where the magenta region is missing. A portion of the following 4th helix of R13-SCLS is now incorporated into the linker between the two helices. This deletion and helix bending is a significant change to the structure of R13 and warranted further investigation. The impact on the 3rd helix was predicted, much like the effect of R1R2 17b, to reduce bundle stability, and was also predicted to increase expose of VBS and other cryptic ligand binding sites as a result.

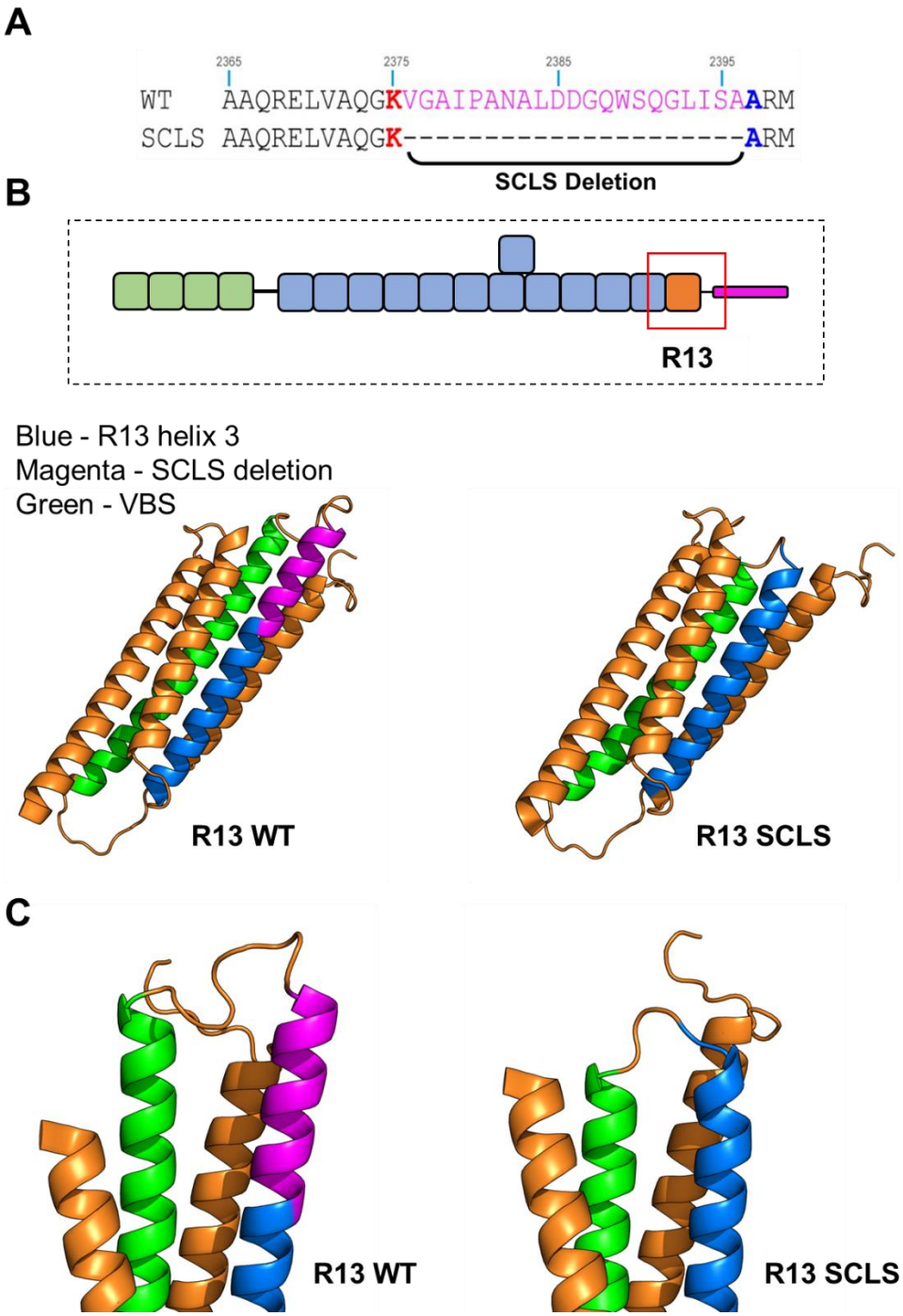


Figure 30: Predictive modelling of the effect of exon 54 skipping

Predictive modelling of the effect of SCLS induced skipping of exon 54 on talin 1 R13. A). A sequence alignment of WT talin 1 and talin affected by the exon 54 skipping. The residues either side of the deletion (K2375 and A2397) are shown in red and blue respectively, the section deleted is shown in magenta. B). PyMol images (as well as a simplified depiction of the whole talin protein) showing the region affected by the exon skipping, (left) crystal structure of talin 1 R13(orange), the affected 3rd helix(blue), the R13 VBS (green) and the region lost due to exon 54 skipping (magenta) are highlighted. (right) AlphaFold 2 generated model of R13-SCLS. C). Close-up of the R13 3rd helix (blue) in WT and SCLS variants. Exon 54 deletion also leads to loss of most of the linker between the 3rd and 4th helices, incorporating a portion of that helix into linker.

3.2.3 SCLS biophysical characterisation

After producing predictive models of the two proteins, synthetic genes of both R12-DD SCLS and R12-DD WT were produced by GeneArt (Thermo Fisher Scientific). These plasmids were transformed into *E. coli* BL21 competent cells. Both proteins expressed nicely and were purified using nickel and ion chromatography.

The first analysis was to confirm the proteins were folded and to examine their stability and compare differences between variant and WT, to do this, CD was used. Fig.31A shows the far UV spectra of both R12-DD SCLS (red) and R12-DD WT (blue). The spectra shows both to be predominantly alpha-helical in nature at 20°C (solid red and blue lines). However, when the far UV was conducted at 90°C (dashed red and blue lines), there was a notable difference between the structure of the two proteins. R12-DD WT and SCLS had similar CD minima during scans at 20°C, but when heated to 90°C, R12-DD WT displayed a lower minima of -80 mdeg while R12-DD SCLS showed a minima of -30 mdeg. This difference in absorbance at 90°C indicates that R12-DD WT is still partly folded at 90°C, while the R12-DD SCLS mutant is fully unfolded by that temperature. Meaning a region of the WT that is usually stable at very high temperatures has been destabilised by the deletion seen in the SCLS variant. This destabilised region was predicted to be the R13 helical bundle as it is the region affected by the SCLS deletion.

To further support the prediction that the unfolded region is R13 are the results of the melting curves of both proteins shown in fig.31B. R12-DD WT shows a two-step unfolding process with two different unfolding events at 60°C, and a second beginning at 90°C likely due to the two helical bundles present in the R12-DD WT. The TM of the second unfolding step cannot be calculated as the full unfolding event likely occurs beyond 100°C where the

sample would boil and the machine no longer able to obtain data. The R12-DD WT spectra is in contrast to the spectra that is seen in the R12-DD SCLS mutant. R12-DD SCLS is only seen to have a single unfolding step which also occurs at ~60°C. As the only difference between the two proteins is the deletion within R13, the loss of the second high temperature unfolding event, implies the R13 in R12-DD SCLS now unfolds at 60°C. This aligns with the predictions of both the AlphaFold models and the far UV spectra of R12-DD SCLS, as the deletion has reduced the stability of the R13 helical bundle causing it to unfold at a significantly lower energy than normal (coinciding with the temperature required to unfold R12).

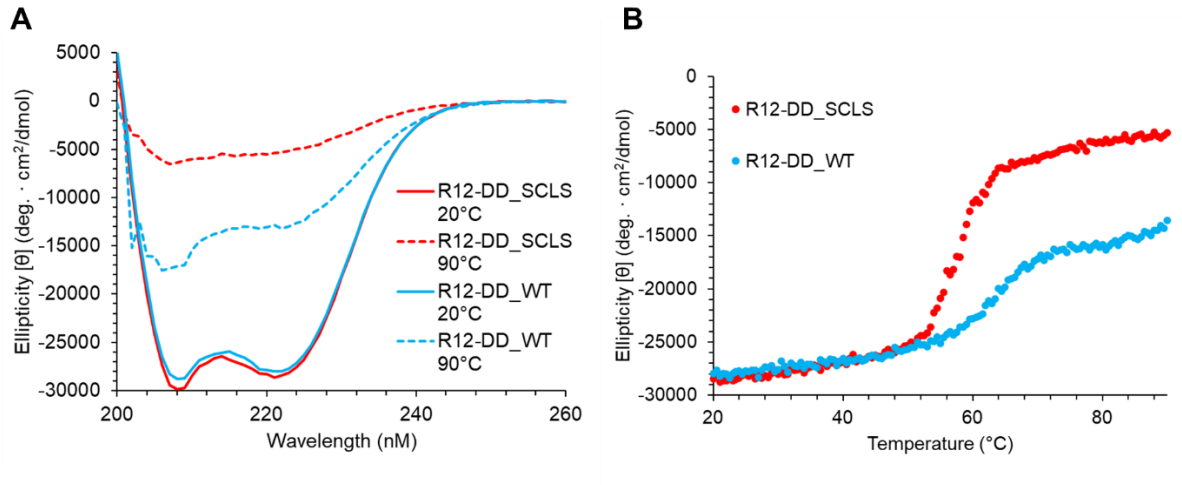


Figure 31: Circular dichroism comparison R12DD-WT/SCLS

Circular dichroism data from R12-DD-WT and R12-DD-SCLS. A). Far-UV data comparison between WT and SCLS variant proteins, conducted between 200 and 260 nm wavelengths. Samples were in Q column affinity buffer A and at 0.5 mg/mL protein concentration. R12-DD-WT (blue) and SCLS (red) show the 'double dip' motif (at 20°C) between 205-225 nm wavelengths. This indicates that both possess an alpha helical structure. Far-UV spectra taken at 90°C shows WT (blue dashed line) minima at -18000 ellipticity while R12-DD-SCLS (red dashed line) minima is -6000 ellipticity. Indicating that the WT is not completely unfolded while the SCLS variant is. B). Melting curve data comparison between R12-DD-WT/SCLS, conducted between 20-90°C. R12-DD-WT (blue) shows an unfolding event at 60°C and one beginning at 90°C, ending with a minima of -15000 ellipticity. R12-DD-SCLS (red) shows only a single unfolding event also occurring at 60°C, which results in a minima of -5000 ellipticity, implying R12-DD-SCLS is fully unfolded at 90°C while WT is still partially folded.

3.2.4 SCLS biochemical characterisation/functionality

With conformation that the deletion in R13 caused by the SCLS mutation does in fact reduce the stability of the helical bundle, I next wanted to examine the effect of the deletion on functionality of the variant R12-DD SCLS protein. Based on the results of the CD, as well as the effects seen in R1R2 by the 17b insertion, I predicted that due to lowered stability of the R13 helical bundle, the availability of internal binding sites such as VBS would be increased.

3.2.5 Initial size exclusion analysis of R12R13-SCLS

To test the hypothesis that the R12-DD SCLS VBS functionality would be affected, I first needed a baseline for the SEC protein-protein analysis. To generate the baseline, both R12-DD WT and SCLS variants were run alone at 100µM concentration (fig.32). It was found that R12-DD SCLS (fig.32, red) eluted earlier at 12mL than the WT at 12.3mL (fig.32, blue) despite being smaller in molecular weight (R12-DD WT, 43 kDa, R12-DD SCLS, 42 kDa). The proteins were run by themselves to provide a baseline for later ligand binding experiments, but the experiment also reveals more evidence that R13 in R12-DD SCLS is not as folded as the R13 in the WT protein. For a protein to elute earlier than another from an SEC column it must be larger (usually meaning a higher MW), but if the protein is less folded (having a larger surface area) it will elute earlier than its MW would suggest. R13 in R12-DD SCLS must therefore not be as folded as R12-DD WT.

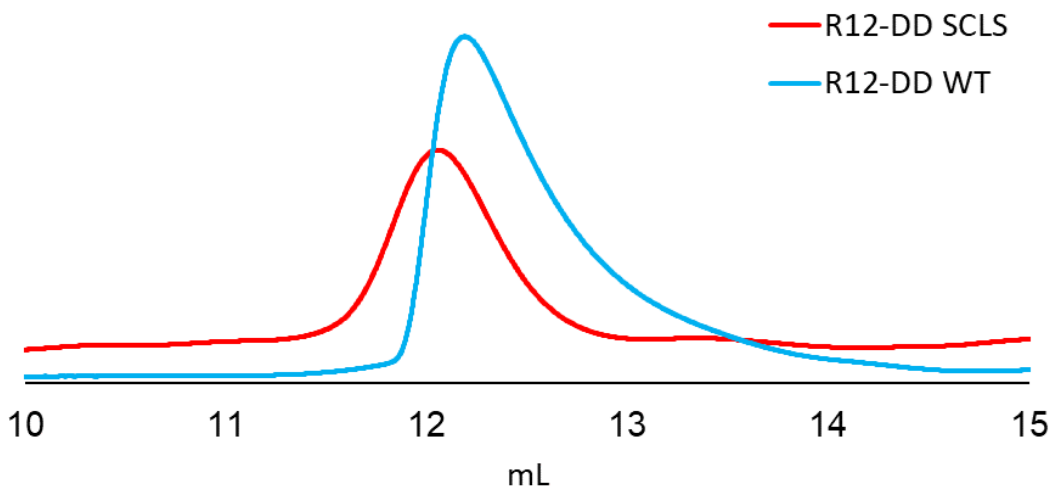


Figure 32: Gel filtration baseline data for R12DD-WT/SCLS

Gel filtration data comparing elution positions of R12-DD WT and SCLS. Samples were 100µL of 100µM talin protein in Gel filtration buffer (20mM Tris (hydroxymethyl), 150mM NaCl, 3mM DTT, pH 8). Both R12-DD WT (blue) and R12-DD SCLS produced a single monomer peak.

3.2.6 Comparison of vinculin binding between R12R13-SCLS and WT

With a baseline to compare against, SEC was used to compare the vinculin binding ability of R12-DD WT and the SCLS variant. R12-DD WT was run at 100 µM in a 1:1 ratio with VD1 (100 µM) after 1 hour incubation at room temperature. There is only 1 VBS present in the R12-DD region so only 100 µM vinculin (VD1) was required to satisfy all present binding sites in the samples 100 µM talin. Shown in fig.33 is the elution of the 1:1 VD1 ratio (fig.33A, solid blue line) compared to the R12-DD WT run alone (fig.33A, blue dashed line), the talin peak has reduced in intensity significantly and the free VD1 peak (which elutes at 15mL) is no longer present. A new, high intensity, 1:1 complex peak is seen at 10.5mL. R12-DD SCLS was also incubated at room temperature with a 1:1 ratio of VD1 as shown in (fig.33B, solid red line). The elution profile was overlayed with the baseline (fig.33B, red dashed line). When vinculin is added to the R12-DD SCLS sample, the intensity of the free talin peak was reduced and a larger, 1:1 complex elutes at 11mL. The free VD1 peak at 15mL has also

reduced in intensity significantly. It is interesting to note that the 1:1 R12-DD SCLS to VD1 complex is smaller than the 1:1 complex of the WT despite free R12-DD SCLS eluting earlier when unbound to VD1, this suggests that VD1 still locks the SCLS variant in a completely unfolded state as observed in the WT, this unfolding of R12-DD SCLS therefore results in the protein eluting at the position expected from its smaller molecular weight.

The SEC data confirms that at room temperature R12-DD SCLS is still functional and able to bind vinculin, however, the presence of a free talin peak means that some VBS remain cryptic but the majority have bound VD1. There is little difference in intensity of the complex peaks between R12-DD SCLS and WT, suggesting that neither variant binds vinculin more efficiently than the other.

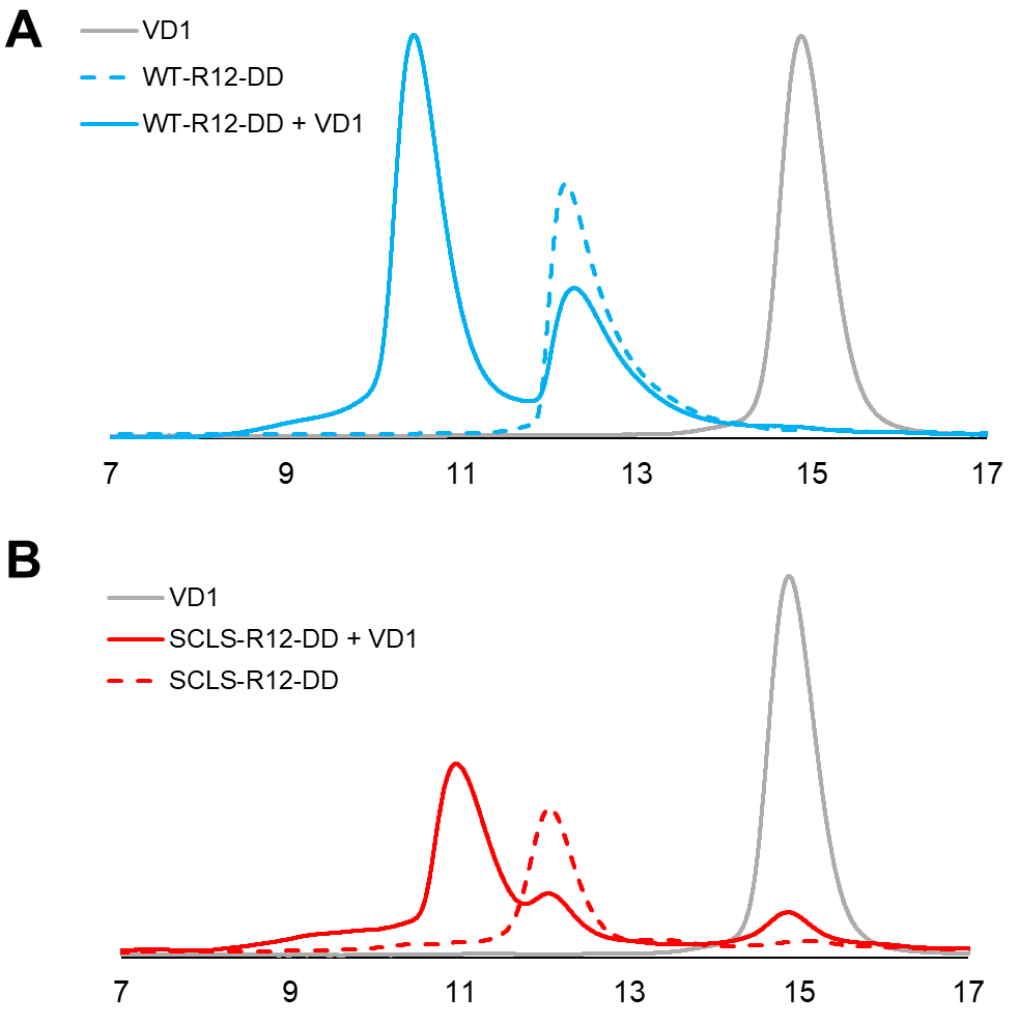


Figure 33: Vinculin binding ability between R12DD-WT/SCLS

Gel filtration data comparing vinculin binding ability of R12-DD WT/SCLS. Samples were 100 µL of 100 µM talin protein and 100 µM VD1 in Gel filtration buffer. A). R12-DD WT alone (blue dashed line) and R12-DD WT +100 µM VD1 (blue). VD1 was run alone at 100 µL for reference. A complex peak can be seen at 10.5 mL, indicating binding of VD1. B). R12-DD SCLS alone (red dashed line) and R12-DD SCLS + 100 µM VD1 (red). Complex peak is formed as seen by new peak at 11 mL, indicating bound VD1. The data suggests both proteins have similar affinities for vinculin due to the similarity in complex peak intensity.

SCLS discussion

3.2.7 R13-SCLS less stable than WT, but the VBS not more active/less cryptic

Overall, the data from the biophysical and biochemical characterisation suggests that the mutation causing the exon skipping event of exon 54 results in the production of a variant form of talin 1 R13 which is significantly destabilised and unfolds at a lower temperature. Based on the data from the R1R2-17b experiments a destabilised helical bundle will have altered force loading capability, requiring less tension within a FA to unfold the protein. The stability changes were not shown to affect the binding sites of the helical bundle as the SEC data shows that VBS within R13 are functional.

Unlike the alteration of R1R2 seen in the exon insertion of 17b, which resulted in a talin region with more active vinculin binding capability, the SCLS variant of R13 does not exhibit more vinculin than the WT. This suggests that although R13-SCLS stability is reduced, the helical bundle is no more unfolded at room temperature than the WT protein.

3.2.8 Preclusion of R12R13-SCLS from actin binding assay

It must be noted here that the R13 region contains the key actin binding region for talin 1, ABS3. The mutation causing SCLS, alters the helices within the R13 region. Therefore, it would seem pertinent to examine the effects of this variation to the talin R13 structure and if it alters the actin binding ability of the helical bundle. I did not perform any experiments on the function of R13 actin binding as previous experiments have been conducted on the R12R13-SCLS protein, which showed it was not viable for actin experimentation. The standard assay to measure the affinity of a protein for actin is an actin co-sedimentation assay. The assay consists of F-actin filaments and the protein of interest being spun at high

speeds (100,000 G) (Heier, Dickinson and Kwiatkowski, 2017). If the protein binds actin it will pellet out together with the filaments (unbound protein will stay in solution) and can be analysed via SDS-PAGE. This assay was conducted prior to my work on the SCLS mutation and revealed that R12R13-SCLS aggregated and pelleted out regardless of actin presence at high speeds. This precluded the protein from all standard methods to experiment actin binding.

3.2.9 Reduced stability in SCLS talin leads to increased vinculin localisation

The biochemical analysis of the protein showed its reduced stability, which should imply a reduced force loading capability, and increased exposure of VBS. The SEC data did not show the expected result as VBS exposure was similar in both WT and SCLS variants of R13. However, to confirm if the VBS activity increases within an SCLS affected FA, the proteins' effects must be observed within a live cell FA.

My collaborators were able to validate these predictions with cell culture experiments. My collaborator, G. Rouni produced mouse endothelium cell lines which expressed the R2510A variant of talin 1 (Elefant *et al.*, 2022), a mutation that mimics the predicted reduced stability and altered actin binding observed in the biochemical data. The vinculin and VE (vascular endothelial) cadherin junctions within these cells were fluorescently tagged to show localisation of these proteins, the cells were imaged under a microscope as shown in fig.34. The localisation of vinculin (fig.34A, magenta) was much more concentrated at the membrane of the cells expressing R13-SCLS as opposed to the membrane of WT cells. The opposite was seen in the localisation of VE cadherin junctions (fig.34A, white and green). VE-cadherin junctions were more prevalent in the WT cells and were also more tightly organised around the junctions between the cells, with a higher ratio of vinculin localised

with them in SCLS cells (fig.34B). In comparison, the cells expressing R13-SCLS had fewer overall cadherin junctions which were not tightly organised at cell junctions, instead being spread out throughout the cytoplasm. The vinculin in SCLS cells was also more spread out, despite the total concentration of cellular vinculin being consistent between both SCLS and WT cells (fig.34C)

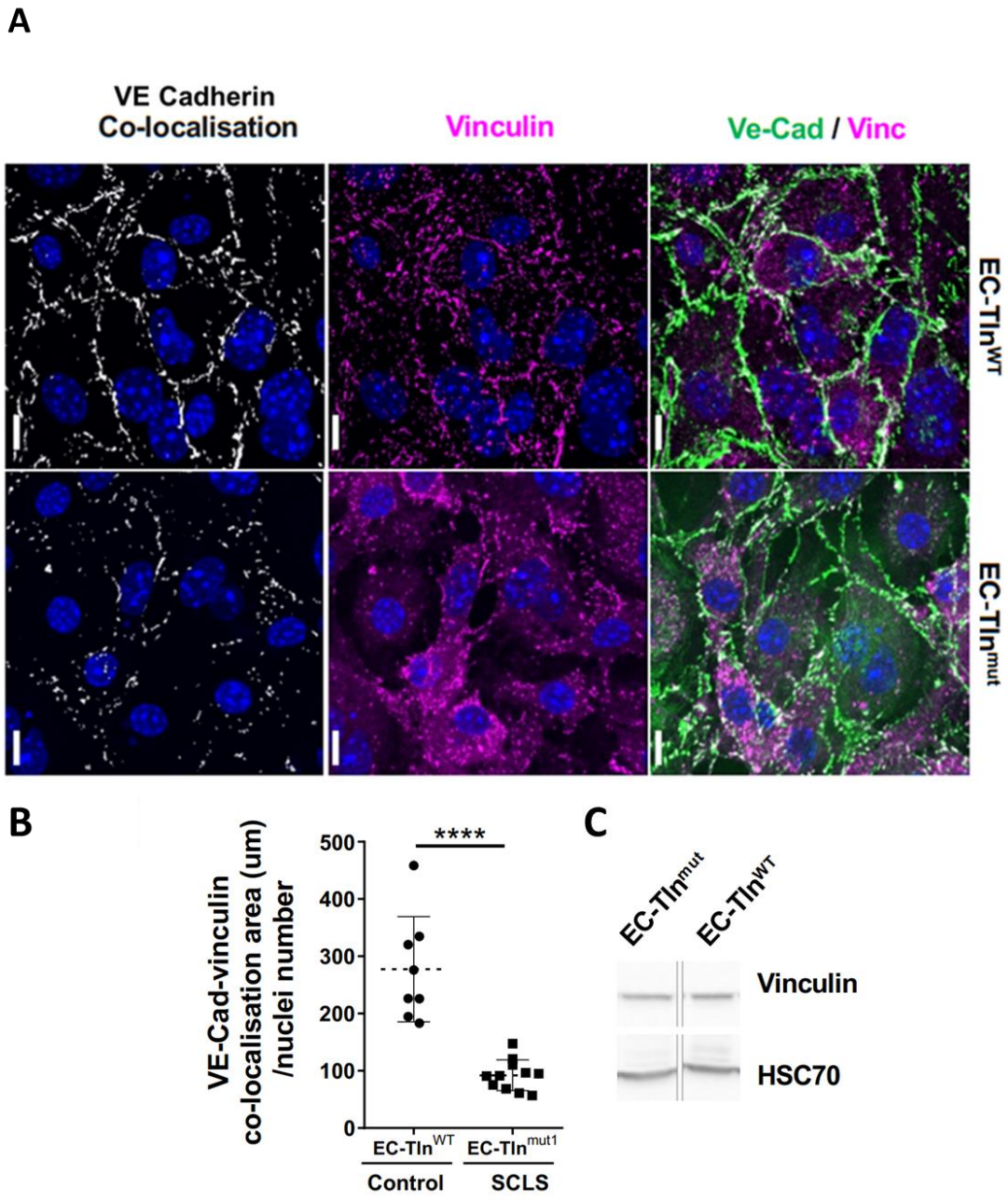


Figure 34:In vitro localisation of FA components in WT/SCLS expressing cells

Data from (Elefant *et al*., 2022)

A). Microscopy images produced by our collaborators of mouse primary endothelial cells (monolayers) produced by G. Rouni, transfected to express either WT talin 1(top row) or talin 1 R2510A, a mutation mimicking the exon skipping effects of the SCLS mutation (bottom row). Vinculin (magenta) and VE-cadherin junctions (green) were immunostained and nuclei (blue) stained with DAPI. B). western blot analysis of the total area (µm^2) of co-localisation of vinculin and VE-cadherin proteins per cell nuclei. The data shows that in cells transfected with the SCLS mimicking mutation, the area of co-localisation is significantly reduced. C). SDS-PAGE gel showing the total concentration of expressed vinculin within cells transfected with the WT and SCLS talin variants (HSC70 was used a control), gel data shows expression of vinculin was similar in both WT and SCLS expressing cells.

3.2.10 Over-active talin outcompetes cadherin junctions for stabilising vinculin

The cell data supports my own experimental data and reinforces that in SCLS patients the less stable R13 region alters the mechanical properties of the protein, resulting in lower mechanical force required to expose VBS than in the WT. This over-active talin causes vinculin to localise at focal adhesions where the SCLS talin is expressed rather than at other junctions where it is needed.

Adherens junctions (AJs), which contain VE-cadherin, are responsible for cell-cell adhesion and cellular remodelling (Huveneers *et al.*, 2012). AJs also require vinculin to form mature and stable adhesion complexes. The loss of vinculin at adherens/cadherin junctions results in decreased cell-cell adhesion (Bays and DeMali, 2017). Therefore, in WT cells both cell-cell (cadherin junctions) and cell-matrix (FAs) adhesions compete for vinculin.

If the talin in the focal adhesions unfolds more easily (as seen with SCLS talin), then the vinculin is more likely to localise at the FAs instead of the AJs, 'leeching' the vinculin away from AJs. This would result in less mature and less overall AJs.

3.2.11 R13-SCLS causes weakened adherens junctions leading to capillary leakage

It can be proposed that the lower stability talin 1 protein found intermittently expressed in patients suffering from SCLS competes with cadherin proteins for the formation/stabilisation of their respective cellular junctions (FA and AJ) by competing for the available vinculin. AJs, of which cadherins are a part, have been shown, like FAs, to be highly dynamic and shown to remodel during tissue strain and movement (Baum and Georgiou, 2011). In a WT cell, when stress is put on the vascular endothelium such as during the consumption of alcohol or when suffering from dehydration (both commonly reported

by patients prior to onset of symptoms), the force on vascular cell-cell junctions increases. In WT cells the AJs remodel within the cell along specific regions of the membrane to reinforce the vascular tissue. However, this rearrangement of AJs, like FAs, requires vinculin to bind actin fibres to adhesion proteins (α/β catenins) and stabilise maturing AJs. In episodes of SCLS the SCLS talin variant is expressed, this limits vinculin due to over-localisation at FAs, the lack of vinculin prevents vascular tissue from remodelling correctly (as shown in fig.35) and thereby destabilising the cell-cell bond in the vascular endothelium. This destabilisation leads to the breakdown on the cell-cell tight junctions causing the fluid contained within the capillaries to leak out into the surrounding tissue, resulting in the symptoms observed in patients of SCLS.

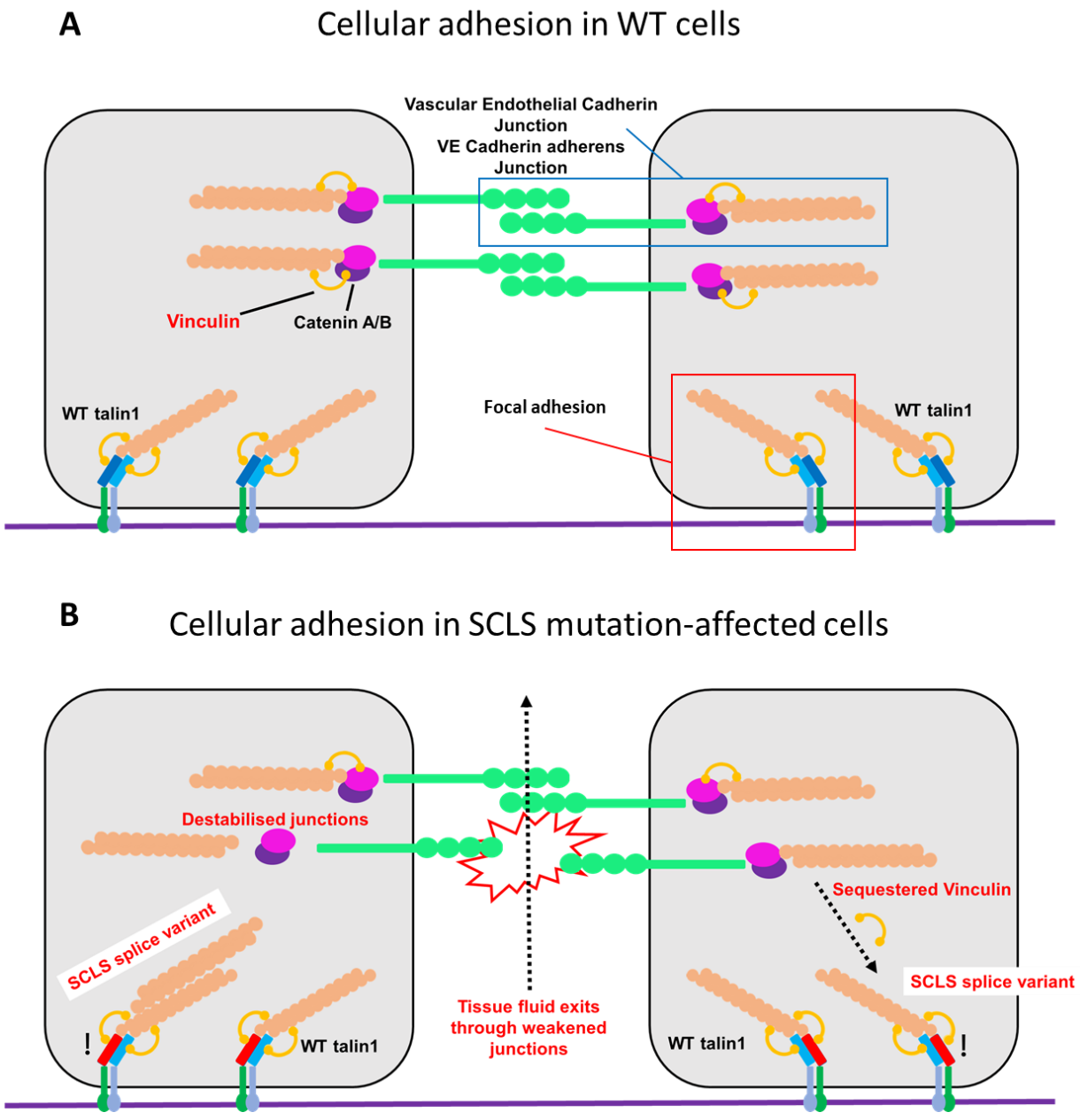


Figure 35: Predicted effect of exon 54 skipping in FA and adherens junction stability

Cartoon showing the predicted effect of the SCLS mutation on the formation and survival of cadherin junctions and focal adhesions. A). Cells expressing WT talin (blue), with expressed vinculin (orange) recruited in normal fashion to both VE-cadherin junctions (highlighted in blue) and focal adhesions (highlighted in red). B). Cells expressing SCLS talin 1. Dashed arrows mark where vinculin is sequestered from cadherin junctions by overly active SCLS talin 1 vinculin binding sites due to lower force limits required for unfolding of R13, as well as the lack of vinculin resulting in destabilised VE-cadherin junctions, leading to tissue fluid leakage.

Chapter 3.3 intrabodies

3.3.4 The process of intrabody evolution

A wide selection of highly varied possible intrabodies are produced as the residues of the concave surface region of the fibronectin are allowed to mutate. To separate non-binding proteins from functional intrabodies a type of phage display is used. Magnetic beads coated in streptavidin (a bacterial protein purified from *Streptomyces avidinii* bacteria) are used in the binding assay. The intended intrabody target, which in the case of this project was R1R2 and R11 regions of talin 1, is biotinylated (via exposed lysine residues) using biotin ligase (Go *et al.*, 2021). The attached biotin forms an incredibly strong bond with streptavidin (Liu, Zhang and Mei, 2016), anchoring the target protein to the bead.

The possible candidate intrabodies are then incubated with these tagged beads, those that do not bind are washed off and precluded from further testing (fig.36). Those that do show interaction are considered viable, proteins displaying binding sequences are then re-

integrated back into the pool and allowed to 'evolve', and are run through this phage display multiple times, each time selecting the most successful binders for the next round.

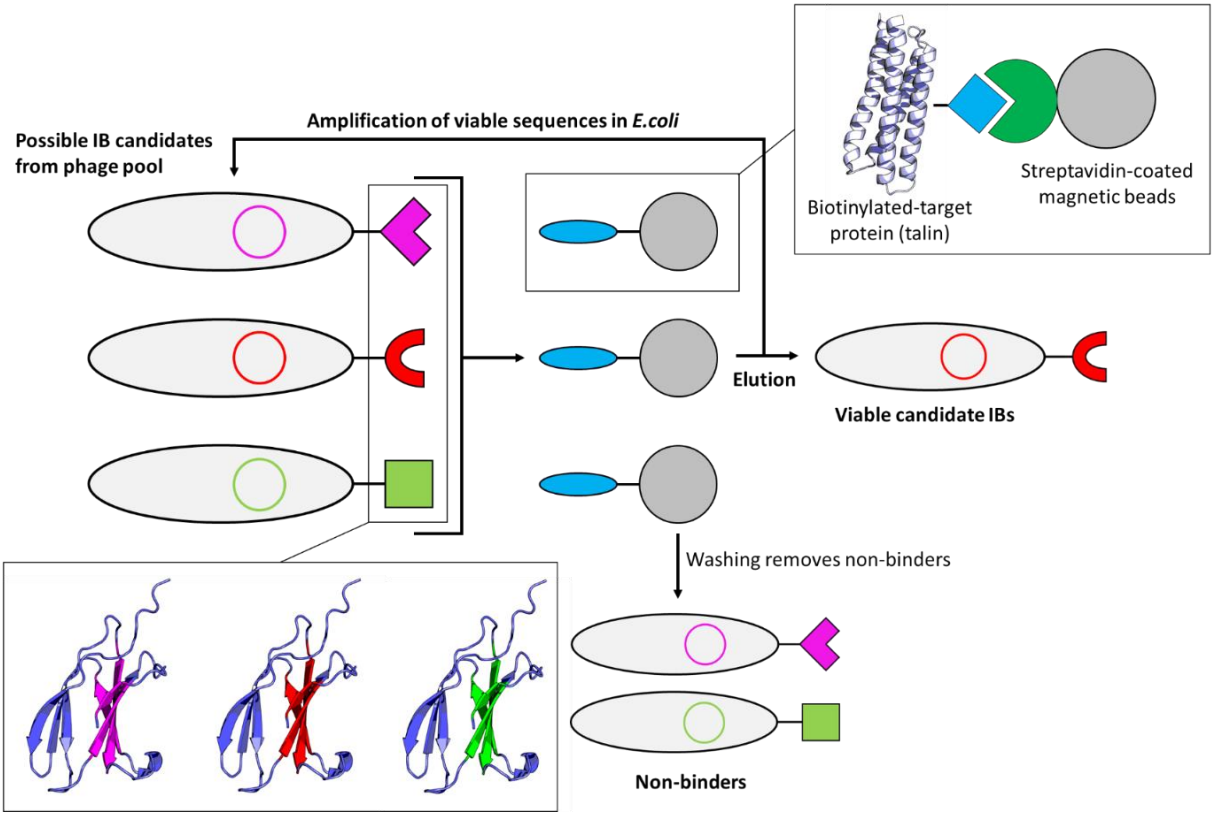


Figure 36: Phage display process

The process of phage display used to test the viability of intrabody candidates. The fibronectin domains from three possible bacterial candidates (shown in magenta, red and green) are spliced into the filamentous bacteriophage gene III, being expressed on the surface of the phage. These member of the 'phage library' are suspended with paramagnetic beads with biotin-streptavidin anchored protein target (talin domain in blue). Candidates capable of binding are noted and their sequences are amplified and allowed to alternate to increase binding capacity. Unbound intrabody candidates are washed out and precluded from further testing.

For this project I was given 3 of these intrabodies, XPH238,241 and 242, all of which had been specialised through phage display. The sequence of each of these bacterial proteins could be expressed in BL21 competent cells, and their binding properties examined. I aimed to confirm the results of the phage ELISA run by the Bordeaux team and validate their findings, as well as show the functionality and viability of these proteins as a potential

competitor to the monoclonal antibody, which involved characterising the epitope of each of the 3 intrabodies.

3.3.5 Intrabody functionality

The plasmids from the Bordeaux group, were transformed into BL21 competent *E. coli* cells and grown and purified following standard protocols. Each plasmid was designed (following the procedure outlined in the introduction) to bind to a specific region of talin 1 (XPH238 to R1R2 and XPH241/242 to R11). However, the intrabodies were found by the Bordeaux team to also have weak interactions with other regions of talin (e.g. R8 which was found to interact weakly with all 3 intrabodies). Once all 3 intrabodies had been purified, the first stage was to confirm the intrabodies bound to their intended regions of talin 1. This was done using NMR.

^{15}N labelled talin 1 R1R2 and R11 were produced following standard protein expression and purification protocols in minimal media. A ^{15}N TROSY HSQC experiment was run with each ^{15}N talin 1 protein at 100 µM concentration (with 2 mM DTT) to establish a baseline spectrum for each. As I did not know for certain that there was only 1 intrabody binding site within the target talin region, so a 1:3 talin to intrabody ratio was used (100 µM talin to 300 µM intrabody) in hope that this would saturate any available binding sites.

3.3.6 Validation of XPH 238 binding talin 1 R2

The first intrabody examined was XPH238, which was evolved to bind to R1R2 on talin 1. As shown in fig.37A, a ^{15}N TROSY HSQC experiment using a 1:3 ratio of talin to XPH238 showed a contrast between the R1R2 baseline (fig.37A grey spectra) compared to R1R2 in the presence of XPH238 (fig.37A Blue spectra), many peaks had shifted and some peaks were

lost completely. The chemical shift indicates the intrabody does interact with the R1R2 region, which agreed with the Bordeaux teams' data showing that XPH238 bound R1R2. However, to map the XPH238 intrabodies epitope I needed to know if the protein bound to both, or a single helical domain within R1R2. ^{15}N talin 1 domains R1 and R2 were individually expressed and purified. Samples containing a 1:3 ratio of talin to XPH238 were prepared and the experiment was repeated separately with these two proteins, R1, fig.37Ai, showed almost no chemical shift between the R1 baseline (fig.37Ai, grey) and R1 + XPH238 (fig.37Ai, green). R2 showed significant shift between the baseline spectra (fig.37Aii, grey) R2 + XPH238 (fig.37Aii, purple). This confirmed that XPH238 had developed a binding site for the R2 region of R1R2 rather than binding both domains.

3.3.7 Validation of XPH 241 binding talin 1 R11

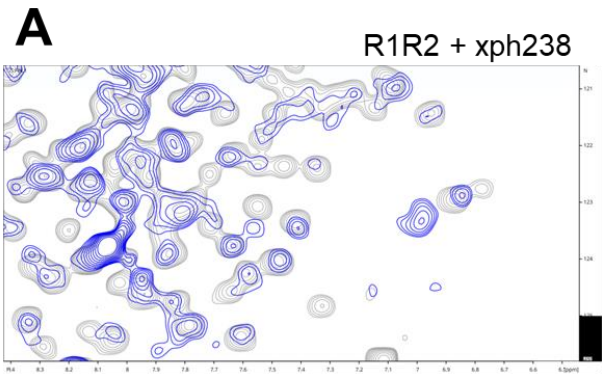
The XPH241 intrabody had been evolved to target the R11 domain of talin 1. To examine this interaction, ^{15}N labelled R11 was produced and run as a TROSY HSQC experiment, the resulting baseline spectra is shown in fig.37B (grey spectra). ^{15}N R11 was then run with a 1:3 ratio of XPH241 intrabody. The resulting differences between the baseline (fig.37B, grey) and the 300 μM intrabody sample (fig.37B, teal) was stark. With the majority of the peaks seen in the baseline spectra disappearing, with only a few peaks remaining unaffected. The loss of so many peaks confirmed that the intrabody did indeed interact/bind R11, however due to the loss of peaks, it was not possible to determine an epitope from this data. One possible reason for the loss of peaks is the known self-assembling property of R11 to polymerise into long chains of unfolded helical bundles. It is possible the interaction of XPH241 causes the unfolding of the R11 protein, resulting in polymer chains.

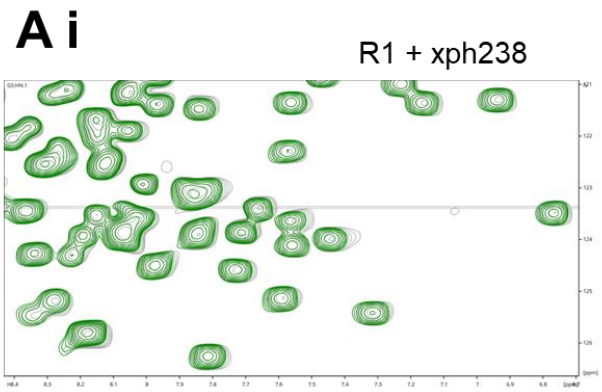
To counteract this issue, and allow identification of the epitope, a mutant of talin 1 R11 was used. R11 V2078K, is a mutant produced by the Goult lab previously in the study of R11 self-assembly. This form of mutant R11 produces a helical bundle that is unable to unfold, removing the risk of polymerisation. The baseline spectra of R11 V2078K is shown in fig.37Bi, grey spectra. When the R11 V2078K mutant was run in a 1:3 ratio XPH241, the resulting spectra (fig.37Bi, red spectra) showed more peaks than were seen in the WT R11. Most peaks were broadened, and many had still disappeared entirely, which showed that the XPH241 intrabody was still able to interact with the R11 V2078K mutant. However, compared to the WT R11, the resulting spectra of R11 V2078K and XPH241 contained enough visibly shifted peaks to predict an epitope for the intrabody.

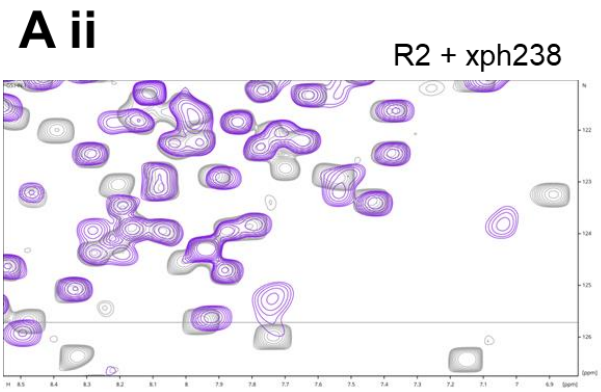
3.3.8 Validation of XPH 242 binding talin 1 R11

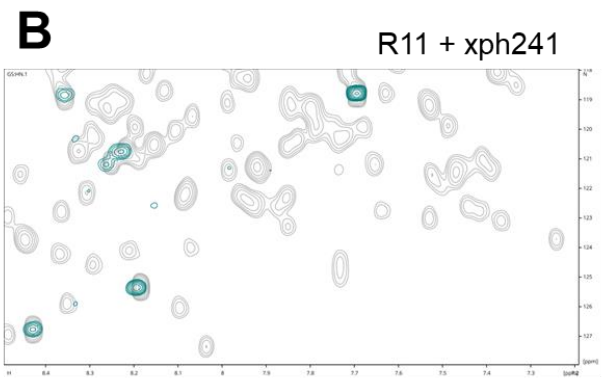
To confirm the interaction between XPH242 and its target, the same ^{15}N labelled R11 was used as both intrabodies share the same intended target. As with the other proteins R11 was run alone in a ^{15}N TROSY HSQC experiment to acquire a baseline (fig.37C, grey spectra). The experiment was then repeated with a 1:3 ratio of talin to XPH242, which is shown in fig.37C, magenta spectra. The addition of XPH242 causes significant chemical shift, with few peaks lost entirely, confirming that the XPH242 does bind its intended target. The type of shift seen also makes it easier to identify an epitope. Although many peaks were seen to have shifted, some had shifted quite significantly making it difficult to identify which peaks were their new positions. To alleviate the issue the TROSY HSQC experiment in fig.37C was repeated with a 1:0.5 ratio of talin to XPH242 instead of a 1:3 ratio. This experiment was run to generate 'intermediate peaks' between the initial and final positions of each peak. The resulting spectra fig.37Ci, can be seen in blue, overlayed on the baseline and 1:3 ratio

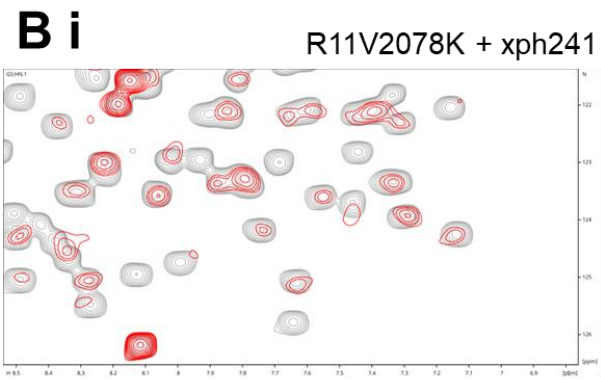
spectra. Instead of seeing a peak in between the initial and final positions, the intensity was split across both positions or had not moved, with a peak at both the initial and final position. This confirmed that the method of interaction was ‘slow exchange’ (shown in fig.38). It also aided in identifying the XPH242 epitope.

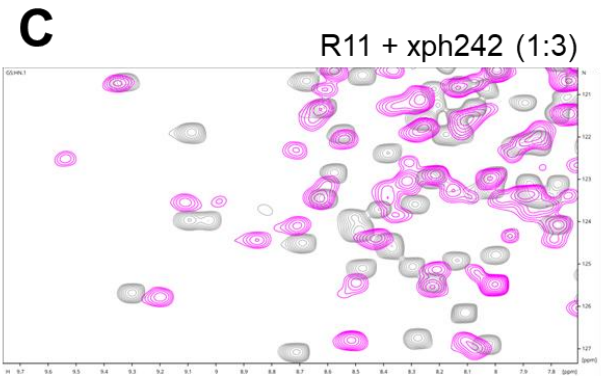

A
R1R2 + xph238


A i
R1 + xph238


A ii
R2 + xph238


B
R11 + xph241


B i
R11V2078K + xph241


C
R11 + xph242 (1:3)

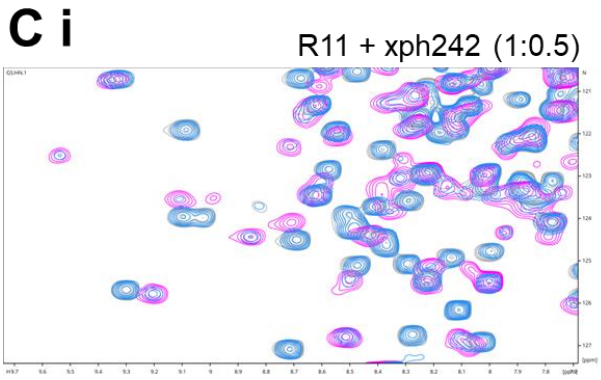

C i
R11 + xph242 (1:0.5)

Figure 37:NMR spectra comparing binding ability of intrabodies XPH238,421,242

Nuclear magnetic resonance data examining binding capability of intrabodies XPH238,241 and 242 for their intended targets (talin1 R1R2 and R11). All samples contained 300 µM intrabody and 100 µM talin in S column affinity buffer A with 1 mM DTT and 5% V/V to allow locking on of the sample, all experiments conducted were ^{15}N TROSY HSQC (2D). A). Overlayed NMR spectra showing interaction between XPH238 and talin1 R1R2, R1R2 alone (grey) and R1R2 1:3 ratio with XPH238 (blue), chemical shift indicates interaction, further examination was done with R1 and R2 separately. Ai). R1 alone (grey) and R1 1:3 ratio with XPH238 (green), few peaks were shown to shift in this spectra. Aii). R2 alone (grey) and R2 1:3 ratio with XPH 238 (purple), a significant number of peaks underwent chemical shift indicates R2 contains the main binding site for XPH238. B). examination of interaction between XPH241 and intended target talin 1 R11, R11 alone (grey) and R11 in a 1:3 ratio with XPH241 (cyan) shows high peak loss indicating interaction but binding site unidentifiable. Bi). Repeat of experiments shown in B, using variant R11 V2078K mutant, unable to unfold. R11 V2078K alone (grey) and R11 V2078K in a 1:3 ratio with XPH241 (red), chemical shift and peak loss observed showing interaction still occurs with R11 V2078K. C). examination of XPH242 binding to talin 1 R11. R11 alone (grey) and R11 in a 1:3 ratio with XPH242 (magenta), chemical shift indicates interaction, but high shifting makes identifying binding site difficult. Ci). R11 alone (grey), in a 1:3 ratio (magenta) and a 1:0.5 ratio with XPH242 (blue), chemical shift observed but usage of both spectra allows binding site identification.

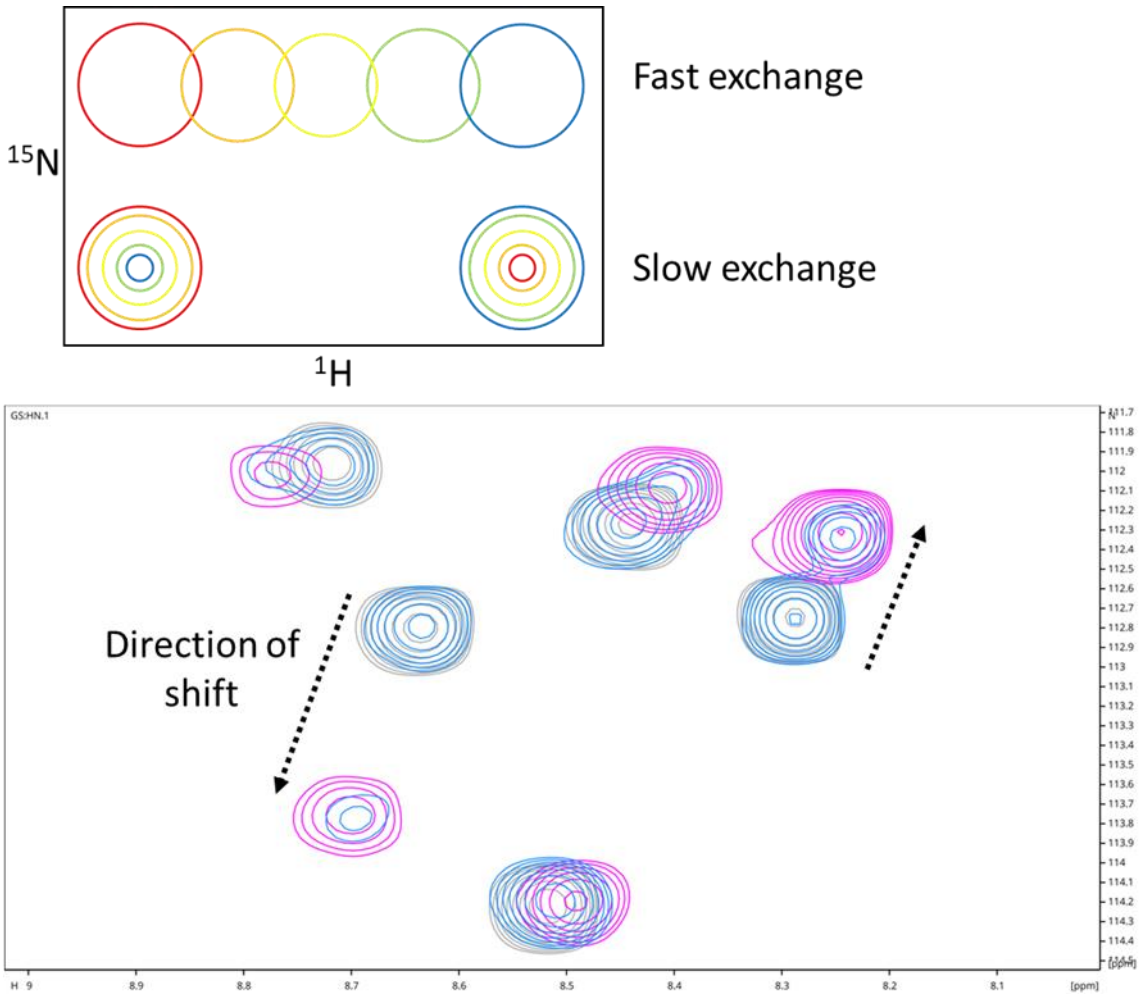


Figure 38: XPH242 binding and identification of slow exchange interaction

Conformation of slow exchange interaction between R11 and XPH242. (top) image showing slow and fast exchange types and how they are observed in NMR spectra, either initial and final shift positions simultaneously (slow) or intermediate peaks (fast). (bottom) close-up overlayed spectra of R11 interaction with XPH242, R11 (grey), 1:3 ratio (magenta) and 1:0.5 ratio (blue), both an initial and final peak can be seen in the 1:0.5 ratio spectra as the binding site on R11 is not fully satisfied by intrabody concentration, the two visible peaks rather than intermediate peaks indicate slow exchange type.

3.3.9 Mapping of epitopes

Once conformation of the intrabodies function was complete, the chemical shift data obtained from the 15N TROSY HSQC spectras was used to predict the epitopes of XPH238, 241 and 242. To do this, the positions of each peak in the baseline spectra was logged. These peak positions were then adjusted to align with their new positions in the talin +

intrabody spectra. The chemical shift value of the post-intrabody spectra (PI) was then subtracted from the value of the baseline spectra. This was done for both the N and H chemical shift values per peak. To counteract negative values, both values were squared then square rooted. These values were then put through the equation shown in table 3. The resulting chemical shift values were plotted on a bar graph displaying the amount of shift seen for each peak (e.g. >0.05, >0.075) values over 0.05 PPM were considered 'significant' and were represented on AlphaFold 2 models produced of the target proteins.

Table 2

Table showing equations used to convert raw 15N and 1H chemical shift value (PPM) (blue box) into a final value of how far each peak has shifted (AU)

Baseline spectra 15N value (**BN**) 1H value (**BH**)	PI spectra 15N value (**PIN**) 1H value (**PIH**)	BN-PIN= **XN** BH-PIH= **XH**	SQRT(**XN**^2) SQRT(**XH**^2)	SQRT(((**XN**/6.5)^2)+((**XH**)^2))

3.3.10 Mapping the epitope of XPH238

The chemical shift of peaks within R2 caused by the binding of XPH238 are shown in fig.40, the assignment for this region of talin, as well as the other proposed epitopes for the antibodies, come from (Goult *et al.*, 2013c)(Banno *et al.*, 2012).Many of the R2 peaks exceeded the shift value needed to be considered significant (0.05) and overall displayed the highest shift values in comparison to the other intrabodies. This showed that XPH238 interacted strongly with its talin target.

All significant values were then ranked (this was arbitrary and was done based on the maximum values observed); >0.05, >0.2 and >0.3. Each of these was given a colour, green for >0.05, blue for >0.2 and red for >0.3. To visualise these shifts (fig.40A), the crystal structure of talin 1 R1R2 (6R9t) was used. Shown in fig.40B, the residues corresponding to

the shifted peaks were coloured green, blue or red according to their chemical shift value (with values below the significant threshold coloured grey). A clear concentration of high chemical shift can be seen in the lower region of the R2 domain where it meets R1, with the majority >0.03 shifted peaks present here.

To allow the modelling of the XPH238/R1R2 interaction, an AlphaFold2 model of XPH238 was generated. The prediction data for the model is shown below in fig.39. The pLDDT data plot (fig. 39 A) shows with the exceptions of the end regions of the model (pLDDT <50) the majority of the structure shows a high pLDDT, and therefore, a high confidence value. These pLDDT values are mapped on the AlphaFold2 model in fig. 39 B. The PAE scores for the XPH238 model are shown in fig. 39 C, which compounds the data shown in fig. 39 A, that the model has low positional error and can be considered accurate.

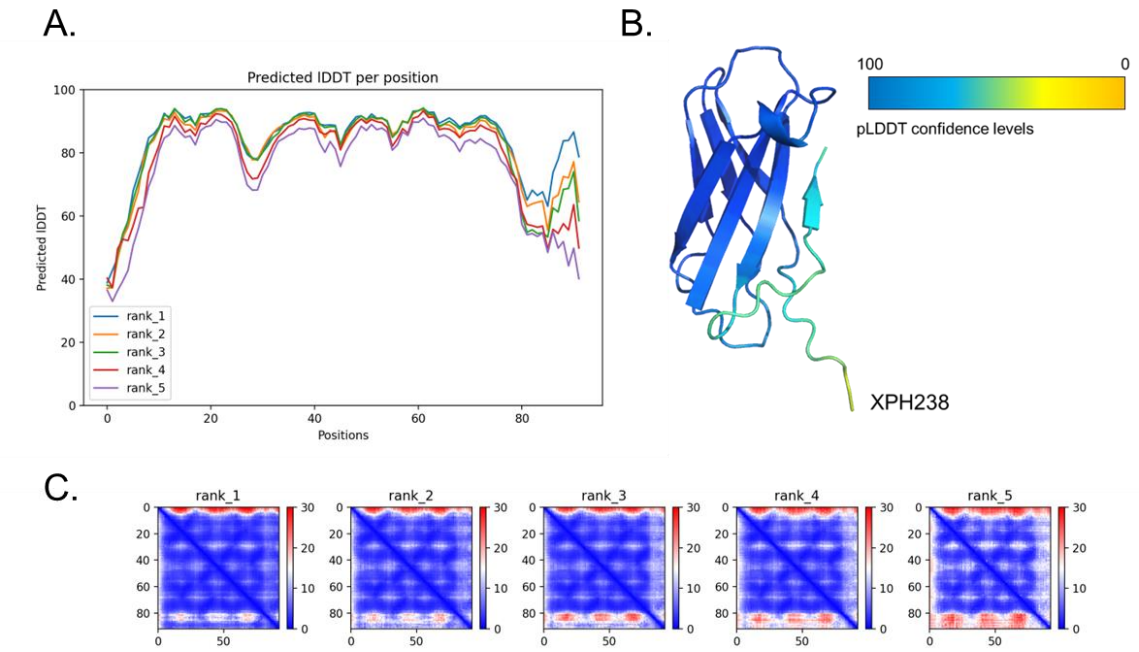


Figure 39: Reliability of AlphaFold2 XPH238 prediction

A,C). predicted Local Distance Difference Test (plDDT) data and Predicted Aligned Error (PAE) data provided by AlphaFold2 on the accuracy of the XPH238 generated structure. B). The pLDDT scores are shown mapped onto their respective residues on the model for XPH238, with higher pLDDT scoring residues coloured blue and lower scoring residues coloured yellow/orange.

The AlphaFold 2 model of XPH 238 was positioned with its hyper-variable concave side facing the approximate binding site. Fig.40C shows this modelled interaction between talin (grey) and intrabody (magenta) with both a cartoon (left) and a surface (right) representation. This figure was generated by manually docking the AlphaFold 2 model of XPH238 and with the crystal structure of talin R1R2. The concave region of the predictive model of XPH238 fits well to the predicted epitope, strongly supporting this as the binding site of XPH238.

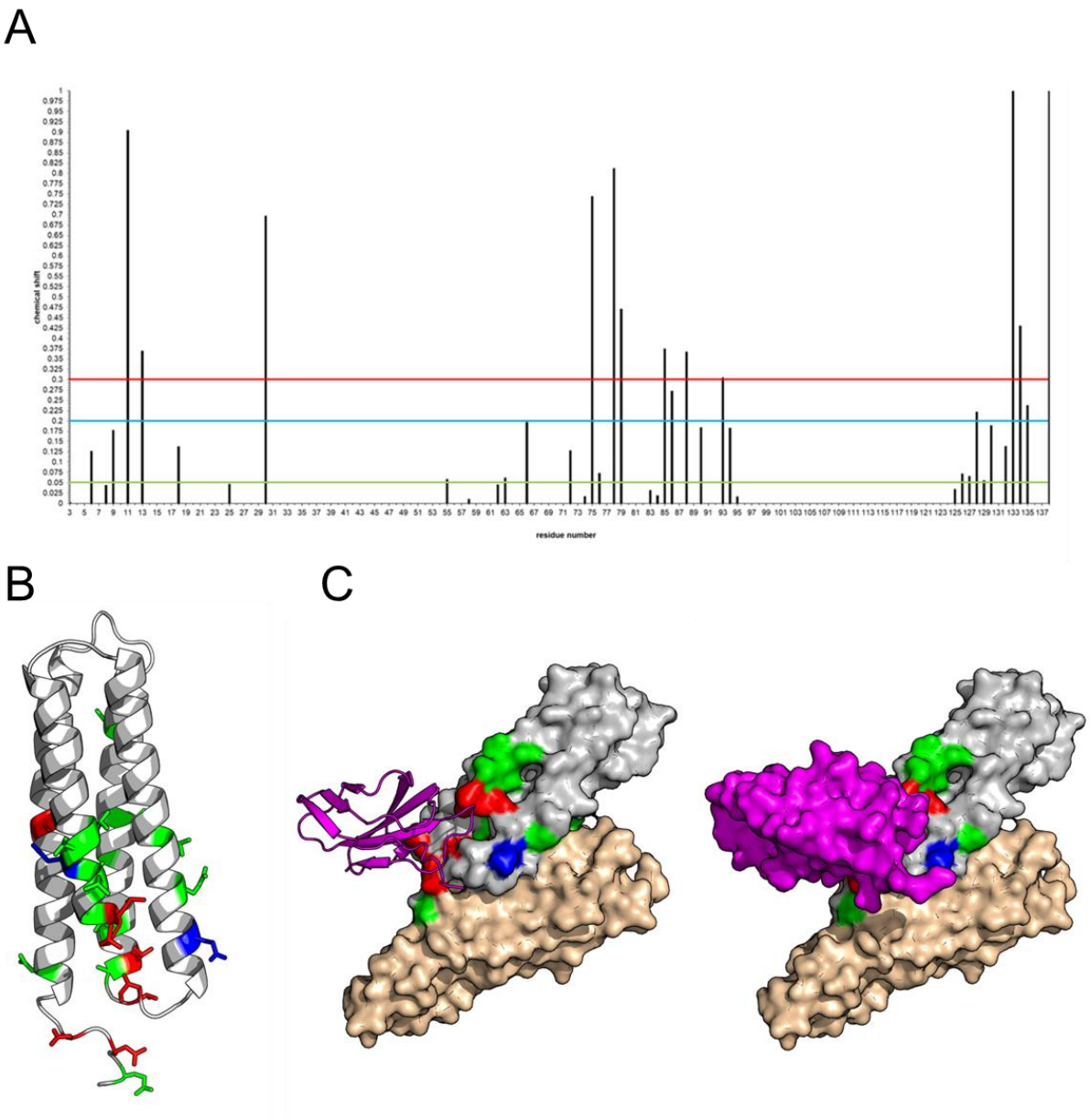


Figure 40: Analysis of XPH238 binding and identification of R2 binding site

Results of chemical shift analysis of interaction between talin 1 R1R2 and XPH238 intrabody. A). Plot showing the chemical shift (AU) of each residue with R1R2, with values over 0.05 (green line) considered significant. B). Significant chemical shift values mapped onto PyMol model of R1R2, with values >0.05 (green), >0.2 (blue) and >0.3(red) coloured to show likely binding site for XPH238 based on areas of high chemical shift. C). PyMol models showing R1 (grey) and R2(beige) interacting with XPH238 (magenta).

3.3.11 Mapping the epitope of XPH241

The epitope of XPH 241 was more difficult to locate with certainty due to the volume of peaks that disappeared, even the replacement of R11 with R11 V2078K did not leave many peaks that had clearly shifted. Peaks that disappeared completely were not considered to be significant as they added too much uncertainty, as peak loss could be the result of lower sample quality rather than true binding-induced chemical shift.

As with XPH238, an AlphaFold2 model of XPH241 was generated to allow the modelled docking of XPH241 to R11. Shown in fig. 41 is the AlphaFold2 generated model for XPH241 and the confidence values for its structure. The pLDDT values shown in fig. 41 A, show that a significant amount of the sturcture, including the C and N terminal regions and a portion of the β-barrel have pLDDT scores of ~50 and below. This implies that the structure of this region could be inaccurate. The areas of low pLDDT confidence are also seen to have higher positional error (fig. 41 C). As a result of this data, the docking model of XPh241 and R11 should be considered less accurate than that of XPH238.

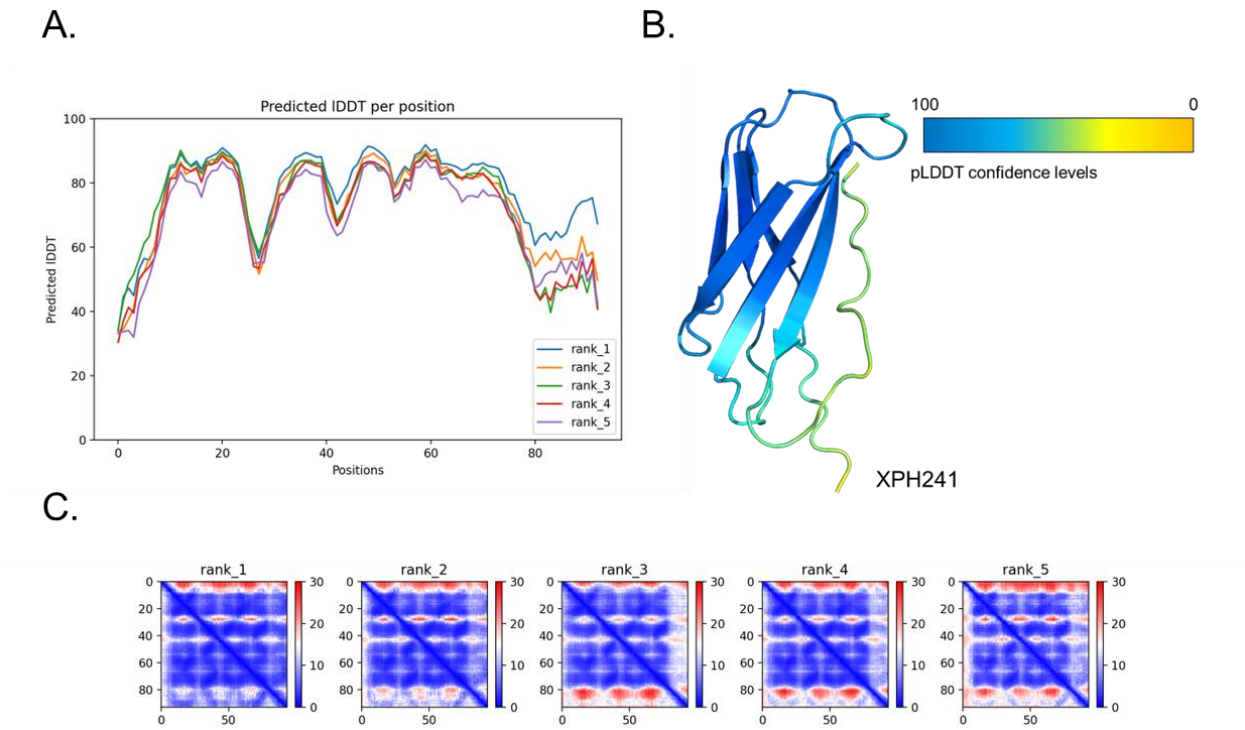


Figure 41: Reliability of AlphaFold2 XPH241 prediction

A,C). predicted Local Distance Difference Test (plDDT) data and Predicted Aligned Error (PAE) data provided by AlphaFold2 on the accuracy of the XPH241 generated structure. B). The pLDDT scores are shown mapped onto their respective residues on the model for XPH241, with higher pLDDT scoring residues coloured blue and lower scoring residues coloured yellow/orange.

The peaks that shifted significantly are shown in fig.42A. When mapped onto a cartoon representation of R11 and coloured according to shift values (Fig.42B), it can be seen that the majority of the shifts (even though low in intensity) were on a single face of the helical bundle. An AlphaFold 2 model of XPH241 generated and docked manually (in AlphaFold 2) to the predicted epitope (fig.42C). The variable concave surface fitted well around the curve of the helical bundle face suggesting a correct epitope prediction. However, due to the low chemical shift values, without a crystal structure conformation of the XPH241 epitope is difficult.

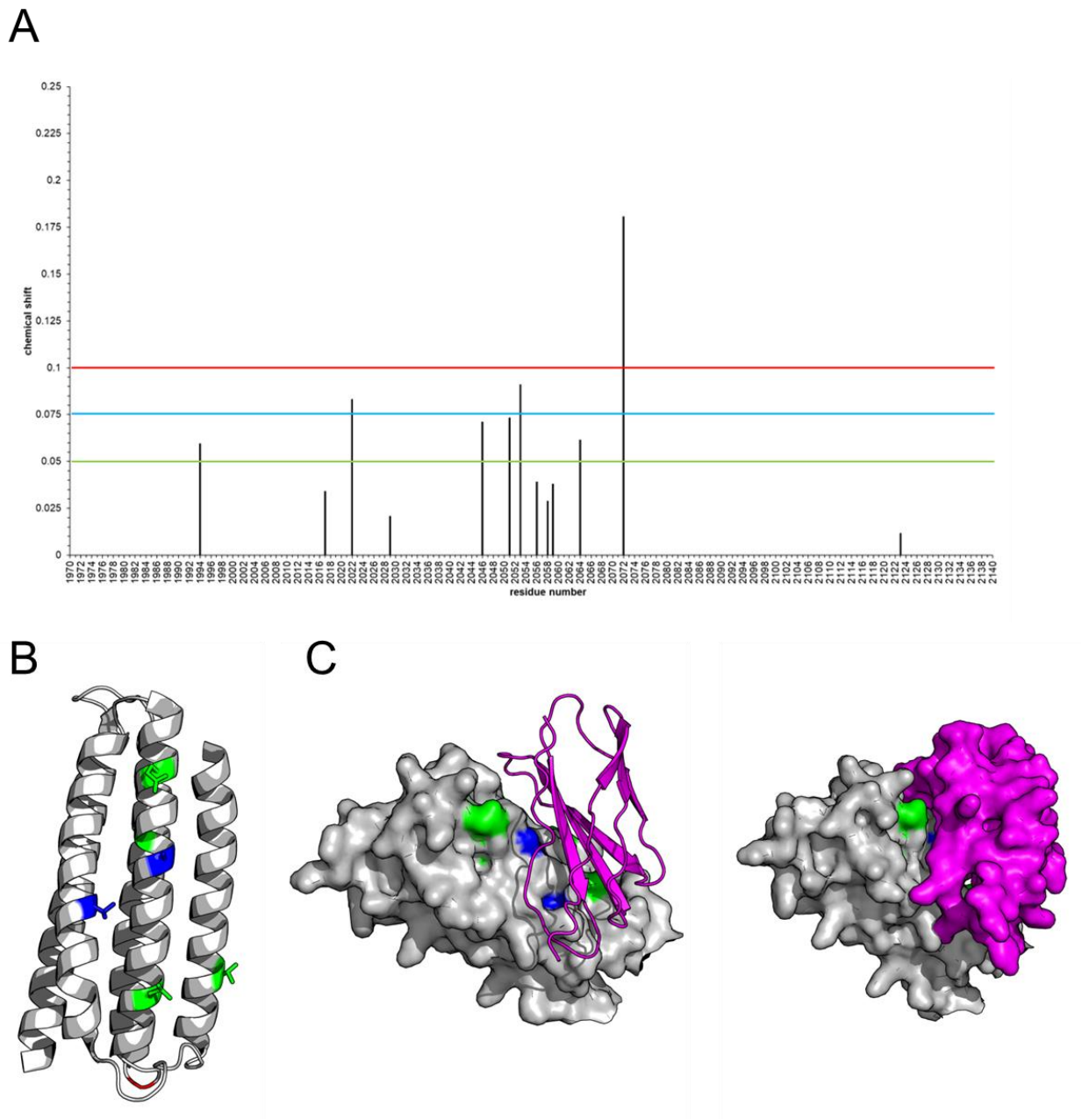


Figure 42: Analysis of XPH241 binding and identification of R11 binding site

Results of chemical shift analysis of interaction between talin 1 R11 and XPH241 intrabody. A). plot showing the chemical shift (AU) of each residue with R11, with values over 0.05 (green line) considered significant. B). significant chemical shift values mapped onto PyMol model of R11, with values >0.05 (green), >0.075 (blue) and >0.1(red) coloured to show likely binding site for XPH238 based on areas of high chemical shift. C). PyMol models showing R11 (grey) interacting with XPH241 (magenta).

3.3.12 Mapping the epitope of XPH242

In comparison to locating the epitope of XPH241, the spectra of XPH242 was sharp and shifted peaks were easily identifiable with the help of the intermediate spectra. However, the number of peaks displaying high shift did make it more difficult to identify a clear epitope.

The AlphaFold model of XPH242 generated to model the interaction between it and R11 is shown below in fig. 43 along with the analysis of its accuracy. Fig. 43 A shows that the pLDDT score for most of the model is >60, therefore, with the exception of both the C and N terminals, the model canbe considered accurate. When the pLDDT is mapped onto the AlphaFold2 model (fig. 43 B), the area of low pLDDT scores sit outside the core β-barrel structure. The PAE, fig. 43 C scores agree with the pLDDT scores, showing low positional error across the central region of the XPH242 model.

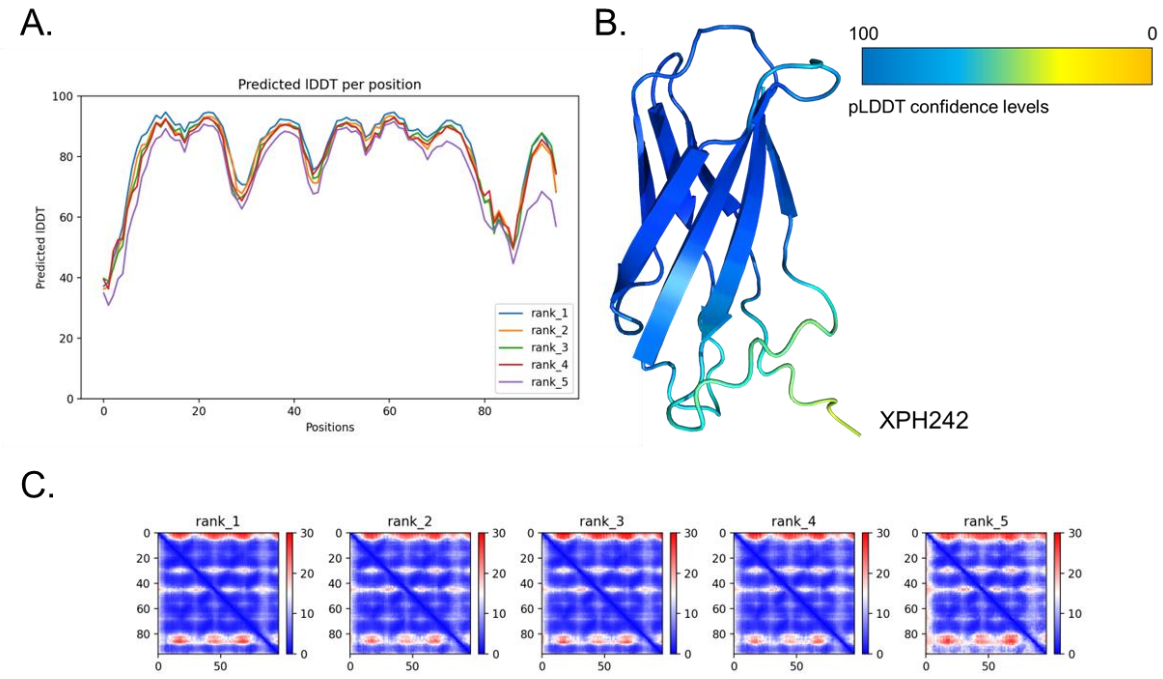


Figure 43: Reliability of AlphaFold2 XPH242 prediction

A,C). predicted Local Distance Difference Test (plDDT) data and Predicted Aligned Error (PAE) data provided by AlphaFold2 on the accuracy of the XPH242 generated structure. B). The pLDDT scores are shown mapped onto their respective residues on the model for XPH242, with higher pLDDT scoring residues coloured blue and lower scoring residues coloured yellow/orange.

The table of chemical shift for XPH242 is shown in fig.44A and are mapped onto the crystal structure of R11 (6R9T). Many high intensity shifts (>0.1) shown in red, are present at either ends of the helical bundle making identification of a binding face difficult. However, on closer inspection of the model, many of shifted peaks have side chains facing into the helical bundle, which would prevent them from interacting with an intrabody if that area was the binding face. This suggests that these residues have shifted due to the perturbation of R11 caused by a similarly sized intrabody binding the talin region. The residues with high chemical shift and outward facing side chains were concentrated towards one end of the helical bundle, as well as some on the flexible linker regions. An AlphaFold 2 model of XPH242 was generated and docked manually to the R11 surface representation as shown in fig.44C, it fits well onto the predicted epitope region and does not interfere with R12 either

(shown in all grey). This makes it highly likely that this is the binding site of the XPH242 intrabody.

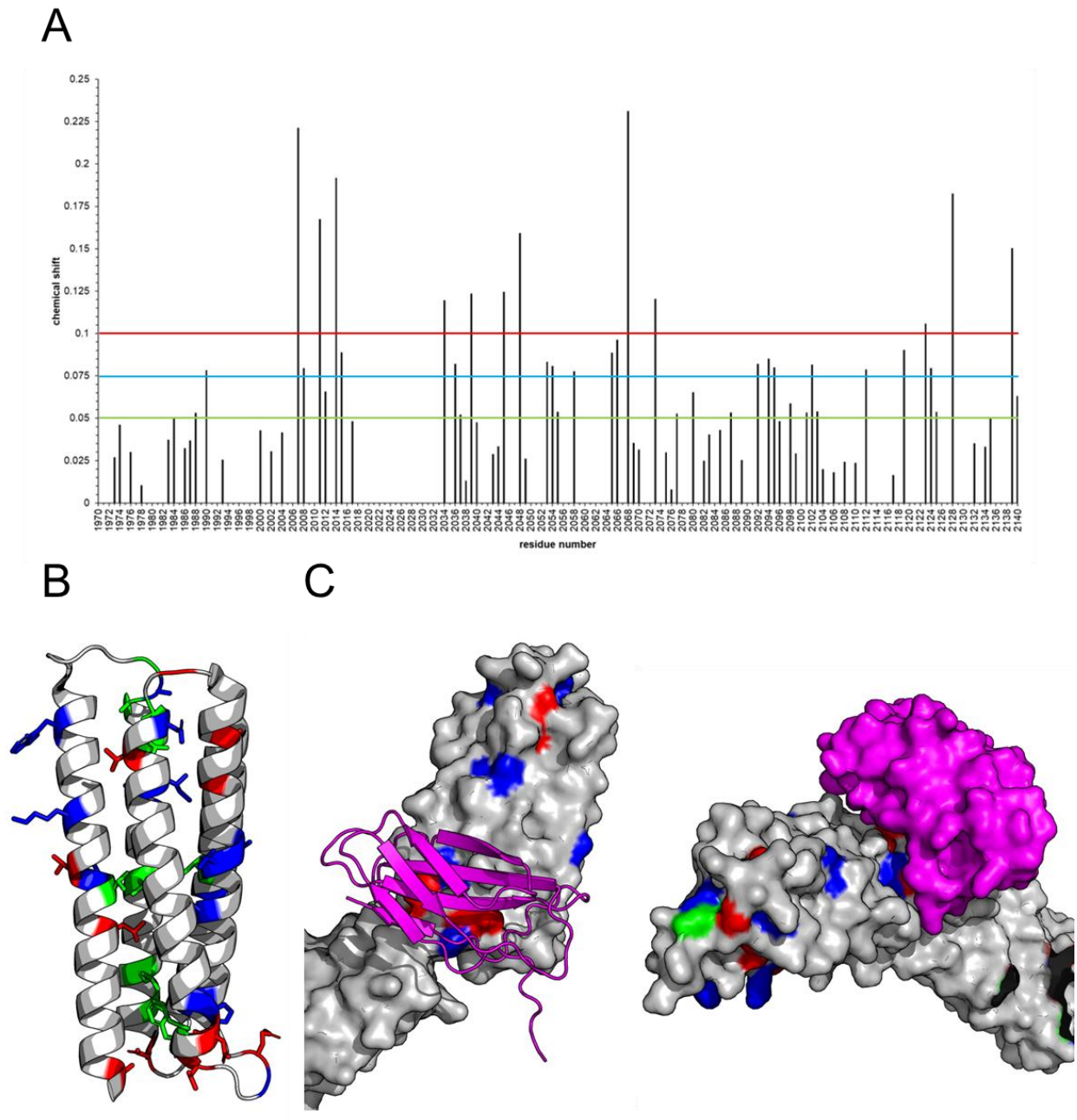


Figure 44: Analysis of XPH242 binding and identification of R11 binding site

Results of chemical shift analysis of interaction between talin 1 R11 and XPH242 intrabody. A). plot showing the chemical shift (AU) of each residue with R11, with values over 0.05 (green line) considered significant. B). significant chemical shift values mapped onto PyMol model of R11, with values >0.05 (green), >0.075 (blue) and >0.1(red) coloured to show likely binding site for XPH238 based on areas of high chemical shift. C). PyMol models showing R11 (grey) interacting with XPH242 (magenta).

3.3.13 Validating intrabody specificity

If intrabodies were to be used for any kind of analytical process, they would need to be highly specific. To confirm the specificity of the discussed intrabodies, further NMR data was gathered. To validate intrabody specificity, each intrabody was combined with a region of talin that it was not raised against. In the case of XPH238 (fig.45), XPH238 was combined with R11 in a 1:3 ratio of talin to XPH238. This sample was run in a ^{15}N labelled TROSY HSQC experiment, with the R11 baseline shown in grey and R11 + XPH238 shown in blue. When the resultant spectra were compared to R11 titrated with XPH241 (grey and red spectra), an intrabody raised against R11, a clear difference is visible. XPH241 causes (as shown previously) massive peak loss and shifting to such a degree as to not leave many peaks behind. In comparison, when XPH238 was added, the majority of R11 peaks do not show any shift, suggesting very low levels of interaction between XPH238 and R11.

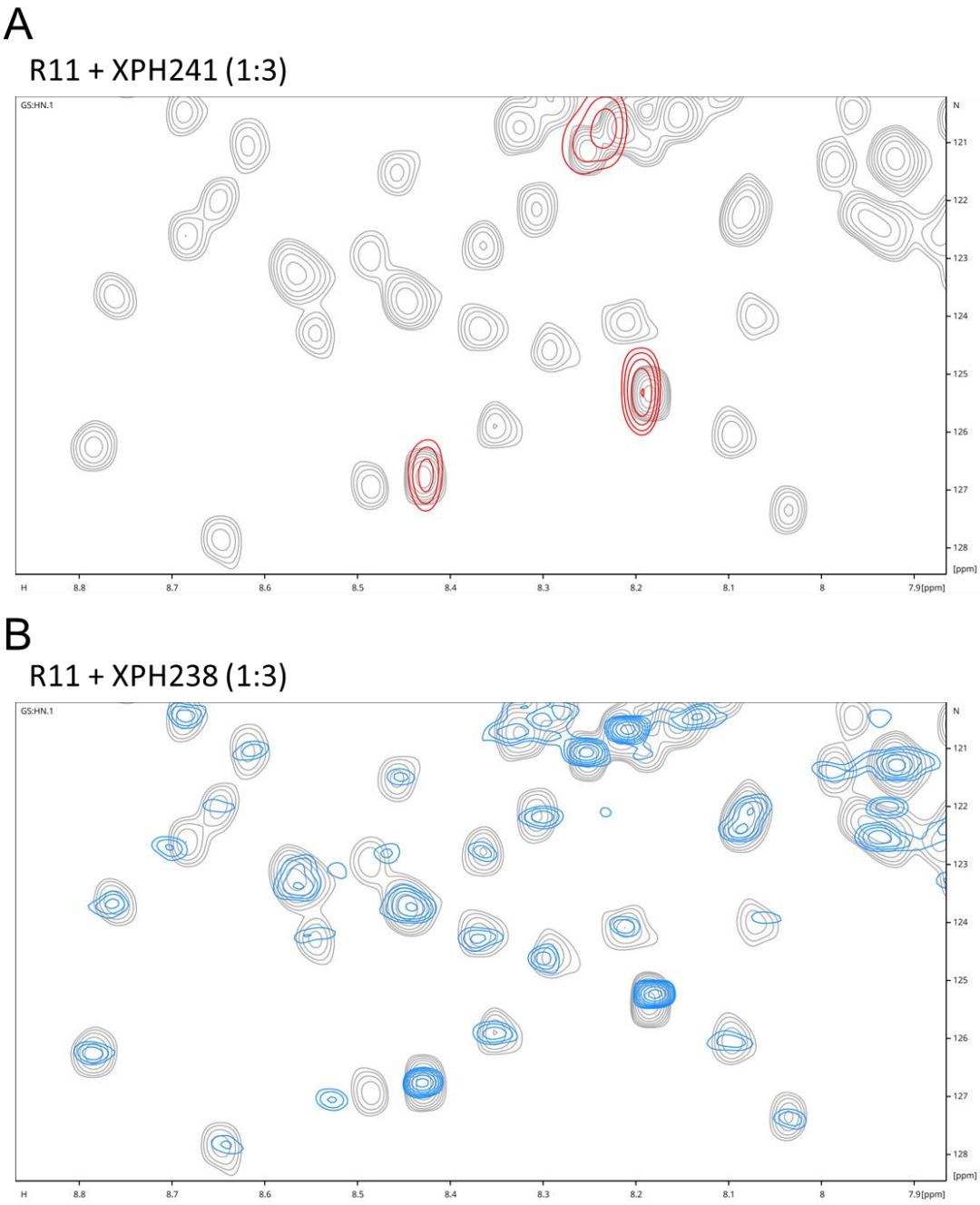


Figure 45: Intrabody specificity, inability to bind other talin domains

Nuclear magnetic resonance data examining the specificity of the XPH238 intrabody for talin regions other than the target R1R2 domains. All samples contained 300µM intrabody and 100µM talin in S column affinity buffer A with 1mM DTT and 5% V/V to allow locking on of the sample, all experiments conducted were 15N TROSY HSQC (2D). A). R11 alone (grey) and with an intrabody intended to bind R11, XPH241 (red), chemical shift shows intended interaction. B). R11 alone (grey) and in a 1:3 ratio with XPH238 (blue), this spectra shows significantly reduced chemical shift indicating lack of/reduced binding. Data shows XPH238 is highly specific to its intended target domain.

Having shown that XPH238 displayed high specificity for its intended talin region, it was decided to test if the XPH238 protein could bind variant versions of intended target region. There was a unique opportunity here as earlier research dealt with R1R2-17b which is an altered form of the intended target of XPH238 (R1R2). If the intrabody couldn't bind R1R2-17b it would provide a method of identifying the presence of such talin variants. A 1:3 ratio of ^{15}N labelled R1R2-17b and XPH238 was run in a ^{15}N TROSY HSQC NMR experiment shown in fig.46, with R1R2-17b baseline in grey and R1R2-17b + XPH238 in green. Very few peaks were seen to have disappeared or shifted in the presence of XPH238, suggesting that XPH238 couldn't strongly interact with R1R2-17b. When compared to the R1R2-WT interacting with XPH238, fig.46, grey spectra (R1R2-WT baseline) and fig.46, purple spectra (R1R2-WT + XPH238), the level of shift seen in R1R2-17b is negligible, suggesting that XPH238 is specific enough to differentiate between two splice variants of the same talin domain.

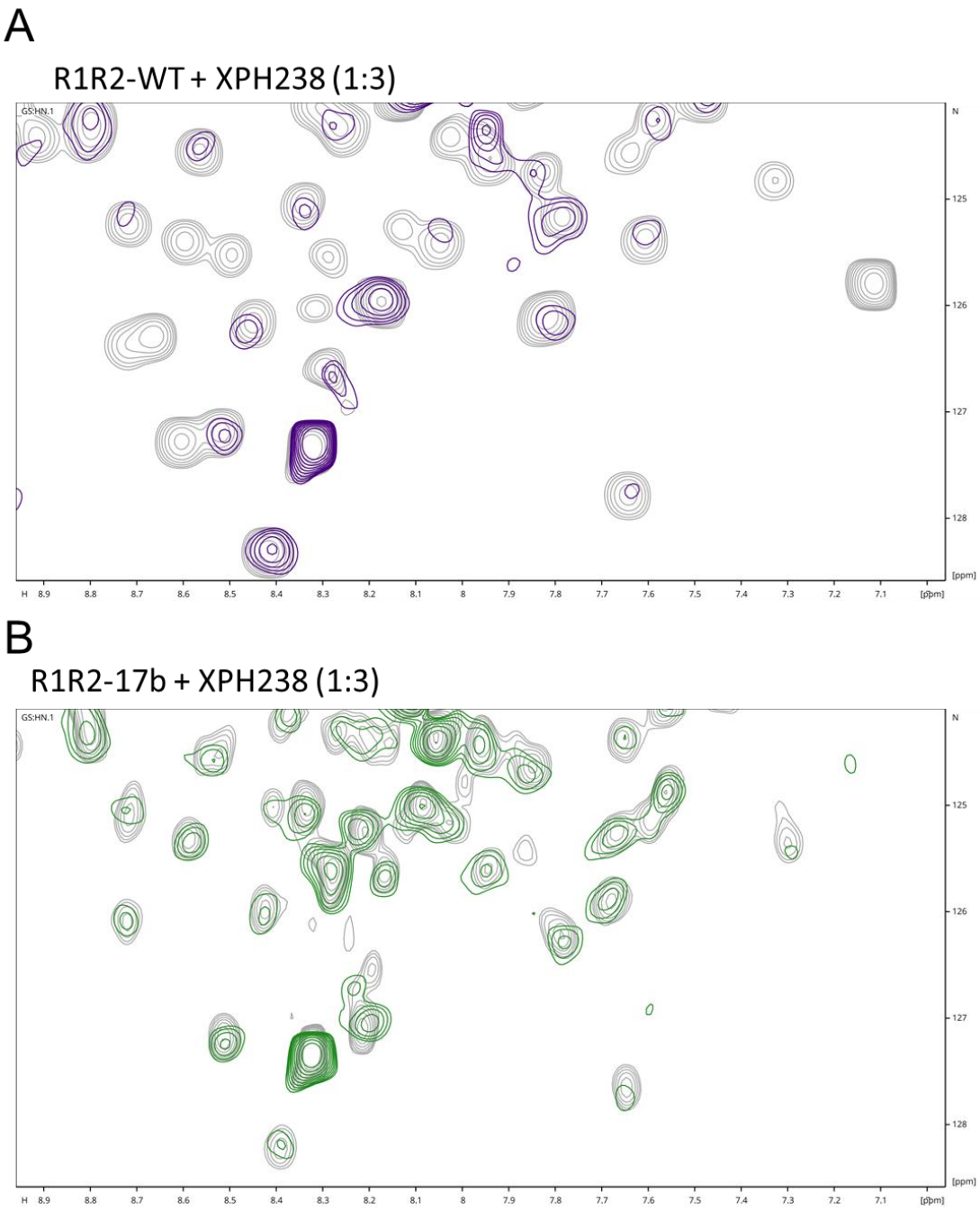


Figure 46: intrabody specificity, lower affinity for talin domain splice variants

Nuclear magnetic resonance data examining the specificity of the XPH238 intrabody for alternate variants (R1R2-17b) of its intended target (R1R2). All samples contained 300µM intrabody and 100µM talin in S column affinity buffer A with 1mM DTT and 5% V/V to allow locking on of the sample, all experiments conducted were 15N TROSY HSQC (2D). A). R1R2 alone (grey) R1R2 1:3 with XPH238 (purple), shifted peaks show binding. B). R1R2-17b alone (grey) R1R2-17b in a 1:3 ratio with XPH238 (green), significantly reduced chemical shift indicates reduced/no binding of XPH238 to the R1R2 variant protein.

Intrabodies discussion

3.3.14 Functionality as an antibody alternative

The data presented here shows the 3 intrabodies that were examined in this project (XPH 238,241,242) bind their intended target regions. They showed varied levels of interactivity as seen in the difference between 242 and 241, with XPH242 causing significantly more chemical shift than XPH241 when titrated with their targets. The data also shows that these novel intrabodies are a functional protein targeting molecule that is both highly specific and binds tightly to its intended target, furthermore, the usage of a bacterial vector to express them makes them far more ethical in their production than the usage of animal-based vectors such as rabbits.

3.3.15 Usage of intrabodies in future talin experiments

The intrabody data shows the three examined intrabodies are capable of forming tight, highly specific interactions with individual target talin regions. The usage of these intrabodies in future talin experiments could help to further elucidate the role of talin within cells. One use of the intrabodies would be the production of ‘reader molecules’, a set of unique intrabodies each targeted to an individual region of talin (e.g., R1, R2, R3 etc.).

Each of these intrabodies would include a unique fluorophore within their structure (such as GFP and RFP (red fluorescent protein)), here intrabodies have an advantage over custom antibodies which must be fluorescently tagged with a fluorophore. Intrabodies being artificial bacterial proteins, can have the desired fluorophore encoded into the protein structure, saving time and expense. These ‘reader molecules’ would then target either the folded or unfolded forms of each helical bundle domain.

Reader molecules could be used to see the effects of mechanical force on talin in real time *in vitro* using light microscopy or *in vivo* using equipment such as the C-Trap® (LUMICKS), which uses optical tweezers to stretch a an individual talin protein. Shown in fig.47A-B, as force is exerted on talin, bound reader molecules would release or bind as helical domains unfolded due to strain.

The fluorescently tagged reader molecule intrabodies could be expressed in live cell cultures to identify target proteins in cells. Allowing visualization of the folding state of FA talin regions in real time as unfolded domains and folded domains would recruit different reader molecules (fig.47B), displaying a different combination of fluorophores depending on the folding state of talin within the adhesion. The expression of the reader molecules in cells could also be used to identify the expression of talin splice variants, such as the SCLS talin variant, which is only expressed during episodes of the disease, therefore excluding the SCLS variant from studies where GFP is encoded into the talin protein itself. Instead, a specific reader molecule designed to bind the SCLS variant would tag the protein during a replicated SCLS episode in cell culture. (fig.47C)

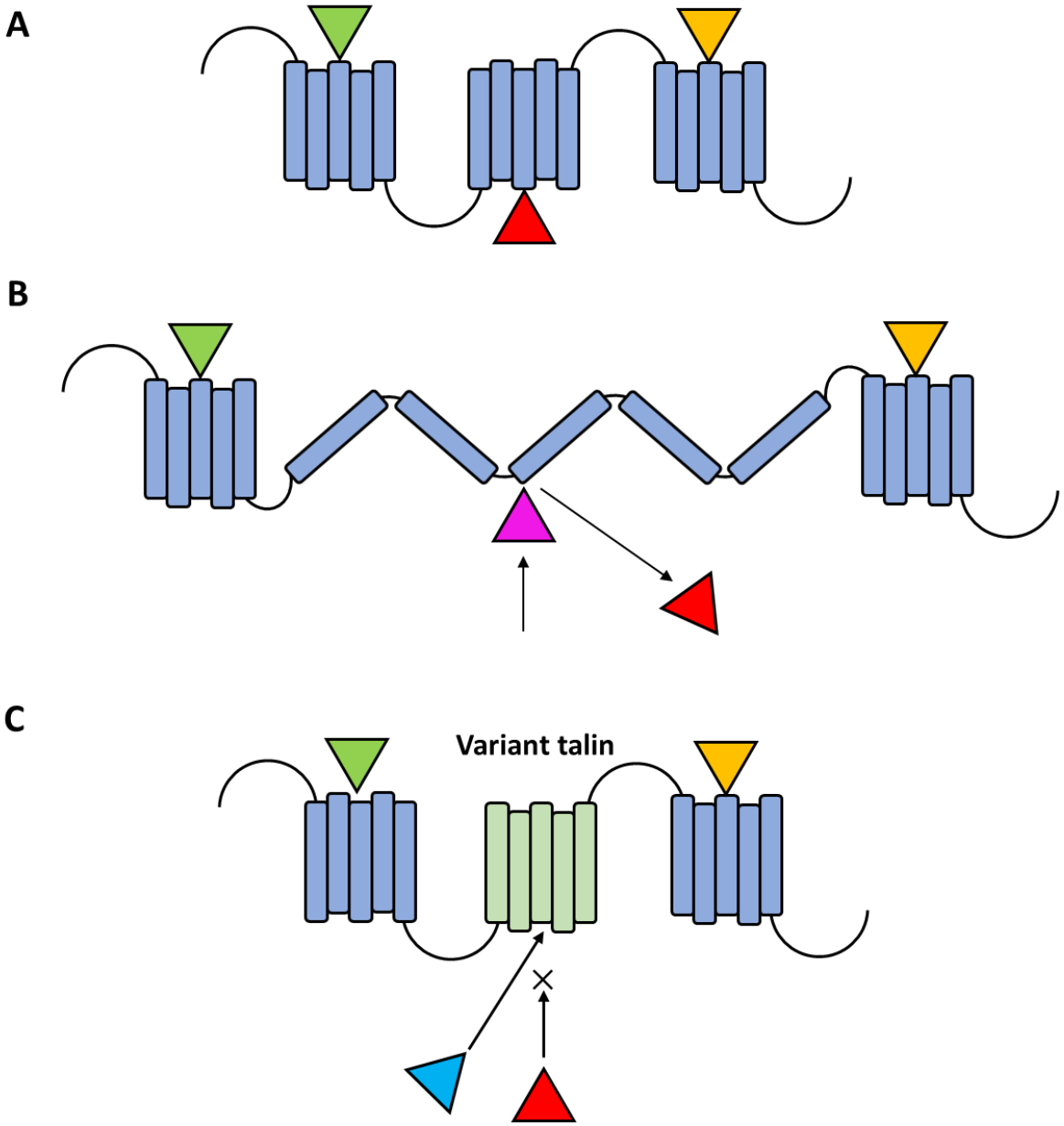


Figure 47: Use of intrabodies as 'reader' molecules

Cartoon showing possible applications of intrabodies as 'reader molecules' for future talin work. A). talin rod domains (blue) and custom intrabodies with fluorophores designed to each bind a single talin domain. B). function of intrabodies as reader molecules, as unfolded talin domain is no longer viable binding site for intrabody (red), causing dissociation of reader molecule, producing detectable data for unfolding of domains. C). usage of reader molecules to examine presence of mutant talin protein in vitro, mutant/variant talin region (green) is no longer viable target for WT region intrabody (red), instead allowing binding of intrabody raised against alternate variant (light blue).

Chapter 4: Future work

4.1 Generation of crystal structures for R1R2-17b and Intrabodies

Throughout the project attempts were made to obtain a crystal structure of R1R2-17b, which were unsuccessful. In future work on the R1R2-17b protein, a crystal structure would be one of the main efforts, as a crystal structure of R1R2-17b would confirm the accuracy of AlphaFold predictive model and show with atomic level detail, the structure of R1R2 affected by the insertion. A confirmed crystal structure would provide further confirmation of the predictions on the effect of exon 17b splicing, as well as being important for any further investigation of the alternative splicing event in both WT and cancerous cells.

With more available time a priority would be to produce crystal structures of all 3 intrabodies (XPH238,241,242) bound to their respective talin targets, as this would validate the predicted binding site based on the NMR data. A crystal structure would be useful for validating the epitope of intrabodies such as XPH242 which likely showed peak shift outside of its epitope region due to the similar size of the intrabody compared to its target talin region.

4.2 Acquiring a Kd of the intrabodies using micro-scale thermophoresis

The intrabodies have shown a high affinity for their target regions, however, precise data on the affinity of these interactions would be useful in the production of future, more efficient 'second generation' intrabody proteins. Specifically, the dissociation constant (Kd) of each intrabody interaction would be needed. The ELISA run by the Bordeaux group established roughly which intrabodies bound the tightest, however, with more time micro-scale-thermophoresis (MST) experiments could be conducted. MST is a technique that uses the

movement of proteins in response to a temperature gradient to accurately establish the kD of tightly interacting proteins.

4.3 R1R2-17b GFP pulldown in cancer lysate

In the future work on exon 17b, a GFP 'fishing' experiment could be conducted to investigate if there is a change in talin interacting proteins during the splicing in of exon 17b. This process is similar to GST pull down experiment(Einarson, 2004), a GFP fusion protein would be created containing R1R2-17b/WT and a GFP tag. The GFP-fusion protein is then purified via nickel and ion chromatography following standard protocol. The purified protein would then be incubated in cell lysate from a known cancer cell line such as BT20, a breast cancer cell line known to express high levels of exon 17b. The R1R2-17b GFP and WT GFP fusion proteins, would then act as 'fishing rods', interacting with proteins within the lysate. Once incubated, the proteins bound to the WT and 17b fusion proteins could be examined, via proteomics, which would show if known cancer associated proteins interact with R1R2 once the 17b exon (and GFP) have been expressed.

4.4 Investigation of the relationship between exon 17b and CLSTN1

During initial discovery of exon 17b, it was found that the splicing of the 11^{th} exon of the calysntenin 1 (CLSTN1) gene was negatively correlated with the splicing of exon 17b in both breast and lung cancer cell lines. This discovery was supported by data from the exon 17b expressing breast cancer cell lines, which were analysed with reverse transcription PCR (RT-PCR), which suggested that talin 1 exon 17b and calsyntenin 1 exon 11 were mutually exclusive splicing events. When 17b-expressing cancer cell lines were stimulated with TGF-β/EGF (predicted regulators of 17b spicing), 17b splicing was down regulated and CLSTN1

exon 11 splicing was upregulated (Gallego-Paez *et al.*, 2022). This suggests the two are part of a TGF-β/EGF regulated alternative splicing pathway.

This pathway would be a central part of future work on exon 17b as CSTLN1 plays a role in stabilizing APP metabolism (CLSTN1 - Calsyntenin-1 - Homo sapiens (Human) | UniProtKB | UniProt, no date). APP is known to be a possible cause of Alzheimer's disease, implying that exon 17b may affect diseases other than cancer. Characterisation of the role of CSTLN1 in adhesions and the effect of its alternate splicing in relation to exon 17b, would lead to new insight into what appears to be a coordinated, pathogenically associated splicing program.

Chapter 5: Final conclusions

This project has centred around talin and has shown that it is an important protein in the working of a cell, and without talin, a cells ability to interact with its environment is significantly impaired. Alterations to talin, either via mutation or by intentional splicing of alternate exons, can produce variants of the talin protein that function differently from the WT. The results of this project have shown that these talin splice variants are prevalent in pathogenic conditions, both purely genetic conditions like SCLS or conditions with both genetic and environmental components such as exon 17b-expressing cancers.

Examination of two examples of disease-associated talin splice variants have revealed the presence of these alternate talin proteins are necessary to the pathogenicity of their associated conditions, showing that talin plays a major role in diseases that involve the dysregulation of cell-cell and cell-matrix interactions.

The intrabodies discussed and examined in this project are shown to function as intended with tight binding to their intended targets. This suggests that they could be raised against

other proteins targets of interest without the involvement of any animals or animal cell lines (therefore removing any existing ethical barrier)

This project has also tested a selection of artificial binding proteins that could serve as monoclonal antibody replacements in future biochemical analysis. These intrabodies display high affinity and highly specific binding for any protein they are raised against. They can differentiate between talins' folded states, and alternate splice variants, making them a useful, and ethical tool in the future of talin research.

Finally, during my work on two separate disease-associated talin splice variants, I have established a pipeline with which to examine the effects of alternate splicing within a protein, from which I can provide biochemical data on the role of a variant protein in its associated disease. This pipeline showed effectiveness with both mutations and newly discovered splice variants, demonstrating the pipelines' viability for future work on talin.

Bibliography

Aarabi, S. *et al.* (2007) 'Mechanical load initiates hypertrophic scar formation through decreased cellular apoptosis', *The FASEB Journal*, 21(12), pp. 3250–3261. doi: 10.1096/FJ.07-8218COM.

Abdellah, Z. *et al.* (2004) 'Finishing the euchromatic sequence of the human genome', *Nature 2004 431:7011*, 431(7011), pp. 931–945. doi: 10.1038/nature03001.

Alfaleh, M. A. *et al.* (2020) 'Phage Display Derived Monoclonal Antibodies: From Bench to Bedside', *Frontiers in Immunology*, 11, p. 1986. doi: 10.3389/FIMMU.2020.01986/BIBTEX.

Atherton, P. *et al.* (2015a) 'Vinculin controls talin engagement with the actomyosin machinery', *Nature Communications 2015 6:1*, 6(1), pp. 1–12. doi: 10.1038/ncomms10038.

Atherton, P. *et al.* (2015b) 'Vinculin controls talin engagement with the actomyosin machinery', *Nature Communications 2015 6:1*, 6(1), pp. 1–12. doi: 10.1038/ncomms10038.

Banno, A. *et al.* (2012) 'Subcellular localization of talin is regulated by inter-domain interactions', *Journal of Biological Chemistry*, 287(17), pp. 13799–13812. doi: 10.1074/jbc.M112.341214.

Baum, B. and Georgiou, M. (2011) 'Dynamics of adherens junctions in epithelial establishment, maintenance, and remodeling', *The Journal of Cell Biology*, 192(6), p. 907. doi: 10.1083/JCB.201009141.

Bays, J. L. and DeMali, K. A. (2017) 'Vinculin in cell–cell and cell–matrix adhesions', *Cellular and Molecular Life Sciences*, 74(16), p. 2999. doi: 10.1007/S00018-017-2511-3.

Borgi, I. and Gargouri, A. (2008) 'A spontaneous direct repeat deletion in the pGEX fusion vector decreases the expression level of recombinant proteins in Escherichia coli', *Protein*

expression and purification, 60(1), pp. 15–19. doi: 10.1016/J.PEP.2008.03.012.

Bork, P. and Doolittle, R. F. (1992) 'Proposed acquisition of an animal protein domain by bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 89(19), pp. 8990–8994. doi: 10.1073/PNAS.89.19.8990.

Butcher, D. T., Alliston, T. and Weaver, V. M. (2009) 'A tense situation: forcing tumour progression', *Nature Reviews Cancer 2009 9:2*, 9(2), pp. 108–122. doi: 10.1038/nrc2544.

Calderwood, D. A. *et al.* (2002) 'The Phosphotyrosine Binding-like Domain of Talin Activates Integrins *', *Journal of Biological Chemistry*, 277(24), pp. 21749–21758. doi: 10.1074/JBC.M111996200.

Campbell, I. D. and Spitzfaden, C. (1994) 'Building proteins with fibronectin type III modules', *Structure*, 2(5), pp. 333–337. doi: 10.1016/S0969-2126(00)00034-4.

Carisey, A. and Ballestrem, C. (2011) 'Vinculin, an adapter protein in control of cell adhesion signalling', *European Journal of Cell Biology*, 90(2–3), pp. 157–163. doi: 10.1016/J.EJCB.2010.06.007.

Clarkson, B. *et al.* (1960a) 'Cyclical edema and shock due to increased capillary permeability', *The American journal of medicine*, 29(2), pp. 193–216. doi: 10.1016/0002-9343(60)90018-8.

Clarkson, B. *et al.* (1960b) 'Cyclical edema and shock due to increased capillary permeability', *The American Journal of Medicine*, 29(2), pp. 193–216. doi: 10.1016/0002-9343(60)90018-8.

CLSTN1 - Calsyntenin-1 - Homo sapiens (Human) | UniProtKB | UniProt (no date). Available at: https://www.uniprot.org/uniprotkb/O94985/entry (Accessed: 24 August 2022).

Cohen, D. M. *et al.* (2005) 'Two distinct head-tail interfaces cooperate to suppress activation of vinculin by talin', *Journal of Biological Chemistry*, 280(17), pp. 17109–17117. doi: 10.1074/jbc.M414704200.

Crick, F. (1970) 'Central Dogma of Molecular Biology', *Nature 1970 227:5258*, 227(5258), pp. 561–563. doi: 10.1038/227561a0.

Critchley, D. R. (2000) 'Focal adhesions – the cytoskeletal connection', *Current Opinion in Cell Biology*, 12(1), pp. 133–139. doi: 10.1016/S0955-0674(99)00067-8.

Einarson, M. B. (2004) 'Detection of protein-protein interactions using the GST fusion protein pull-down technique', *Nature Methods 2004 1:3*, 1(3), pp. 275–276. doi: 10.1038/nmeth1204-275.

Elefant, N. *et al.* (2022) 'Talin1 dysfunction is genetically linked to systemic capillary leak syndrome', *medRxiv*, p. 2022.10.17.22280833. doi: 10.1101/2022.10.17.22280833.

Elliott, P. R. *et al.* (2010) 'The Structure of the Talin Head Reveals a Novel Extended Conformation of the FERM Domain', *Structure(London, England:1993)*, 18(10–13), p. 1289. doi: 10.1016/J.STR.2010.07.011.

Frantz, C., Stewart, K. M. and Weaver, V. M. (2010) 'The extracellular matrix at a glance', *Journal of Cell Science*, 123(24), pp. 4195–4200. doi: 10.1242/JCS.023820.

Gallego-Paez, L. M. *et al.* (2022) 'A novel cancer-associated cassette exon in TLN1 alters Talin 1 mechanosensitivity', *bioRxiv*, p. 2022.09.06.506802. doi: 10.1101/2022.09.06.506802.

Gasteiger, E. *et al.* (no date) *Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy*

Server. Available at: http://www.expasy.org/tools/. (Accessed: 4 April 2021).

Gilmore, A. P. and Burridge, K. (1996) 'Regulation of vinculin binding to talin and actin by phosphatidyl-inositol-4-5-bisphosphate', *Nature 1996 381:6582*, 381(6582), pp. 531–535. doi: 10.1038/381531a0.

Gingras, A. R. *et al.* (2008) 'The structure of the C-terminal actin-binding domain of talin', *The EMBO Journal*, 27(2), pp. 458–469. doi: 10.1038/SJ.EMBOJ.7601965.

Go, C. D. *et al.* (2021) 'A proximity-dependent biotinylation map of a human cell', *Nature 2021 595:7865*, 595(7865), pp. 120–124. doi: 10.1038/s41586-021-03592-2.

Gough, R. E. (2018) *Interaction of Talin with Proteins Containing LD-Motifs*.

Gough, R. E. and Goult, B. T. (2018) 'The tale of two talins – two isoforms to fine-tune integrin signalling', *FEBS Letters*, 592(12), pp. 2108–2125. doi: 10.1002/1873-3468.13081.

Goult, B. T. *et al.* (2010) 'Structure of a double ubiquitin-like domain in the talin head: a role in integrin activation', *The EMBO Journal*, 29(6), pp. 1069–1080. doi: 10.1038/EMBOJ.2010.4.

Goult, B. T. *et al.* (2013a) 'RIAM and vinculin binding to talin are mutually exclusive and regulate adhesion assembly and turnover', *Journal of Biological Chemistry*, 288(12), pp. 8238–8249. doi: 10.1074/JBC.M112.438119/ATTACHMENT/4F92C301-4898-4990-A4C8-1E69A19C5A34/MMC1.PDF.

Goult, B. T. *et al.* (2013b) 'RIAM and vinculin binding to talin are mutually exclusive and regulate adhesion assembly and turnover', *Journal of Biological Chemistry*, 288(12), pp. 8238–8249. doi: 10.1074/JBC.M112.438119/ATTACHMENT/4F92C301-4898-4990-A4C8-1E69A19C5A34/MMC1.PDF.

Goult, B. T. *et al.* (2013c) 'RIAM and vinculin binding to talin are mutually exclusive and regulate adhesion assembly and turnover', *The Journal of biological chemistry*, 288(12), pp. 8238–8249. doi: 10.1074/JBC.M112.438119.

Goult, B. T., Yan, J. and Schwartz, M. A. (2018) 'Talin as a mechanosensitive signaling hub', *Journal of Cell Biology*, 217(11), pp. 3776–3784. doi: 10.1083/JCB.201808061.

Gousseff, M. *et al.* (2011) 'The systemic capillary leak syndrome: A case series of 28 patients from a European registry', *Annals of Internal Medicine*, 154(7), pp. 464–471. doi: 10.7326/0003-4819-154-7-201104050-00004.

Guck, J. *et al.* (2005) 'Optical Deformability as an Inherent Cell Marker for Testing Malignant Transformation and Metastatic Competence', *Biophysical Journal*, 88(5), pp. 3689–3698. doi: 10.1529/BIOPHYSJ.104.045476.

Hahn, R. G. (2021) 'Fluid escapes to the "third space" during anesthesia, a commentary', *Acta Anaesthesiologica Scandinavica*, 65(4), p. 451. doi: 10.1111/AAS.13740.

Heier, J. A., Dickinson, D. J. and Kwiatkowski, A. V. (2017) 'Measuring Protein Binding to F-actin by Co-sedimentation', *Journal of Visualized Experiments : JoVE*, 2017(123), p. 55613. doi: 10.3791/55613.

Hemmings, L. *et al.* (1996) 'Talin contains three actin-binding sites each of which is adjacent to a vinculin-binding site', *Journal of Cell Science*, 109(11), pp. 2715–2726. doi: 10.1242/JCS.109.11.2715.

Hsu, P. *et al.* (2015) 'Idiopathic Systemic Capillary Leak Syndrome in Children', *Pediatrics*, 135(3), pp. e730–e735. doi: 10.1542/PEDS.2014-2268.

Humphrey, J. D., Dufresne, E. R. and Schwartz, M. A. (2014) 'Mechanotransduction and

extracellular matrix homeostasis', *Nature Reviews Molecular Cell Biology 2014 15:12*, 15(12), pp. 802–812. doi: 10.1038/nrm3896.

Humphries, J. D. *et al.* (2007) 'Vinculin controls focal adhesion formation by direct interactions with talin and actin', *Journal of Cell Biology*, 179(5), pp. 1043–1057. doi: 10.1083/JCB.200703036/VIDEO-4.

Huveneers, S. *et al.* (2012) 'Vinculin associates with endothelial VE-cadherin junctions to control force-dependent remodeling', *Journal of Cell Biology*, 196(5), pp. 641–652. doi: 10.1083/JCB.201108120/VIDEO-9.

Iskratsch, T., Wolfenson, H. and Sheetz, M. P. (2014) 'Appreciating force and shape — the rise of mechanotransduction in cell biology', *Nature Reviews Molecular Cell Biology 2014 15:12*, 15(12), pp. 825–833. doi: 10.1038/nrm3903.

Jansen, K. A. *et al.* (2013) 'Cells Actively Stiffen Fibrin Networks by Generating Contractile Stress', *Biophysical Journal*, 105(10), pp. 2240–2251. doi: 10.1016/J.BPJ.2013.10.008.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature 2021 596:7873*, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.

Kalluri, R. and Weinberg, R. A. (2009) 'The basics of epithelial-mesenchymal transition', *The Journal of Clinical Investigation*, 119(6), p. 1420. doi: 10.1172/JCI39104.

Khalili, A. A. and Ahmad, M. R. (2015) 'A Review of Cell Adhesion Studies for Biomedical and Biological Applications', *International Journal of Molecular Sciences*, 16(8), p. 18149. doi: 10.3390/IJMS160818149.

Khan, R. B. *et al.* (2021) 'Biochemical Characterization of the Integrin Interactome', *Methods in Molecular Biology*, 2217, pp. 115–147. doi: 10.1007/978-1-0716-0962-0_9.

Kim, D. H. and Wirtz, D. (2013) 'Focal adhesion size uniquely predicts cell migration', *The FASEB Journal*, 27(4), p. 1351. doi: 10.1096/FJ.12-220160.

Kisseleva, T. *et al.* (2012) 'Myofibroblasts revert to an inactive phenotype during regression of liver fibrosis', *Proceedings of the National Academy of Sciences of the United States of America*, 109(24), pp. 9448–9453. doi: 10.1073/PNAS.1201840109/-/DCSUPPLEMENTAL.

Klapholz, B. and Brown, N. H. (2017) 'Talin - The master of integrin adhesions', *Journal of Cell Science*, 130(15), pp. 2435–2446. doi: 10.1242/JCS.190991/265176/AM/TALIN-THE-MASTER-OF-INTEGRIN-ADHESIONS.

Kumar, A. *et al.* (2016) 'Talin tension sensor reveals novel features of focal adhesion force transmission and mechanosensitivity', *Journal of Cell Biology*, 213(3), pp. 371–383. doi: 10.1083/JCB.201510012.

Lee, H. S. *et al.* (2004) 'Characterization of an actin-binding site within the talin FERM domain', *Journal of Molecular Biology*, 343(3), pp. 771–784. doi: 10.1016/j.jmb.2004.08.069.

Lin, H. H. *et al.* (2015) 'Mechanical phenotype of cancer cells: cell softening and loss of stiffness sensing', *Oncotarget*, 6(25), p. 20946. doi: 10.18632/ONCOTARGET.4173.

Lipman, N. S. *et al.* (2005) 'Monoclonal Versus Polyclonal Antibodies: Distinguishing Characteristics, Applications, and Information Resources', *ILAR Journal*, 46(3), pp. 258–268. doi: 10.1093/ILAR.46.3.258.

Liu, F., Zhang, J. Z. H. and Mei, Y. (2016) 'The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations', *Scientific Reports 2016 6:1*, 6(1), pp. 1–11. doi: 10.1038/srep27190.

Liu, J. K. H. (2014) 'The history of monoclonal antibody development – Progress, remaining

challenges and future innovations', *Annals of Medicine and Surgery*, 3(4), p. 113. doi: 10.1016/J.AMSU.2014.09.001.

Madden, T. (2003) 'The BLAST Sequence Analysis Tool'. Available at: https://www.ncbi.nlm.nih.gov/books/NBK21097/ (Accessed: 29 August 2022).

Maddugoda, M. P. *et al.* (2007) 'Myosin VI and vinculin cooperate during the morphogenesis of cadherin cell–cell contacts in mammalian epithelial cells', *Journal of Cell Biology*, 178(3), pp. 529–540. doi: 10.1083/JCB.200612042.

Mariani, V. *et al.* (2013) 'lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests', *Bioinformatics*, 29(21), p. 2722. doi: 10.1093/BIOINFORMATICS/BTT473.

Marks, J. and Shuster, S. (1973) 'Disorders of capillary permeability', *The British journal of dermatology*, 88(6), pp. 619–621. doi: 10.1111/J.1365-2133.1973.TB08029.X.

Martel, V. *et al.* (2001) 'Conformation, Localization, and Integrin Binding of Talin Depend on Its Interaction with Phosphoinositides *', *Journal of Biological Chemistry*, 276(24), pp. 21217–21227. doi: 10.1074/JBC.M102373200.

McCafferty, J. *et al.* (1990) 'Phage antibodies: filamentous phage displaying antibody variable domains', *Nature 1990 348:6301*, 348(6301), pp. 552–554. doi: 10.1038/348552a0.

McLachlan, A. D. *et al.* (1994) 'Analysis of Repeated Motifs in the Talin Rod', *Journal of Molecular Biology*, 235(4), pp. 1278–1290. doi: 10.1006/JMBI.1994.1081.

Mirdita, M. *et al.* (2022) 'ColabFold: making protein folding accessible to all', *Nature Methods 2022 19:6*, 19(6), pp. 679–682. doi: 10.1038/s41592-022-01488-1.

Mureddu, L. and Vuister, G. W. (2019) 'Simple high-resolution NMR spectroscopy as a tool in molecular biology', *The FEBS Journal*, 286(11), pp. 2035–2042. doi: 10.1111/FEBS.14771.

Papagrigoriou, E. *et al.* (2004) 'Activation of a vinculin-binding site in the talin rod involves rearrangement of a five-helix bundle', *EMBO Journal*, 23(15), pp. 2942–2951. doi: 10.1038/SJ.EMBOJ.7600285.

Pearson, M. A. *et al.* (2000) 'Structure of the ERM Protein Moesin Reveals the FERM Domain Fold Masked by an Extended Actin Binding Tail Domain', *Cell*, 101(3), pp. 259–270. doi: 10.1016/S0092-8674(00)80836-3.

Pierce, R. *et al.* (2019) 'Whole exome sequencing of adult and pediatric cohorts of the rare vascular disorder Systemic Capillary Leak Syndrome', *Shock (Augusta, Ga.)*, 52(2), p. 183. doi: 10.1097/SHK.0000000000001254.

Rodius, S. *et al.* (2008) 'The talin rod IBS2 α-helix interacts with the β3 integrin cytoplasmic tail membrane-proximal helix by establishing charge complementary salt bridges', *Journal of Biological Chemistry*, 283(35), pp. 24212–24223. doi: 10.1074/JBC.M709704200/ATTACHMENT/21E95667-ECEF-43E9-8DAA-D6B36C96245D/MMC1.PDF.

Schaefer, L. and Schaefer, R. M. (2010) 'Proteoglycans: From structural compounds to signaling molecules', *Cell and Tissue Research*, 339(1), pp. 237–246. doi: 10.1007/S00441-009-0821-Y/TABLES/3.

Siddall, E., Khatri, M. and Radhakrishnan, J. (2017) 'Capillary leak syndrome: etiologies, pathophysiology, and management', *Kidney International*, 92(1), pp. 37–46. doi: 10.1016/J.KINT.2016.11.029.

Sigismund, S., Avanzato, D. and Lanzetti, L. (2018) 'Emerging functions of the EGFR in cancer', *Molecular Oncology*, 12(1), p. 3. doi: 10.1002/1878-0261.12155.

Singh, R. K. and Cooper, T. A. (2012) 'Pre-mRNA splicing in disease and therapeutics', *Trends in Molecular Medicine*, 18(8), pp. 472–482. doi: 10.1016/J.MOLMED.2012.06.006.

Singh, R. N. and Singh, N. N. (2018) 'Mechanism of Splicing Regulation of Spinal Muscular Atrophy Genes', *Advances in neurobiology*, 20, p. 31. doi: 10.1007/978-3-319-89689-2_2.

Smith, G. P. and Petrenko, V. A. (1997) 'Phage display', *Chemical Reviews*, 97(2), pp. 391–410. doi: 10.1021/CR960065D/ASSET/IMAGES/LARGE/CR960065DF00004.JPEG.

Smith, M. L. *et al.* (2007) 'Force-Induced Unfolding of Fibronectin in the Extracellular Matrix of Living Cells', *PLOS Biology*, 5(10), p. e268. doi: 10.1371/JOURNAL.PBIO.0050268.

Smith, S. J. and McCann, R. O. (2007) 'A C-terminal dimerization motif is required for focal adhesion targeting of Talin1 and the interaction of the Talin1 I/LWEQ module with F-actin', *Biochemistry*, 46(38), pp. 10886–10898. doi: 10.1021/BI700637A/ASSET/IMAGES/LARGE/BI700637AF00010.JPEG.

Stacey, D. W. *et al.* (1985) 'Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface', *Science*, 228(4705), pp. 1315–1317. doi: 10.1126/SCIENCE.4001944.

Stricker, J., Falzone, T. and Gardel, M. L. (2010) 'Mechanics of the F-actin cytoskeleton', *Journal of Biomechanics*, 43(1), pp. 9–14. doi: 10.1016/J.JBIOMECH.2009.09.003.

Sun, Z., Guo, S. S. and Fässler, R. (2016) 'Integrin-mediated mechanotransduction', *Journal of Cell Biology*, 215(4), pp. 445–456. doi: 10.1083/JCB.201609037.

Suresh, S. (2007) 'Biomechanics and biophysics of cancer cells', *Acta Biomaterialia*, 3(4), pp. 413–438. doi: 10.1016/J.ACTBIO.2007.04.002.

Tschumperlin, D. J. *et al.* (2018) 'Mechanosensing and fibrosis', *The Journal of Clinical Investigation*, 128(1), pp. 74–84. doi: 10.1172/JCI93561.

Varadi, M. *et al.* (2022) 'AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research*, 50(D1), p. D439. doi: 10.1093/NAR/GKAB1061.

WANG, Y. *et al.* (2015) 'Mechanism of alternative splicing and its regulation', *Biomedical Reports*, 3(2), p. 152. doi: 10.3892/BR.2014.407.

Weber, J., Peng, H. and Rader, C. (2017) 'From rabbit antibody repertoires to rabbit monoclonal antibodies', *Experimental & Molecular Medicine 2017 49:3*, 49(3), pp. e305–e305. doi: 10.1038/emm.2017.23.

Wegener, K. L. *et al.* (2007) 'Structural Basis of Integrin Activation by Talin', *Cell*, 128(1), pp. 171–182. doi: 10.1016/j.cell.2006.10.048.

Will, C. L. and Lührmann, R. (2011) 'Spliceosome Structure and Function', *Cold Spring Harbor Perspectives in Biology*, 3(7), pp. 1–2. doi: 10.1101/CSHPERSPECT.A003707.

Wolfenson, H. *et al.* (2009) 'The heel and toe of the cell's foot: A multifaceted approach for understanding the structure and dynamics of focal adhesions', *Cell Motility and the Cytoskeleton*, 66(11), pp. 1017–1029. doi: 10.1002/CM.20410.

Wozniak, M. A. *et al.* (2004) 'Focal adhesion regulation of cell behavior', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1692(2–3), pp. 103–119. doi: 10.1016/J.BBAMCR.2004.04.007.

Wu, C. (2007) 'Focal Adhesion: A Focal Point in Current Cell Biology and Molecular Medicine', *Cell Adhesion & Migration*, 1(1), p. 13. doi: 10.4161/CAM.1.1.4081.

Yan, J. *et al.* (2015) 'Talin Dependent Mechanosensitivity of Cell Focal Adhesions', *Cellular and Molecular Bioengineering*, 8(1), pp. 151–159. doi: 10.1007/S12195-014-0364-5/FIGURES/4.

Yao, M. *et al.* (2014) 'Mechanical activation of vinculin binding to talin locks talin in an unfolded conformation', *Scientific Reports*, 4. doi: 10.1038/SREP04610.

Yao, M. *et al.* (2016) 'The mechanical response of talin', *Nature Communications 2016 7:1*, 7(1), pp. 1–11. doi: 10.1038/ncomms11966.

Yarden, Y. and Sliwkowski, M. X. (2001) 'Untangling the ErbB signalling network', *Nature Reviews Molecular Cell Biology 2001 2:2*, 2(2), pp. 127–137. doi: 10.1038/35052073.

Zahavi, D. and Weiner, L. (2020) 'Monoclonal Antibodies in Cancer Therapy', *Antibodies*, 9(3), pp. 1–20. doi: 10.3390/ANTIB9030034.