Filamin: A Family of Mechanosensory Scaffold Proteins

Barry Alan Kesner

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Cell and Developmental Biology (School Of Medicine).

Chapel Hill
2009

Approved by,

Advisor: Nikolay V. Dokholyan, PhD

Committee Chair: Doug M. Cyr, PhD

Ken A. Jacobson, PhD

Sharon L. Milgram, PhD

Garegin A. Papoian, PhD

Brenda R. S. Temple, PhD

Abstract

The vertebrate filamin family (A, B, and C) are critically involved in development of brain structure, cardio-vasculature, and skeleton. Filamins are large F-actin cross-linking proteins containing an actin-binding domain, followed by 24 Immunoglobulin(Ig)-like domains. Filamins play critical cellular roles as mechanical and scaffold proteins. As mechanical proteins they cross-link F-actin into gels or fibrils. As scaffold proteins they bind over 70 critical proteins. Filamins have overlapping and distinct roles, however it is not understood why. Nor well understood, are the mechanisms by which filamin act as mechanical proteins. Our work focuses on understanding their function; by analyzing site-specific functional divergence, using the evolutionary trace (ET) method, over vertebrate developmental periods —— Teleostei, Amphibian, and Mammalian; and by analyzing filamin behavior under mechanical stress. We find, isoforms diverge from one gene between urochordate and vertebrate lineages; most divergence occurs in Teleostei; and that filamin C diverged least. In addition, the heterogeneous spatial pattern of functional divergence we observe is not correlated with scaffold protein activity either in binding interfaces or across domains. Our results also suggest isoforms have diverged with regard to specificity for binding partners or regulatory function. To elucidate the structure-function relationship of filamin, we used constant force (0-315 pN) simulations to derive both the critical unfolding force and the unfolding pathways at biological levels of force (35-70 pN). Despite a large heterogeneity in

the population of force-induced intermediate states, we find a common initial unfolding intermediate in all the Ig-like domains of filamin, where the N-terminal β strand unfolds. We also study the simulated thermal unfolding of several filamin Ig-like domains. We find that thermally-induced unfolding has an early-stage intermediate state similar to the one observed in force-induced unfolding and characterized by the N-terminal strand being unfurled. We propose that the N-terminal strand may act as a conformational switch that unfolds under physiological forces leading to exposure of cryptic binding sites, removal of native binding sites, and modulating the quaternary structure of domains. This work provides insights into both isoform distinctive and mechanical properties of vertebrate filamin.

This dissertation is dedicated to my wife, Julie, whose love, support, and patience have given me the strength to attain my goals while reminding me to take time along the way to enjoy the journey.

Acknowledgements

I wish to thank my advisor, Dr. Nikolay Dokholyan, for accepting me into his laboratory. He has been supportive of my work, and I am grateful to have him as my mentor. In addition, I want to thank other members of my committee including Dr. Brenda Temple, Dr. Doug Cyr, Dr. Sharon Milgram, Dr. Garegin Papoian, and Dr. Ken Jacobson. I would like to thank Dr. Sharon Milgram for letting me work in her lab and taking a chance on me. She and her associates taught me the basics of bench work, and I will forever be in their debt. I also thank Dr. Martin Playford for advising me on a personal level and for taking a great interest in my studies. In addition, I wish to thank members of the lab, both past and present: Dr. Feng Ding, Dr. Shuangye Yin, Dr. Adrian Serohijos, Srinivas Ramachandran, and Pradeep Kota. Their support, discussions and guidance have been most helpful. I am especially grateful to Dr. Brenda Temple for taking me under her wing when I was most in need. She is a great person and I am glad that I have had the pleasure of working with her. Her approach to teaching and science is genuine and sincere. She has also shown great patience with me, and for this I am grateful.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Aa	amino acid
AO	Atelosteogenesis
ABD	actin-binding domain
AFM	Atomic Force Microscopy
BD	Boomerang dysplasia
BSS	Bernard-Soulier syndrome
cAMP	cyclic AMP
CF	Constant Force
CH	calponin homology
CIP	Chronic idiopathic intestinal pseudo-obstruction
C-terminal	carboxyl-terminal
ddFilamin	*Dictyostelium discoidiem* filamin
DMD	discrete molecular dynamic
ECM	extracellular matrix
ET	evolutionary trace
F-actin	Filamentous actin
FILGAP	Filamin associated GAP
FMD	Frontometaphyseal dysplasia
FN	fibronectin
$FNIII_{10}$	fibronectin domain 10, of fibronectin
GPCR	G-protein coupled receptors
GPIbα	platelet glycoprotein 1, α subunit

GTPase	guanine triphosphate hydrolysis activating protein
H1/2	filamin hinge 1/2
id	identity
Ig-like	Immunoglobulin-like
K	Kelvin
LS	Larsen syndrome
MD	molecular dynamic
MEKK4	map kinase kinase
MFPT	mean first passage time
NMR	nuclear magnetic resonance
N-terminal	amino-terminal
OPD	otopalatodigital syndrome
Q-value	percent of native contacts
PDB	protein data bank
PV	Periventricular nodular heterotopia
pN	picoNewton
SD	standard deviation
TNI27	titin Ig-like domain 27
TGF	transforming growth factor
WHAM	weighted histogram analysis method

List of Symbols

A	Arrhenius prefactor
F	force
F_c	critical unfolding force
K_b	Boltzmann constant
T	temperature
τ	mean first passage time
$\Delta G^{\ddagger}$	activation energy
ΔH	alternate splice variant of filamin lacking a hinge
$\Delta Var1$	alternate splice variant of filamin lacking 41 in β strand 19 to 20
x	end-to-end extension length

Chapter 1
The filamin family of proteins

The vertebrate filamin family (A, B, and C) are critically involved in development of brain structure, cardio-vasculature, and skeleton. Filamins play critical cellular roles as mechanical and scaffold proteins. As mechanical proteins they cross-link F-actin into gels or fibrils. As scaffold proteins they bind over 70 critical proteins. Filamins have overlapping and distinct roles, however it is not understood why. Nor well understood, are the mechanisms by which filamins act as mechanical proteins. This work addresses these questions.

1.1 Background

All vertebrate filamin family members are architecturally composed of an N-terminal actin-binding domain (ABD) followed by a series of 24, ~96 residue Immunoglobulin(Ig)-like domains (Figure 1.1). The 24th domain is the self-association domain giving a filamin dimer the ability to cross-link F-actin. Invertebrate filamin that have fewer domains exist. For example, the single homologous fly filamin has an ABD followed by only 20 Ig-like domains (6); and the single *Dictyostelium discoidiem* filamin (ddFilamin) has an ABD, but is followed by only six Ig-like domains (Table 1) (7). All family members have been associated with mechanical roles such as cellular motility, cell morphology, and involvement in stress areas of cells (8-11). In addition, family members act as scaffold proteins and are involved in many critical signaling pathways including inside-out and outside-in signaling. Signaling also results in cytoskeletal rearrangements associated with

cell motility and cell morphology modulation (12-14). Thus, filamins are a structural protein and are also involved in signaling structural changes within a cell.

The N-terminal ABD gives filamin a mechanism for both binding actin, and regulating this interaction. The ABD consists of two calponin-homology (CH) domains connected by a negatively charged linker. CH domains are ball-like, consisting of seven alpha helices. CH 1 has two putative actin-binding sites, and CH 1 has been shown to directly bind F-actin (15). CH 2 has one putative actin-binding site, and CH 2 is believed to play a regulatory role in binding F-actin. The interaction of filamin with calmodulin on CH 1 is reported to also regulate its interaction with F-actin (2). However, the role of CH 2 is less well understood. Several missense mutations in CH 2 of both filamin A and B cause a spectrum of similar skeletal disorders. These mutations are believed to enhance binding to F-actin (170), and have been considered gain-of-function mutations (17). However, the function of CH 2 is not well understood.
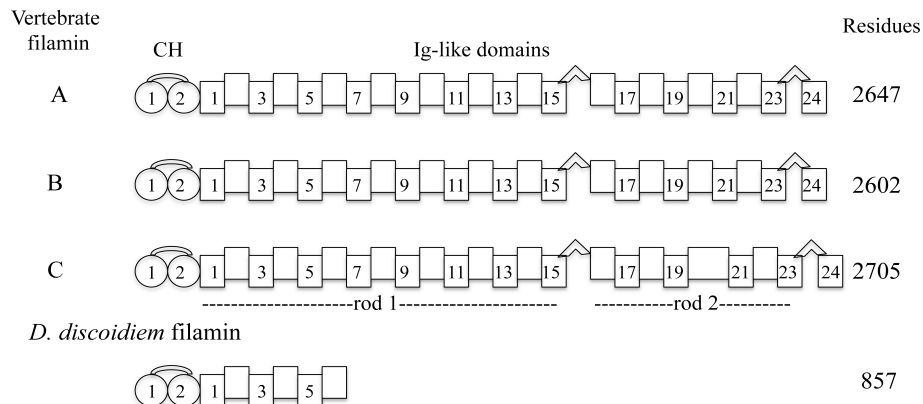
Figure 1.1. Filamin family of proteins. Models of vertebrate filamin A, B, and C showing their Ig-like and calponin homology domains; and regions rod 1 and rod 2. *D. discoidiem* is an invertebrate filamin having only six Ig-like domains.

Ig-like domains give filamin their strength and long reach. Ig-like domains are β sandwich structures consisting of seven β strands (A-G) (Figure 1.2). The two terminal edges

of Ig-like domains are composed of β strands AG and CD. Each Ig-like domain is able to

withstand large forces (50-220 pN) (16). The ability to withstand force, along with the

overall domain architecture, is associated with giving cells their rheological properties (17).

A filamin dimer can be extended end-to-end over 160 nm (18). The long extension comes

from the linear arrangement of Ig-like domains in rod regions. An elongated series of 15 Ig-

like domains, connected by short linkers, make up the first rod region. A hinge (H1) (> 30

aa) connects rod 1 to a second rod region giving filamin high flexibility (19). Rod 2 is more

compact and made up of Ig-like domains 16 and above. A second hinge (H2) connects

domain 23 to the dimerization domain, 24 (20). Ig-like domains also allow filamins to act as

scaffold proteins. Most binding partners occur though interaction on rod 2, and binding

isolated to the CD face. The reasons for these specificities are not understood.
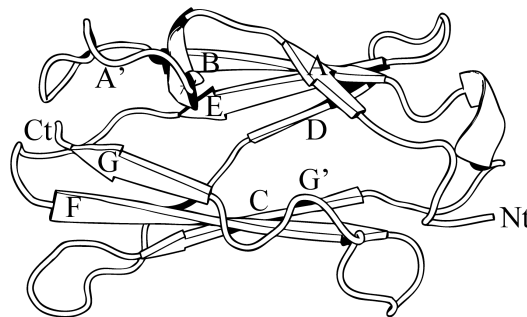
Figure 1.2. Filamin Ig-like domain. This domain
consists of a seven-stranded β sandwich structure labeled
A-G, with β strands AG located on faces opposite to CD.
Both β strands A and G are bisected, having components
A, A', G', and G respectively.

The diversity of vertebrate filamin is increased by alternate splice-variants (21) that

have differing structural roles in cells. Internal insertions and deletions in filamin isoforms

vary by tissue type and development stage. While H2 exists in all filamin isoforms, H1 is

3

absent in some forms of filamin B (ΔH1) and all forms of Filamin C. In filamin A and B there is a deletion of 41 residues in the last β strand of domain 19 and the first two β strands of domain 20 (ΔVar1)(21). These two variants are widely expressed at low levels. Filamin C has an insertion of ~80 residues at the end of β strand A in domain 20, which is associated with its localization in muscle.

Filamins play a dual role, functioning as both mechanical and scaffold proteins. As mechanical proteins they cross-link F-actin into gels and fibrils. As scaffold proteins they bind over 70 critical proteins, including transmembrane proteins as well as proteins involved in trafficking, signaling, F-actin remodeling, and transcription (Table 1).
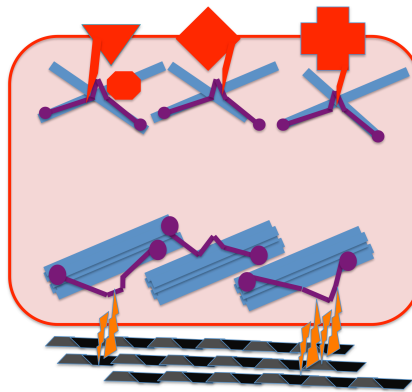
Figure 1.3. The role of filamin in a cell. Filamin (purple) cross-links actin (blue) into orthogonal gels and fibers. It also is a scaffold interacting with many transmembrane proteins (red) including integrins (orange) which interact with the extracellular matrix (ECM-black/gray). Filamin also interacts with cytosolic effector proteins (red).

Table 1. Filamin-associated function

Filamin-associated function	Reference
Outside-in signaling	(22-25)
Ion channels	(26, 27)
Vesicle trafficking	(28)
Endocytosis	(29-31)
Soluble receptor	(32-34)
Transcription factor	(35, 36)
Cytokinesis	(37, 38)
Muscle regeneration	(39, 40)
Muscle contractile	(41, 42, 43)
Cellular adhesion	(2, 5, 44-50)
Actin remodeling	(51-53)
Differentiation	(21, 54-57)
Mechanoprotection	(58-61)

1.1.1 Filamins as mechanical proteins

Filamin is one of the most potent F-actin cross-linkers. *In vitro*, at low concentration,

filamin produces the same sized F-actin structures as other F-actin cross-linking proteins do

at much higher concentrations (15). At low concentrations filamin and F-actin produce gels,

but at higher concentration they produce bundles. *In vitro*, gels have rheological properties

similar to those of cells (17). A recent hypothesis suggests that filamin not only cross-links F-

actin, but also acts in cells as a template to polymerize actin into orthogonal interconnects.

This hypothesis comes from recent work which shows that filamin has a secondary site of

actin binding in N-terminal domains; and as a dimer, filamin monomers have relative

orientation of ~90° degrees (18, 62).

Heterologous filamin-null cells suggest a critical role for filamin A in morphology. *In

situ*, melanoma cells with diminished levels of filamin A (M2 cells) have defects in motility,

cell shape, and membrane blebbing. Also, filamin A is critical to the disc morphology of resting platelets (63). Recently, primary filamin B-null mutant fibroblasts were shown to have significantly impaired motility (64). Similarly, filamin C-null mutant myocytes have an aberrantly rounded phenotype (65). Thus, it has been postulated that all filamin isoforms play critical structural roles.

As mechanical proteins, filamins play important roles in cell structure *in vivo*. Different isoforms express in distinct tissue. Filamin A and B are ubiquitously expressed and filamin C is expresses in muscle. All isoforms are localized to stress areas of cells. Non-muscle filamin localization includes the leading edge of motile cells (66) and stress fibers (67, 68). Filamin C is localized to the Z-disc of myotubes (39). Splice-variants cause variations in subcellular localization. For example, the filamin B H1-ΔVar1 splice-variant exclusively localizes to the tips of stress fibers in developing muscle cells (21).

1.1.2 Filamins as scaffold proteins

Actin-filamin gels are localized to the submembranous cortex of a cell, where filamins act as scaffold proteins by linking transmembrane proteins to downstream signaling cascades (25). Filamins bind to a diverse set of proteins (Figure 1.3) (4, 69). These include actin remodeling associated proteins and their effectors including GTPases (Rac, Rho, Cdc42, and RalA), GTPase-activation proteins (e.g., FILGAP) and GTPase exchange factors (trio) (51, 53, 70). GTPases have been implicated in many aspects of cell physiology, including proliferation, differentiation, cytoskeletal organization, and vesicle trafficking(71-73). GTPases are also emerging as important molecules in cellular response to exogenous and intracellular force by coordination of cytoskeletal rearrangements and gene expression

(74, 75). Other binding partners include G-protein coupled receptors (e.g. dopamine D2

receptor) (25), ion channels (e.g. cystic fibrosis transmembrane conductance regulator) (27),

transcription factors(e.g., MEKK4)(35), kinases (e.g., cAMP-dependent protein kinase

(PKA)) (76-80), phosphatases (e.g., PEST-motif containing protein tyrosine phosphatase)

(38, 81), vesicle trafficking associate proteins (e.g., BIG2)(28), and proteases (e.g., calpain)

(82-84).

Filamin A has been shown to mediate protein trafficking (27, 82) and protein

localization (23), and it may create signaling subdomains on the plasma membrane (85).

Filamins also bind adhesion proteins, including $\beta 1$ integrin family members B_{1A}, B_2, B_3, and

B_7 (21, 47, 86), and the GPIbα subunit of the von Willebrand receptor (87-89). The

interaction of filamin with adhesion proteins is associated with cytoskeleton anchoring (90).

Adhesion proteins may also localize filamin in cells (21).
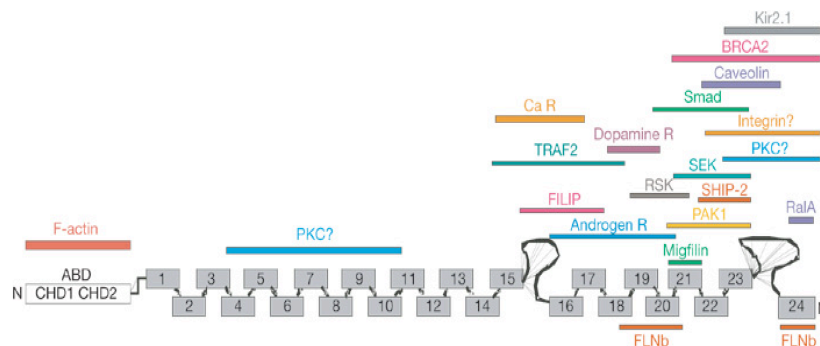
Figure 1.4. Filamin as a scaffold protein (4). As a scaffold protein, filamin interacts with transmembrane proteins, such as receptors and ion channels. Filamin also interacts effector proteins such as kinases and GTPases. Most proteins interact with filamin on domains 16 and above. Binding partners bind specific Ig-like domains; for example, the two best-characterized domains are domain 17 binding to the von Willebrand receptor subunit GPIbα (2), and $\beta 1$ integrin family members binding domain 21 (5).

Filamin is both a substrate and binding partner to several kinases. For example,

PKCα is believed to bind to both the N- and C-terminus domains of filamin and has activity

on the C-terminus region (78). Phosphorylation of filamin is important for its activity, including migration and actin reorganization (79). Many other kinases phosphorylate filamin *in vitro*, including cyclic Ribosomal S6 Kinase, PKA, and CaM kinase II(79). Filamins are also phosphorylated *in vivo* (91, 92). Sites that are phosphorylated range the entire length of the protein (91, 92).

With the exception of kinases, the binding partners of filamin bind almost exclusively to specific Ig-like domains in rod 2 (Figure 1.4). Additionally, a generalized binding mechanism has been identified as β sheet extension of β strand C (2, 5, 49, 93). Many of the binding partners of filamin have been found in interaction assays with Filamin A. It is not known to what extent these binding partners also bind to filamin B or C.

1.1.3 Filamin as both scaffold and mechanical proteins

Several reports suggest that filamin acts as both a scaffold and a mechanical protein in several stress pathways. These include mechanoprotection (58, 61), extracellular matrix (ECM) compliance matching (57), and cell-fate determination (55, 94).
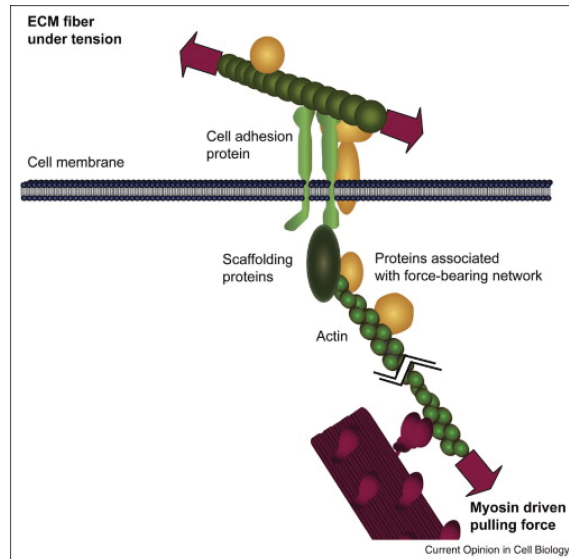
Figure 1.5. ECM compliance matching(3). Extracellular matrix (ECM) stiffness is matched to intercellular stress by acromyosin motor proteins that that pull on actin. F-actin is connected to scaffolding proteins that are connected to cell adhesion proteins.

Mechanoprotection is the process by which cells reinforce their cytoskeleton in response to high tensile exogenous force. Several investigators showed that filamin A accumulates at cell adhesion sites in response to mechanical tension, suggesting a role for filamin A in mechanoprotection (58, 95). Shifrin et al. showed that FILGAP is targeted to filamin A in response to exogenous force applied through β1 integrins (61). There, FILGAP deactivates Rac, leading to mechanoprotection.

Another stress related pathway associated with filamin is ECM compliance matching (Figure 1.5). This is the process by which cells match internal stress with ECM stiffness. Gehler et al. showed that filamin A acts as a stress sensor in ECM compliance matching through a feedback loop that matches extracellular stiffness with intracellular stress. They further showed that filamin A is the structural linkage connecting ECM-bound integrins to acromyosin contractility (57). It has been suggested that involvement of filamin in ECM

compliance matching is related to its scaffold activity, since Filamin A binds GTPase effector proteins that regulate acromyosin contractility (53). Furthermore, Filamin A is reported to regulate myosin-mediated contractility *in vitro* (96, 97).

ECM compliance matching is also a mechanism by which cells use environmental cues to mediate differentiation (98). The role of filamin in ECM compliance matching associated with differentiation could be related to its scaffold activity, since filamin binds several transcription effector proteins (e.g. androgen receptor(32, 33), SMADS(35)) as well as mitogen-associated kinases(80). Furthermore, filamin has been shown to play a critical role in differentiation of breast epithelial cells (filamin A) (57) and maturation of chondrocytes (filamin B) (55). Thus, filamin is believed to be involved in the pathways involving both extracellular and intracellular stress, and the relationship of filamin with scaffold proteins likely regulates transcriptional activity associated with differentiation.

The mechanism by which filamin operates in stress pathways is not yet identified. One hypothesis is that filamin activity is indirectly mediated during stress pathways. In this scenario, stretch-activated calcium channels mediate intracellular calcium levels (99). Calcium levels then mediate both the filamin turnover (78, 100) and F-actin binding (56, 81, 100, 101).

Other studies suggest that the role of filamin A as a scaffold protein is modulated by the quaternary structure of its rod 2 region (93). For example, the crystal structure of domains 19-21 reveals domain-domain interactions resulting in the auto-inhibition of β1 integrin binding (93). It is believed that stress may regulate the quaternary structure of these domains, resulting in enhanced binding of integrins.

1.1.4 The role of filamin *in vivo*

Accumulated evidence suggests a critical role for filamin in cells (12, 17, 102). However, most filamin-null mutant cells from knockout animals appear unaffected by loss of filamin (55, 64, 65, 103-105). Since filamins are highly similar, it is suspected that they have overlapping roles. However, the phenotypes of these knockout animals are considerably different. Filamin A knockout mice have midline defects (103, 104), disorganized heart and endothelial cell gap junction defects, and they die prematurely of bleeding. Filamin B knockout mice have, both microvasculature and skeletal defects, linked to chondrocyte maturation defects (55, 64, 105). Filamin C knockout mice have lethal muscle defects (65).

Furthermore, disease-causing filamin mutations suggest that isoforms have both overlapping and distinct functions (Table 2). Loss-of-function mutations in filamin A result in periventricular nodular heterotopia (PV) (106). Both loss-of-function mutations in filamin B (36, 105) and surface exposed missense mutations in many domains of filamin A and B lead to a spectrum of similar skeletal disorders (107, 108). Deletion of the last 8 C-terminal residues in filamin C leads to late-onset muscular dystrophy (43, 109). Loss of filamin and hence loss of scaffolding of interacting partners (SMADS, BIG2) has been suggested as critical by other studies (28, 36, 110). From these studies we conclude that filamin A is critical to heart, brain, and bone development; filamin B is critical to developing chondrocytes; and filamin C is critical to the function of muscle. However, filamin-associated disease mutations in N-terminal domains, and lack of binding partners in those domains suggest that there is much more to be understood about the function of this protein family.

Table 2. Diseases associated with filamin isoforms (I)

Name	I	Typical manifestation	OMIM
Periventricular nodular heterotopia (PV)	A	Male lethal, Nodules on lateral ventricle of cerebral cortex containing late stage neurons that failed to migrate radially	300049
Otopalatodigital syndrome types 1 (OPD1)	A	Deafness, cleft palate, facial malformations, wide thumbs, long metacarpals, severe scoliosis, and dislocations of the hip and knee	311300
Otopalatodigital syndrome types 2 (OPD2)	A	OPD2 more severe then OPD1.Microcephaly, retardation, malformed skull, ribs, clavicles, scapula, pelvis, curved long bones	304120
Frontometaphyseal dysplasia (FMD)	A	Overgrowth of frontal facial bone	305620
Melnik-Needles syndrome (MNS)	A	Malformed skull, craniofacial structures, irregular ribs, deformed clavicles, scapula, pelvis, and curved long bones	309845
Prolapsed valve	A	Myxomatous degeneration of cardiac valves	
Chronic idiopathic intestinal pseudo-obstruction (CIP)	A	Intestinal obstruction without any mechanical lesion	300048
Bernard-Soulier syndrome (BSS)	A	Excessive bleeding and large fragile platelets	131200
spondylocarpotarsal synostosis (SCT)	B	Disproportionate short stature, scoliosis and lordosis, carpal and tarsal fusion, hearing loss, midline cleft palate	272460
Larsen syndrome (LS)	B	Dislocations of the hip, knee, and elbow; scoliosis and cervical kyphosis; prominent forehead, midline cleft palate, and hearing loss	150250
Atelosteogenesis types I (AOI)	B	Short-limbed dwarfism, LS dislocations, and lethal	108720
Atelosteogenesis types I (AOIII)	B	Short-limbed dwarfism, LS dislocations	108721
Boomerang dysplasia (BD)	B	Dwarfism, short, bowed, rigid limbs, lower spine, ossification, and digits retarded	112310
Filaminopathy	C	Late-onset muscular dystrophy	609524

1.2 Isolation of filamin function using evolution

Much of our accumulated knowledge about the function of filamin is derived from using the yeast two-hybrid screen that identifies filamin interaction with a subject protein. Functional analysis is usually performed in melanoma cells, which lack filamin A. Unfortunately the limitations of current methods, in our opinion, have led to a biased sampling of interacting partners and a bias towards C-terminal binding. Some evidence for these biases is suggested by the presence of kinase activity and a number of disease-causing mutations in filamin rod 1 but the general absence of binding partners there. In addition, cancer cell lines have regulatory defects in GTPase signaling, resulting in aberrant locomotion and morphology (111).

To avoid the current biases in filamin research, it is important to use a top-down approach to understand the function of filamin. Based on vertebrate isoform similarities in sequence and function, we hypothesize that filamin isoforms have diverged from a common ancestor, and those isoforms have maintained shared function, developed new function, or modified existing function. Therefore, it is possible to understand the function of individual filamin isoforms by analyzing site-specific divergence within the filamin family of proteins.

1.2.1 Analyzing functional divergence

Two distinct methods, evolutionary rate-based and evolutionary trace (ET), have been developed to derive functional significance from evolution of protein families (112). Both methods depend upon first deriving the common ancestor, then reconstructing all ancestral filamin forms at clade branch points of the phylogenetic family tree. The rate-based methods discriminate functional vs. non-functional sites by assessing site-specific rate shifts between

13

different subsets of homologous genes. For example, a site would be considered functionally informative if it were stable in one isoform but variable in another. A rate-based approach has been used on Src homology domains (113-115). Since the rate-based approach depends upon the resolving power of statistical tests, it is not well suited for small families such as filamin.

1.2.2 Evolutionary trace

The evolutionary trace (ET) method discriminates functional vs. non-functional sites by screening for site-specific change in amino acid biophysical properties. Since conserved function is usually associated with conserved biophysical properties at particular sites, a shift in residue biophysical property is usually associated with a shift in function. Function could be related to a modulation of a protein-protein interface or a structural change. Additionally, a biochemical shift could occur by chance if the residue is functionally unconstrained. The ET method can distinguish functional vs. nonfunctional shifts based upon rate of change by assuming that functionally unconstrained residues have a higher observed mutational rate. A biochemical shift in a particular clade is also referred to as either a "class-distinctive" or "distinctive " change. ET has been used on protein families, including TGF-β(116), GPCRs (117), and norovirus viral capsid (118). Often an ET is used in combination with a protein structural to provide a spatial context for analyzing functional divergence.

In this work we use a highly constrained ET method (Figure 1.6) to increase the power to discriminate functional vs. non-functional sites. In our implementation of the ET method we exclude from analysis those sites that are not functionally constrained throughout evolution. We also exclude residues if they show signs of being functionally unconstrained

14

due to high variability. Thus, in our implementation of the ET method we try to elucidate only functionally informative sites. We also improve upon traditional ET methods by excluding highly variable sites, whose functionality is indeterminate.
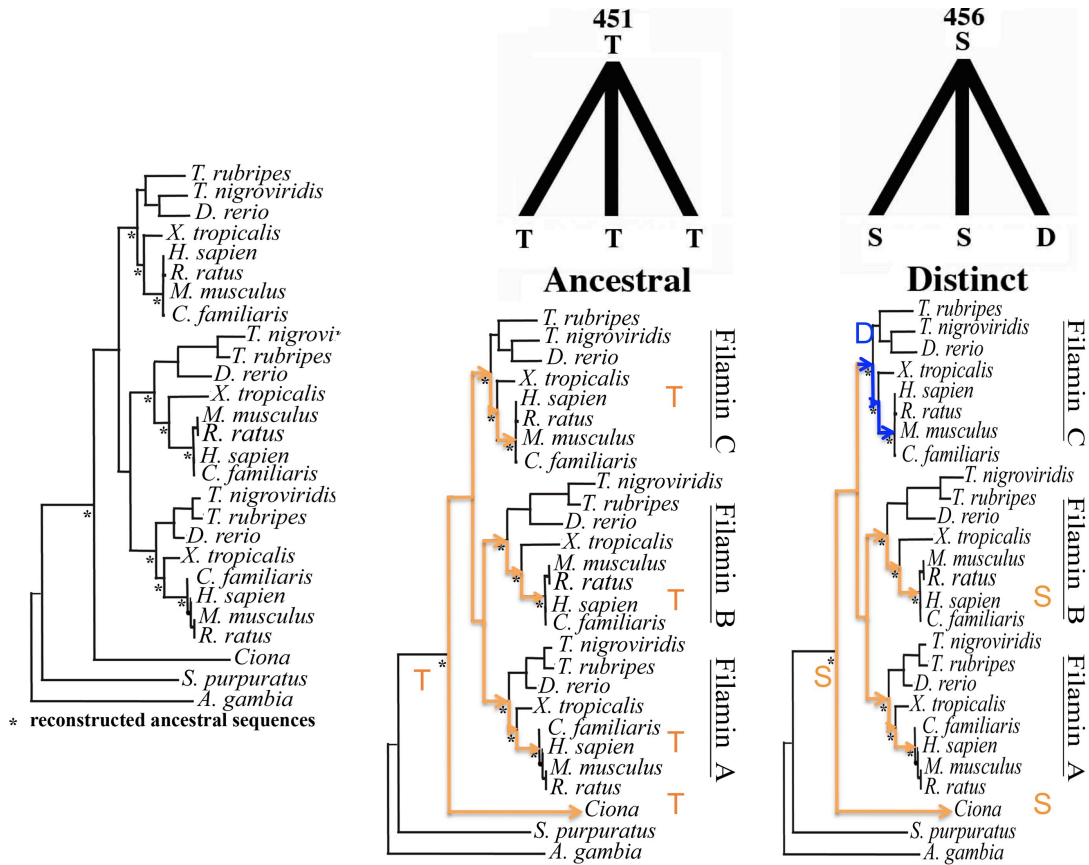
Figure 1.6. Evolutionary trace method. In the evolutionary trace method ancestors are reconstructed at clade branch points. Then specific sites are traced from the common ancestor to human. If there is no change in residue type (site 451) then the site is considered "ancestral." If there is a change in residue type (site 456) and it becomes fixed in a particular isoform, we call this a "distinct" residue in an isoform. Note we require that one isoform remains ancestral for the site to be considered functional.

1.2.3 Results from evolutionary trace

Functional divergence: To study functional divergence of filamin, we performed a phylogenetic analysis of invertebrate, invertebrate chordate, and vertebrate filamin using Bayesian methods. Then we used ancestral reconstruction to derive ancestors at clade branch

15

points of the phylogenetic tree. We used these ancestors to perform our highly constrained ET method. In addition, as part of the ET analysis, we also recorded the period (Teleostei, Amphibian, or Mammalian) in which distinctive divergence of a site in an isoform occurred so that we could discriminate early vs. late divergence. We also examined well-characterized domains associated with critical filamin function, the ABD, and Ig domains 17 and 21.

Subsequently, we derived an all-atom homology model of filamin A to provide a spatial and temporal context for functional divergence. Our homology model of a filamin A monomer was created as a sum of its parts. We first derived all-atom homology models for individual domains using structural templates from homologous Ig-like domain. Then we connected domains using the relative orientation of a pair of ddFilamin Ig domains as a guide. We then mapped onto the structure both conserved and distinctive divergent residues, including an indication of period (Teleostei, Amphibian, or Mammalian) in which divergence occurred. We also analyzed the structure for functional subdomains by observing large regions that have not diverged from the common ancestor or have diverged in just one residue.

Structure: During homology modeling we identified two types of Ig-like domains, six β stranded and seven β stranded. The latter forms the rod I region as a linear arrangement of domains. The six β stranded domains we identified are domains 18 and 20, which cause rod 2 to form a compact region. The resultant hormonology model is very similar in size and features to low resolution electron micrographs, including the compact C-terminal region.

Phylogeny: The filamin family diverged from a common single ancestral gene between urochordate and vertebrate lineages. Filamin isoforms diverged the most during early vertebrate evolution in the period prior to the branching of the Teleostei lineage off the Chordate phylogenetic tree. Filamin C diverged the least from the common ancestor.

Evolutionary trace: From current reports, most filamin functionality is attributed to rod 2, but we did not observe a significant difference between rod 1 and rod 2 with regard to functional divergence. In our examination of well-characterized domains, ABD and Ig domains 17 and 21, we found that although these domains are highly conserved, there are subdomains that have distinctive changes, suggesting that isoforms have diverged with regard to specificity for binding partners or regulatory function. We also observed the potential for regulatory differences among isoforms, since 34% of functionally divergent changes involve a change to or from a serine residue in filamin A and C, but only 11% of such changes in filamin B. There appeared to be a quiescence of divergence towards C-terminal domains of rod 2 (filamin A 21-24), suggesting the possibility that binding interactions in these domains predate vertebrate evolution (integrins for example). Some isoforms have significantly more distinctive residues in some domains than others (e.g. filamin B domains 6 and 15), suggesting that critical divergent events occurred in these domains. Furthermore, we observed co-localization of distinctive residues with missense

17

mutations in both the ABD and some Ig-like domains in rod 1. These suggested the importance of these regions in functional divergence of isoforms.

1.3 Isolation of filamin function using mechanical techniques

Based on studies of the role of filamin in cellular stress pathways, we hypothesized that stress-induced conformational changes in filamin A Ig domains play a direct role in signaling by disrupting existing interactions or by introducing new ones. Therefore, it was important to investigate possible stress-induced conformational changes that could play functional roles.

1.3.1 Single molecule unfolding

Atomic force microscopy (AFM) has been commonly used to study the response of a protein to mechanical stress at the single molecule level. The most common AFM mode is called constant velocity (CV). In CV mode, domains are attached to a substrate. Next, a cantilever, which measures force, is attached to the protein. Then the molecule is pulled apart at constant velocity. Force builds up until the molecule quickly unfolds. The force at this point is considered the critical unfolding force (F_c). Filamin A has been studied using CV AFM and has an F_c in the range of 50-220 pN (16). AFM also has been used in constant-force (CF) mode. CF pulling produces an end-to-end extension vs. time curve. The use of CF mode is a good technique to identify stable intermediate conformations of proteins pulled apart. As molecules are pulled apart they can stabilize into low energy conformations. A stable intermediate is observed when there is an extended period in which its end-to-end length stops increasing.

1.3.2 Molecular dynamics

AFM does not produce an atomistic description of unfolding trajectories, so molecular dynamic simulations (MD) have been used to complement experiments. The use of MD to analyze the forced unfolding pathways of domains has traditionally been too computationally expensive to simulate length of time required to observe unfolding. Strategies implemented to avoid this limitation include using simplified protein models (119, 120), or pulling with high forces or speeds. One major drawback of these strategies is that unfolding pathways can be altered under extreme conditions (121).

Table 3. Ig or similar domains with force-induced stable intermediates

Domain	Unfolding force range	Intermediate conformation
TNI27	100-150 pN	Loss of AB region contacts
FN-III$_{10}$	60-100 pN	Loss of β strands A and B, β sheets slid past one another
ddFilamin 4	50-60 pN	Loss of B strands A and B

Several AFM and MD studies of Ig or similar folds (Table 3) indicate that Ig domains can have force-induced metastable conformations. These include titin Ig-like domain 27 (TNI27)(122-130), type III fibronectin domain 10 (FN-III$_{10}$) (131, 132), and *D. discoideum* domain 4 (ddFilamin 4)(133-135). However, only fibronectin FN-III intermediates are reported to have functionality (132, 136). When cells are cultured on ECM they spread. Using their acromyosin contractile apparatus and their integrins, spreading cells pull on protein components of the ECM, including fibronectin. The forces generated on fibronectin can cause its domains to unfold partially to a stable intermediate, exposing cryptic integrin binding sites (136, 137), which mediate cellular adhesion.

1.3.3 Discrete Molecular Dynamics

The Discrete Molecular Dynamic (DMD) simulation technique (138) is used to sample protein conformational space efficiently (138). DMD is a special variant of the molecular dynamics methods where continuous potential functions are simplified as stepwise functions, thus reducing DMD into collision-driven dynamics. Such an approximation enables DMD to sample protein conformations efficiently with at least three orders of magnitude increase in sampling efficiency over traditional molecular dynamics (138). Over the past years, our lab has developed a hierarchy of simplified protein models, from simple one-bead-per-residue to all-atom representations. Since simplified protein models are not as accurate as all-atom models, the two types of models can be coupled to increase the accuracy of molecular modeling without compromising computational efficiency. Multi-scale (MS) simulations can be achieved by iteratively performing simulations using high- and low-resolution protein models. This idea represents the fundamental approach of multi-scale modeling; low-resolution models are used to sample the conformations of large-scale processes and systems, while higher-resolution models are used for improving accuracy. We have successfully applied MS-DMD methods to study dynamics of large molecules including vinculin (139), dynein (140), nucleosome (141, 142), and cystic fibrosis transmembrane conductance regulator (143, 144).

1.3.4 Results from mechanical techniques

Physical stress on filamin: To study force-induced unfolding of filamin A using DMD we used all-atom models of Ig-like domains derived during the functional ET study. For each of the 24 Ig-like domains that make up filamin, we performed constant-force simulations by applying a set of physiologically relevant ranges of forces (0-315 pN) to the N- and C-terminal Cα atoms of each domain. Based on these systematic DMD simulations, we estimated the critical force of unfolding of each individual domain and characterize the conformational changes under mechanical stress. We found that the critical forces are heterogeneous among the 24 domains due to the sequence differences. We also found that there is no significant difference between the critical unfolding forces of domains with and without N-terminal strands. We found that under small forces up to ~35 pN, which correspond to the physiological forces (145), most Ig-like domains remained in their native states. As the force increased up to the intermediate force levels of ~70 pN, large conformational changes take place. Interestingly, Ig-like domains featured a common initial conformational change, where the first β strand unfurls. Domains that lack their N-terminal β strands appeared similar to one another, maintaining their native-like conformation under low forces. At intermediate forces, they unfolded to a heterogeneous population of intermediate states similar to seven-stranded domains.

Thermal stress on filamin: Additionally, the effect of temperature versus mechanical stress on unfolding pathways has been the subject of much debate (146). The general hypothesis is that thermally- and force-induced unfolding pathways are independent of each other since the thermal fluctuations are exerted on the protein globally, while the effect of mechanical force is localized and non-homogeneous. To examine whether the thermally- and force-induced

21

unfoldings are different, we also performed thermal unfolding of three Ig-like domains, 14, 21, and 24. We found that thermally-induced unfolding shares the same initial unfolding intermediate state as forced unfolding, where the first strand unfolds. This observation suggested that filamin evolved to feature the intermediate state of unfolded N-terminal strand(s). Given the observation that the putative first strand of domain 20 interacts with domain 21 instead of forming a β sheet with the second strand, we proposed that an N-terminal strand may act as a conformational switch that unfolds under stressed physiological conditions leading to exposure of cryptic binding sites, removal of native binding sites, or modulating the ternary structure of domains.

In Chapter 2 we used evolutionary trace method to create hypotheses about the functional importance of residues within the vertebrate filamin family. Our method improved upon the traditional methods by adding constraints that eliminate highly variable sites. These constraints limit hypothesis generation to only those sites in which at least one isoform is conserved throughout evolution. Our work suggested the possibility that filamin isoforms are regulated distinctly both in F-actin binding and in interacting with other proteins such as integrins. It also suggested the possibility of functional roles for N-terminal domains for which few binding partners have been identified to date. In Chapter 3 we analyzed the function of filamin as a mechanical protein using DMD simulation. We obtained mechanical properties of filamin A, including Ig-like domain critical unfolding forces, and unfolding pathways at biological levels of force. In addition, we analyzed how Ig domains react to thermal stress. Our results suggested the possibility that filamin acts as a conformational switch because at intermediate levels of force (35-70 pN), domains unfold to form a heterogeneous set of conformations, many lacking β strands A or B.

Chapter 2
Isoform divergence of the filamin family of proteins

This chapter was submitted as a part of a manuscript to *The Journal of Molecular Biology and Evolution*; and is formatted in the journal style.

2.1 Introduction

Vertebrate filamin is a family of F-actin cross-linking proteins with members named A, B, and C (147). They have a dual role in cells as both mechanical proteins and scaffold proteins. As mechanical proteins, filamins participate in dynamic shaping and maintenance of the cytoskeleton, cross-linking F-actin into gels and fibrils. They also have a role in the physical stress-response and mechanoprotection of cells (58-60). Filamins also have a role in cellular cytoskeletal anchoring by binding the tails of transmembrane proteins such as GPIbα, integrins, channels, and receptors (45-47, 87, 88, 148). Filamin proteins also bind to cytosolic effector proteins and act as scaffold proteins for both inside-out and outside-in signaling, including F-actin remodeling. Filamin family members are greater than 64% similar and have both overlapping and distinct roles. We know much about filamin A, but less about filamin B, and C. Most knowledge has been gained by studying interacting partners and not through filamin centric research. Thus, the proteins studied have biased the knowledge we have about filamin. In this chapter we take a top-down approach to studying the function of filamin. We assume that family members diverged from a common ancestor and that subsequent gene duplication events created all three family members. The questions of interest are: which of the ancestral functional roles of filamin, after gene duplication, are

maintained across all isoforms, distributed between different isoforms, or are newly evolved within an isoform? To answer these questions we use phylogenetic analysis, an evolutionary trace, and an all-atom structural model of filamin A. We analyze spatial and temporal patterns of filamin family divergence to give us insights into the structural elements involved in both isoform-distinctive and common function. The three components of this analysis are development of a structural model of filamin A; a phylogenetic analysis of invertebrate, invertebrate chordate, and vertebrate filamin paralogs; and an evolutionary trace analysis, all three of which we use to observe the divergences of individual filamin sites in a spatial and temporal context.

2.2 Materials and methods

We first use homology, modeling and superimposition, to create an all-atom model of filamin A. Then we use Bayesian phylogenetic methods to derive a phylogenetic tree of the vertebrate filamin family of proteins. We then analyze the site-specific divergence of filamin isoforms by performing a highly constrained evolutionary trace.

Modeling

An all-atom model of filamin A is created from homology models of individual filamin A domains. Most domains are joined by relative positioning pairs of domains onto a multi-domain structure of ddFilamin (PDB ID 1wlh) and then loops are generated to connect domains.

Homology modeling is performed using the Modeler module of the *InsightII* from Accelrys Inc. (www.accelrys.com). Homology modeling uses a template structure from a

related protein to predict a structure for a target protein sequence. The homology models are evaluated for sequence-structure compatibility using the *Verify-3D* function of the Profiles-3D module from *InsightII*. *Verify-3D* scores residues based upon how well they fit in their local environment. *De novo* loop generation is performed in *SYBYL* computational informatics software for molecular modeling from Tripos Inc. (www.tripos.com). To avoid residue clashes and approximate native structures, we relaxed pairs of linker-connected domains with *DMD/Eris/Medusa* (149-151). To create a filamin dimer we superimpose each self-association domain (24) onto a structure of a dimer of IgFLNc24 (PDB ID 1v05) using *PyMOL*.

Actins-binding domain (ABD)

The ABD of filamin is homology modeled using a template of ABD of α actinin in the closed conformation (PDB ID 1wku) (152). The protein threading server *HHPRED* (153) is used to identify suitable templates for the ABD of filamin A. The top hit is to the ABD of α actinin with an E value of 0. The architecture of filamin and the ABD of α actinin are similar except for a longer negatively charged linker between first and second CH domain of filamin (Figure 1 A, Appendix A).

Ig-like domains

Filamin A Ig-like domains are modeled using templates composed of homologous Ig-like domains of human and *Dictyostelium discoideum* filamin (ddFilamin), except for the tri-domain structure of domains 19-21 (PDB ID 2j3s). A multiple sequence alignment (MSA),used for template assignment, is derived from a structural alignment of the human

and ddFilamin Ig-like domains in Figure 1 B, Appendix A. A template for each target domain is chosen based upon its having both a sequence and an insertion pattern that are similar to the query. Table 1 in Appendix B contains the paired target/templates we use.

Domain model validation

Homology models of filamin domains generally have high sequence-structure compatibility scores (Table 2, Appendix B). Quality analysis of each domain is performed using *Verify-3d* module of *InsightII* (Accelrys Inc.). This analysis measures the compatibility of modeled residues with their structural environment. A quality score of greater than 0.1 indicates a valid structural model with a correct fold (154), while a higher score indicates a more accurate model. Typical experimental structures score around 1.0. The mean *Verify-3d* score we obtain for Ig-like domain models is 0.89 with standard deviation 0.30. The score we obtain for the ABD is 0.75. The only model to produce a low score is IgFLNa18 (0.11). However, the homology model of IgFLNa18 has high sequence identity (64% id), and a low RMSD (0.23 Å) when compared to its IgFLNb18 template, suggesting that the IgFLNa18 model is correctly built based on the template. Furthermore, the structure of the template itself has a low quality score of 0.56. Overall, the high mean quality score we obtain for the Ig-like domain models infers their high accuracy. Recently, the crystal structure of the filamin B ABD was released (PDB ID 3fer). This structure and our homology model have a 1.1 Å RMSD, over 168 Cα atoms. The largest architectural difference between our model and this structure is the positioning of a highly ancestral set of residues that lead to the neck region (linker between CH 2 and IgFLN 1). This may have the effect of modifying the relative position of the ABD and IgFLN 1.

Joining domains

We oriented domains relative to one another using the multiple-domain structure of ddIgFLN 4-6. Ig-like domains with short linkers (mean linker length 3±1 (Table 2, Appendix B) have an extended conformation (Figure 2.2, domains 6-8) (62) and have been proposed to have the same relative conformation as ddIgFLN 4 and 5 (PDB ID 1wlh) (155). Popowicz et al. modeled extant ddFilamin using a similar technique (155). Domains surrounding hinge regions are first oriented similar to that of short linkers and then the distance between domains is extended to accommodate the larger linker. An adjustment to the orientation of domains around hinge 2, IgFLN 23 and IgFLN 24, is made based upon visual inspection of a structure of these two domains produced by small angle X-ray light scattering (156). After domain orientation, *de novo* loop generation is used to join domains.

During homology modeling we identified several Ig-like domains that are potentially missing a β strand. As a result, we adjusted our method of modeling these domains. Ig-like domains are composed of two β sheets forming a β sandwich - one sheet with four (A-B-E-D) β strands and one with three (G-F-C) β strands. Our analysis of IgFLNb18 (PDB ID 2dmc) and IgFLNb20 (PDB ID 2e9i) revealed six-stranded Ig-like domains with most of β strand A forming a highly dynamic coil. In addition, the sequence of β strand A for these three domains shows low sequence identity with the canonical filamin β strand A sequence, a possible indication of a structural change. We, therefore, modeled Ig-like domains 18 and 20 lacking β strand A. Structurally the loss of β strand A has the affect of placing the N- and C-termini adjacent to one another thus inducing a compact multi-domain structure (Figure 2.2, Ig domains 17-19). Since IgFLN 18 lacked strand A we used *ad hoc* methods to determine its

relative positioning to domain 17. IgFLN 18 and 19 share a short linker so we used ddFLN superimposition to derive their relative positioning. We modeled IgFLN 19-21 using the structure of IgFLN 19-21 (PDB ID 2j3s) (93). Both tri-domain models (17-19, 19-21) have an elbow-like architecture that when incorporated in our structure produced a zigzag conformation.

An analysis of IgFLNb16 (PDB ID 2ee9) also revealed a six-stranded Ig-like domain. However, sequence differences between filamin A and B, in β strand A, suggest potential isoform-specific conformational differences in Ig-like domain 16. The sequences of IgFLNa16, β strand A has several residues that are conserved to the canonical Ig-like domain (Figure 1 B, Appendix A), in contrast to IgFLNb16 and IgFLNc16. Therefore, we modeled IgFLNa16 with β strand A intact. Since the experimental structure of IgFLNb16, lacks β strand A, there is some ambiguity as to whether IgFLNa16 should have been modeled with or without β strand A.

Phylogenetic analysis of the filamin family

Data

Filamin protein sequences (Table 3, Appendix B) are obtained from a search of genomic databases. The general procedure is to obtain protein sequences using tblastn with the query sequence of hsFilamin A on genomic databases. We only use sequences that have both the ABD and a large quantity of Ig-like domains (all but mosquito and sea urchin have the full complement of 24 Ig-like domains). The species included in this study are vertebrates: mammals (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Canis familiaris*), Amphibians (*Xenopus tropicalis*), and teleosts (*Danio rerio*, *Takifugu rubripes*, *Tetraodon*

nigroviridis); urochordates: *Ciona intestinalis, Ciona savignyi*; and invertebrates: *Strongylocentrotus purpuratus, Anopheles gambiae*. Well-characterized sequences for *Homo sapiens* filamin are obtained from NCBI (http://www.ncbi.nlm.nih.gov/) (hsFilamin A (**NP_001447.2**), B (**NP_001448.2**), and C (**NP_001449.3**)). Vertebrate and urochordate sequences are obtained from the   Ensemble (release 41) database (http://www.ensemble.org) and invertebrates sequences are obtained from the NCBI database. Table 3 in Appendix B lists the sequences we use in the study. To obtain a full length *Ciona* filamin protein sequence with the full complement of 24 Ig-like domains we use the consensus sequence from the following predicted proteins (**ENSCSAVP00000017581**,

ENSCSAVP00000017582, **ENSCSAVP00000017578**, **ENSCSAVP00000017579**,

ENSCSAVP00000017577, **ENSCSAVP00000017583**, **ENSCSAVP00000017585**,

ENSCSAVP00000017580, **ENSCSAVP00000017584**, **ENSCINP00000006732**).

Tree construction

Alignments are generated with the assistance of software alignment tools as well as by visual adjustment using both *JalView* (157) and *Seaview* (158). *Mr. Bayes* (159) is employed to calculate all phylogenetic trees and for ancestral reconstruction. The out-group chosen is a mosquito filamin from *Anopheles gambiae* (**AGAP004335-PA**). We use the rate matrix amino acid model equalin, and fixed rate model invgamma, 125000 generations, sample frequency of 100, and three independent runs. A burn-in of 250 is used to create consensus trees as indicated by convergence of test runs. All but one branch possesses a posterior probability of partition of 1.00. This single exception occurs within the mammalian

clade and possessed a posterior probability of (0.93). Tree generation software *njplot* is

employed to render trees (160).

A Ancestral reconstructon

	451	456	459
A. GAMBIA	TVETFSAGKGNVDVAV		
S. PURPURATUS	TVETFGAGSGNVEVLI		
CIONA	TVETVSAGSGEVLVYL		
FLNa H. SAPIEN	TVETRSAGQGEVLVYV		
FLNa X. TROPICALIS	TVETISAGMGEVLVYV		
FLNa D. RERIO	TVETISAGQGEVLVYV		
FLNb H. SAPIEN	TVDTISAGQGDVMVFV		
FLNb X. TROPICALIS	YIVQISIWSSSVLNLV		
FLNb D. RERIO	TVDTFSAGQGQVMVYV		
FLNc H. SAPIEN	TVQTVDAGVGEVLVYI		
FLNc X. TROPICALIS	TVETINAGLGEVLVFV		
FLNc D. RERIO	TVETLEAGLGEVIVYV		

B Distinctive residues analysis

C Period of divergence

reconstructed sequences

FLNa CHORDATE	TVETXSAGXGEVLVYV
FLNa MAMMALAIN	TVETRSAGQGEVLVYV
FLNa AMPHIBIAN	TVETXSAGXGEVLVYV
FLNa TELEOSTEI	TVETISAGXGEVLVYV
FLNb MAMMALIAN	TVDTISAGQGDVMVFV
FLNb AMPHIBIAN	TVXTISAGSGXVLVLV
FLNb TELEOSTEI	TVDTFSAGQGXVMVYV
FLNc MAMMALIAN	TVQTVDAGVGEVLVYI
FLNc AMPHIBIAN	TVETVDAGVGEVLVYI
FLNc TELEOSTEI	TVETVEAGLGEVLVYV

Figure 2.1. Isoform-specific divergence. The analysis of isoform-specific divergence consists of ancestral reconstruction, aligning ancestral sequences, and an evolutionary trace of ancestral sequences. (**A**) Segment alignment of a section of the extant filamin protein sequences. Lower alignment is of reconstructed ancestral sequences. X is indeterminate residue. (**B**) Evolutionary trace results indicating sites are: "ancestral," "distinctive," or "other." Ancestral implies the same biochemical type is maintained in each isoform. "Distinctive" implies change to a new biochemical type in one isoform and then is fixed, while at least one other isoform is ancestral. "Other" implies a variable site. (**C**) The evolution of extant species are viewed as branching off the evolutionary tree that leads to humans. Ancestors at the branch points are Teleostei (blue), Amphibian (green), and Mammalian (yellow).

Isoform-specific divergence

Determining isoform-specific divergence consists of ancestral reconstruction

followed by a site-specific evolutionary trace (ET) as detailed in Figure 2.1. Ancestors are

reconstructed using *Mr. Bayes* at the clade branch points as indicated in this figure. We used

an amino acid rate model equalin with a fixed rate model invgamma, and either 50000 or

10000 generations, both of which converged well. To eliminate ancestral calls with high

uncertainty a consensus of eighty percent of the generations have to agree on an ancestral

residue. We indicated with an X any residue in the ancestral sequence that did not meet this

criteria. Using an alignment of these ancestral sequences we performed an ET. At each site of

this alignment we monitor at periods Teleostei, Amphibian, or Mammalian, for a change in residue biochemical type that then becomes fixed until the human form. We additionally required that at least one isoform remained ancestral. Throughout this analysis we use a reduced amino acid type set consisting of 16 residue types: (S/T, I/L, K/R, D/E, F/Y), 10 remaining single residues and a type for an insertion. We allow the skipping of at most one period to allow for period specific divergence. We also accout for divergence of Ciona by redefining the ancestral type, if all three Teleostei period ancestors are in agreement for residue type but disagreed with the common *Ciona* ancestor. The latter caveat occurred very infrequently.

Information about polymorphisms leading to disease is derived from filamin A (P21333) and filamin B (075369) in the UniProt database (http://www.uniprot.org).

2.3 Results

All-atom model of filamin A

We construct an all-atom structural model of a human filamin A monomer using homology modeling (see Materials and methods) (Figure 2.2). This monomer has an actin-binding domain (ABD), 24 Ig-like domains, linkers and two hinge regions. In general, model construction consists of homology modeling individual domains, measurement of the quality of each model, positioning domains relative to one another, and generating linkers between domains. We then construct a filamin A dimer model by superimposing the 24th Ig-like domain of each monomer onto the X-ray structure of the dimer of IgFLNc24 (PDB ID 1v05).

Model Description

The predicted structural model of filamin A is consistent with recent electron micrographs and experimental data (62) that suggest that filamin A has two distinct architectural regions, one consisting of domains 1-15 and having a linear extension of domains, and the other (domains 16-24) having a compact conformation. The linear region is similar to the linear conformation of filamin domains that are proposed to exist based upon the high-resolution structure of ddFilamin domains 4-6 (155). The compact region results from two domains, 18 and 20, lacking β strand A. Interestingly, early electron micrographs of filamin A (18, 161) have an overall dimension and appearance quite similar to our model.
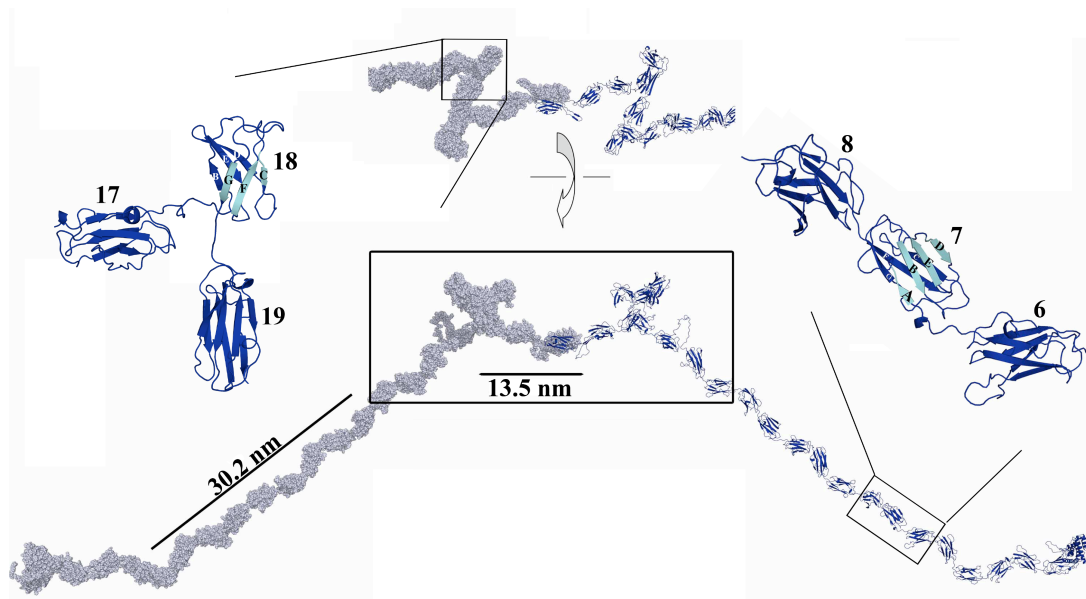
Figure 2.2. Atomic model of a filamin A dimer. Filamin A dimer and component domains are produced by homology modeling. In filamin we propose that there are two types of Ig-like domains, six-stranded because they are missing β strand A (e.g., domain 18) and seven-stranded (e.g., domain 7). We propose that IgFLN 18 and 20 lack their first strand, which results in a conformation such that the N- and C-terminus of the domains coincide, causing a "zigzag" or kink conformation in distal C-terminal domains. The entire dimer is displayed as a hybrid model, a space filling model (left monomer) and a cartoon (right monomer). The structural compaction caused by the kink region results in the length of eight C-terminal domains equal to 15.3 nm compared to the linear N-terminal domains of 30.2 nm. A 103° angle is observed between the filamin monomers and approximates the 90° angle average observed in filamin electron micrographs.

Derivation of vertebrate filamin phylogeny

To derive the phylogeny of vertebrate filamin we perform a Bayesian inference phylogenetic analysis on aligned filamin invertebrate, invertebrate chordate, and vertebrate sequences (Figure 2.3). All nodes of the consensus phylogenetic tree have at least a 90% level of confidence. Within the chordate filamin clade, we observe three monophyletic subclades, each corresponding to a different isoform of filamin. The single ascidian sequence, from the urochordate genus, cluster outside these three vertebrate subclades, while invertebrates are even more distantly related. Therefore, we hypothesize that two gene duplication events occur between the branching of urochordate and teleostei ancestors from the mammalian lineage during early vertebrate evolution to give rise to the three vertebrate filamins.

Our phylogenetic study of vertebrate filamin suggests that the three isogenes of filamin (A, B, and C) originated from a single ancestral gene that existed prior to the advent of vertebrates and that it is also the progenitor to extant urochordate (tunicates) filamin. The tunicate *Ciona intestinalis* has been extensively studied as having a simplified body plan that mimics the ancestor to vertebrates (162). Tunicates have organs and systems in which filamin isoforms have been shown to play a developmentally critical role. These include, a well-developed muscle of their one chambered heart (163), vasculature (163), and immune system that represents the prototype for the innate vertebrate immune system (164). Filamin binding partner and associated pathway orthologs have also been identified in ascidians, including specialized von Willebrand factor type A -like (165) and β1 integrin family –like members (1). The interaction between GPIbα (von Willebrand receptor) and filamin A is critical to platelet activation and adhesion as part of wound healing in vertebrates (2, 166). A recent report suggests that analogs of the von Willebrand factor and many other proteins involved in the vertebrate intrinsic wound healing pathway, do not appear to have a role in ascidian

33

wound healing (167), but play a role in ascidian clot formation during allo-rejection response (167). Thus, it is possible that ascidian filamin have similar protein-protein interactions, but may function in pathways that do not necessarily lead to the same overall biological function.

Figure 2.3. Phylogenetic analysis of the vertebrate filamin family of proteins. Phylogenetic Bayesian analysis of aligned invertebrate and chordate filamin protein sequences suggests that urochordate *Ciona* filamin is closely related to the chordate common ancestor of the three vertebrate filamin isoforms. Gene duplication took place during early vertebrate evolution resulting in three filamin isoforms in Teleostei. The level of confidence in all branches is high. An asterisk (*) denotes branch points in which we derived ancestral sequences for subsequent isoform-specific divergence analysis. (Abbreviations: *Ciona* family: *Ciona Intestinalis* and *Ciona Savignyi*)

Spatial and temporal divergence

The single urochordate *Ciona* filamin ancestor, as the progenitor to the three vertebrate isoforms, gives us a snapshot of parental function and allows us to track, through sequence analyses, the maintenance of parental function and the evolution of neo-functionality in the three isoforms. Bayesian ancestral reconstruction is used to derive the ancestral sequences required for this analysis. Investigating an evolutionary trace for each

34

site in the multiple sequence alignment of reconstructed ancestors is then used to identify

three characteristics. (1) The location and type (see Materials and methods) of an accepted

mutation. (2) The isoform that an accepted mutation occurs in. (3) The ancestor in which the

substitution first occurs. If at a particular site, a reside type is conserved across all isoforms

including the *Ciona* ancestor, then maintenance of parental functionality in all isoforms is

hypothesized to have occurred. We call these residues "ancestral." If, however, at this site at

least one isoform has maintained the residue type of the ancestral *Ciona*, and at least one of

the other isoforms became fixed to a new residue type, this suggests that during evolution

both maintenance of ancestral functionality within one isoform and derivation of neo-

functionality within another isoform occurs. We call these isoform-specific divergent

residues "distinctive." All other residues are called "other." We further annotate the class-

distinctive residues by the period in which the divergence occurs as Teleostei, Amphibian, or

Mammalian, thus a temporal categorization of divergence. Our method of isoform-specific

divergence analysis is more restrictive than other analyses such as the "Evolutionary Trace"

method developed by Lichtarge et al. (168), since we require maintenance of ancestral

functionality in at least one isoform.

Isoform conservation and class-distinctive divergence

We quantify both conserved and class-distinctive residues in the context of the whole

protein. We find that filamin isoforms are 64% similar to one another and the common

ancestor, and 21% of the sites are class-distinctive (Figure 2.4 A). Of these class-distinctive

sites, 78% diverge in the Teleostei period. Class-distinctive changes have few additions in

the Amphibian period and a modest increment in the Mammalian period. The percentage of

class-distinctive sites by isoform is: filamin A 35%, filamin B 43%, and filamin C 22% (Figure 2.4 B). Thus, we hypothesize that filamin B contains the most neo-functionality and filamin C contains the most parental functionality.

Isoform conservation across domains

Quantification of the number of ancestral residues within domains identifies domains that played critical roles in maintenance of parental functionality. We find that conserved residues are not equally distributed implying potential conservation of interfaces in domains also outside the most extensively studied domains 16-24. This is observed from a bimodal distribution (Figure 2, Appendix A) of conservation with the first mode ~55% conserved, and the second mode about ~70% conserved. Between domains 2-15, domains with lower conservation are interspersed with domains of higher conservation. Most domains known to participate in protein-protein interactions (16-24) are contained within mode 2. Surprisingly, a number of domains that currently have little evidence of associated function also have as large a proportion of ancestral residues, suggesting maintenance of ancestral functionality beyond just fold maintenance (e.g., Ig-like domains 3, 4, 9, 12 and 14). It should be noted that non-conserved residues could either be "class-distinctive" or "other."

Isoform class-distinctive divergence across domains

Our evolutionary trace analysis reveals ancestral conservation and isoform class-distinctive divergence down to the spatial granularity of an individual residue and temporal granularity to three periods – Teleostei, Amphibian, and Mammalian. To identify domains that may play a critical role in isoform divergence and the adoption of neo-functionality we quantify their class-distinctive residues (Figure 2.4 C). Since only 22% of the class-

distinctive residues occurred after the Teleostei period, we limit this analysis of divergence to only within the Teleostei period.

The ABD is highly conserved, thus all isoforms share similar ancestral function. Two domains highlighted by this analysis are CH 1 and CH 2 of the ABD. CH 2 with several class-distinctive residues (7, 6, and 1 for A, B, and C, respectively) shows greater isoform differentiation than CH 1 (1, 1, 2 for A, B, and C, respectively). These data indicate that CH 1 is nearly ancestral and has few class-distinctive residues in each isoform, yet CH 2 has a varying divergence pattern with A and B diverging from the common ancestor and C remaining the closest to the parental form. ABD CH 2 has been proposed to play a regulatory role (169), hence it is possible that isoforms differ in their regulation of actin-binding.

Although most protein interactions have been isolated to domains 16 and above, we do not see a clear delineation in divergence pattern between this and the preceding region. This pattern suggests potential limitations in experimental procedure, caused by the extensive use of the yeast two-hybrid method to identify interacting partners with filamin. All isoforms have diverged in almost every Ig-like domain, thus divergence through class-distinctive residues occurred ubiquitously, but with variations in degree, rather than through modes such as localization to a limited number of specific domains. There are however, isoform distinct patterns of divergence and some individual domains are notable for extreme values of class-distinctive residues, a possible indication of isoform-specific functionality. We identify two global patterns. First, in almost every domain filamin A and B have more class-distinctive residues than filamin C. Second, there appears to be two regions of class-distinctive residues over domains. N-terminal domains fluctuate between low and high distinctive-residues counts. C-terminal domains in rod 2 have lower numbers of distinctive residues.
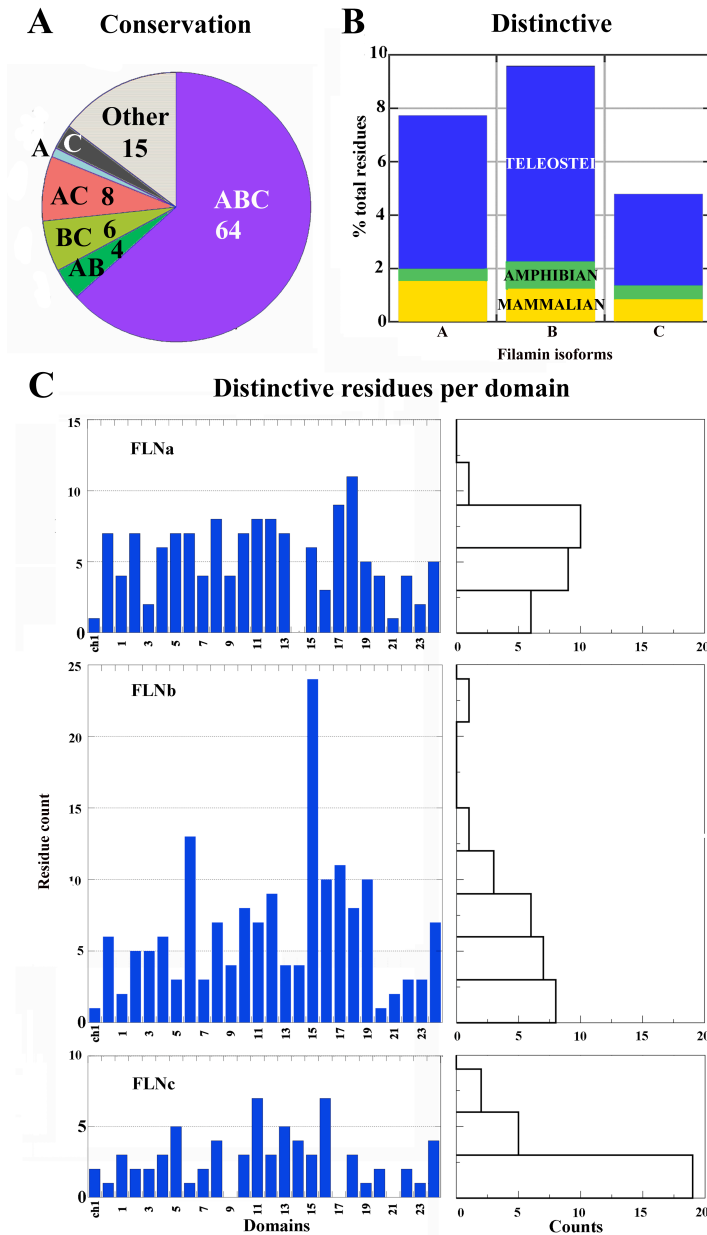
Figure 2.4. Quantifying the divergence of filamin isoforms. (A) Similarities between isoforms and common ancestor. ABC – All three; AB, BC, AC – Two are similar, one distinctive. A, B, C – One similar and one is distinctive; Other – All other sites. Filamin isoforms retained 64% similar residues during evolution. Overall filamin C retained the most similar residues with the common ancestor. **(B)** The number of distinctive residues in each isoform and period (Mammalian, Amphibian, Teleostei), reported as percent of the total number of isoform residues. Most divergence is in Teleostei. Filamin A and B diverge about equally, while C diverge the least. **(C)** Distribution across filamin domains of Teleostei period distinctive residues. Right graph is the histogram of distributions on left. The composition of many individual domains reflects the overall composition (Mean and standard deviation per domain: A 7±3, B 9±5, C 4±2). Domains near the C-terminus have a lower number than proceeding domains (20 and above A, 19 and above B, 16 and above C). Large counts within isoform: A, 17 and18; B: 6 and 15; C: 11 and 16.   Low (1 or 0) counts within isoforms: A, CH 1, 14, 21; B: CH 1 and 20; C: CH 2 and 6, 9, 17, 19, 21, and 23.

These regions differ in location for each isoform (filamin A: 1-20, 21-23; filamin B: 1-19, 20-23; filamin C: 1-16, 17-23). Even within these broad regions there are individual domains with extreme counts of class-distinctive residues. In Filamin A, Ig domains 14 and 21 have very few class-distinctive modifications, and domains 17 and 18, have many class-distinctive residues. In filamin B, domain 20 has hardly diverged, while domains 6 and 15 has extreme counts of class-distinctive residues. In fact, domains 6 and 15, have more class-distinctive residues than any other domain, therefore they may have played a prominent role in developing filamin B neo-functionality. In filamin C, domains 9, 17, and 21 have no class-distinctive residues; whereas, domains 11 and 16, have a relatively large number of class-distinctive residues. Domains 11 and 16 also have a large number of class-distinctive residues in filamin A and B. Thus, in each isoform domains some domains maintain parental function while others have developed neo-functionality.

A structural view of conservation and divergence

To increase our understanding of the pattern of divergences, we place the "ancestral," "distinctive," and "other" residues in the spatial context of our model of filamin A. We also place class-distinctive residues in a temporal context as either from the "Teleostei," "Amphibian," or "Mammalian" period. In addition, to gain further insights into degree of class-distinctive change as well as biological significance, we observe class-distinctive residue relative positioning (such as surface exposed and clustering) while noting the status of surrounding residues, e.g. are class-distinctive residues observed in an ancestral or variable background? Do they co-localize with mutations that lead to disease? A single

accepted mutation in an ancestral background can indicate a minor modification to ancestral function, whereas a cluster of accepted mutations can indicate neo-functionality.

We report our findings for well-characterized domains that have known binding partners in filamin A, IgFLN 21-β7integrin (PDB ID 2BRQ) (5) (Figures 2.5 A and 2.6 A), IgFLN 17-GPIbα (PDB ID 2BP3) (2) (Figures 2.5 B and 2.6 B), and ABD (2) (Figures 2.5 C and 2.7 C). In addition, we examine the spatial distribution of class-distinctive residues for domains in the range 5-10 and 14-18 (Figures 2.8 and 2.9).

Adhesion protein binding domains

During the Mammalian period filamin isoforms distinctively diverge in domain 21 at a site that is relevant to binding the developmentally critical binding partners, integrins. We find one class-distinctive substitution (Figure 2.5 A) along β strand C in the well-characterized ligand-binding interface, an ancestral Ser/Thr in filamin B and C changed to an Ala at residue 2272 in filamin A during the Mammalian period. This class-distinctive site is localized to the end of the ligand-binding pocket. This site is suggested to be critical for binding based on *in-vitro* mutational studies (5) and several NMR perturbation (5, 49, 93) studies. Kiema et al. showed that $A^{2272} A^{2274}$/DK (a phosphomimic) diminished binding, while the substitutions AA/ST (ancestral residue type) did not (5). This result suggests the possibility of isoform distinct regulation of binding integrins, although there is no evidence for this to date.

Domain 17 has two Teleostei period distinct sites in the binding pocket with GPIbα (Figure 2.5 B). On the N-terminal end of the interface, an ancestral Ser (filamin A S^{1899}) to class-distinctive Asp substitution is observed in filamin B. On the C-terminal end, an

40

ancestral filamin C Val to class-distinctive Ile substitution is observed in filamin A and B. The N-terminal class-distinctive site aligns with the class-distinctive site in the integrin-binding interface. Evidence for its critical nature comes from the site-directed-mutagenesis of the adjacent ligand residue (170). The Ser to Asp class-distinctive substitution at this site suggests the possibility of isoform distinct regulation of interactions involving this domain, although there is no evidence for this to date. Class-distinctive sites on domain 17 and 21 also localize outside the well-characterized binding interface (Figures 2.6 A and B), suggesting the possibility of other modes of binding.

Based upon the class-distinctive residue profile of Ig domain 21, we also identify other domains that have a class-distinctive residue at the N-terminus of β strand C (Figure 2.6 C), similar to that of domain 17 and 21. We find that IgFLNc14, IgFLNa19 and IgFLNb19 have the closest match. Of these three domains, IgFLN 19 is most similar to IgFLNa21. In addition, the substitutions of serine/threonine residues in class-distinctive sites appeared to be a recurrent pattern. There is some evidence that ligands bind multiple domains, including domain 19 (2, 5). Therefore, it is possible that isoforms distinctively diverge at analogous sites within different domains because these locations are good targets to modulate binding of the same ligand.

Actin-binding domain

The N-terminal ABD is composed of two CH domains. These domains are connected to the Ig-like domains with a larger linker. The first CH domain is more conserved than the second and is involved in both directly binding F-actin and regulation of F-actin binding, by interacting with calcium-activated calmodulin. The second domain is suggested to have a

regulatory role in binding F-actin (15). Quantification of class-distinctive residues suggests that isoforms have distinctively diverged to a much greater extent in CH 2 than CH 1. Within CH 2, class-distinctive sites are localized to a large surface on helices B, C, E and F (Figure 2.5 C and Figure 2.7 C). The biochemical nature of class-distinctive residues differed by isoform. Class-distinctive changes in filamin A are primarily to serine residues. Most filamin B class-distinctive residues are on helices (except Y^{155}). The only filamin C class-distinctive residue that localizes to this region is a large polar residue that diverged from a charged residue. Transitioning from a more biochemically charged surface to a more uncharged polar surface appears to be the general trend. We also observed mutations in this region that co-localize with class-distinctive residues. Mutations that localize to this region have been reported as "gain-of-function" mutations leading to similar developmental skeletal disorders in both filamin A and B (107). Thus, we have isolated a particular region of CH 2 where all isoforms have distinctly diverged. Furthermore, we show that class-distinctive residues co-localize with mutations causing disease in both filamin A and B. Therefore, we propose class-distinctive residues within this region are important for isoform divergence, and these changes can be involved in pathways that when mutated lead to disease. One possibility is that isoforms distinctly regulate actin-binding, especially since filamin A contains a large number of distinctive serine residues.
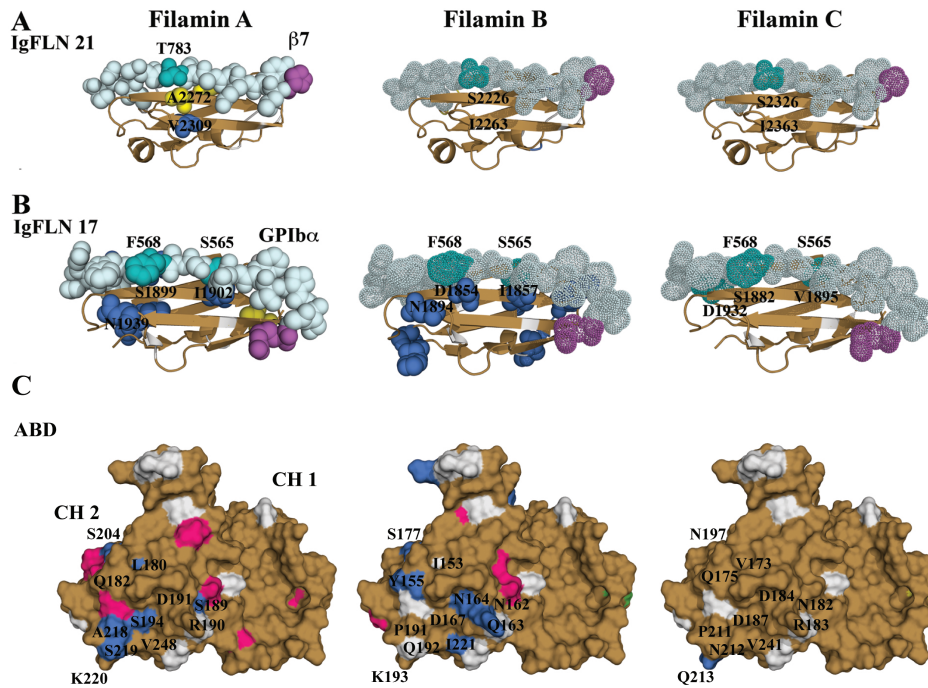
KEY: Ancestral, Distinctive (Teleostei, Amphibian, Mammalian), Disease

Figure 2.5. Analysis of class-distinctive residues in domains with well-characterized interfaces. Each column contains an all-atom model of filamin A, differentially colored based upon isoform class-distinctive divergence (sand = ancestral, class-distinctive (blue = Teleostei, green = Amphibian, yellow = Mammalian), white = other). Ligands are colored in light cyan with residues adjacent to class-distinctive residues in dark cyan. Ligands are space-filled for experimental and dotted for hypothetical complexes. The N-terminus of the ligand is colored magenta. (**A**) IgFLN 21 in complex with β7 integrin. There is one class-distinctive residue in the interface on filamin A. (**B**) IgFLN 17 in complex with GPIbα. There is one class-distinctive residue on filamin A and two on filamin B at the interface. (**C**) Surface model of the filamin A actin-binding domain (ABD), consisting of two calponin homology (CH) domains. CH 1 is highly ancestral, while CH 2 has a few class-distinctive residues that cluster around a few subdomains. Filamin A and B have more class-distinctive residues than C, and those co-localize near mutations that cause disease (pink).
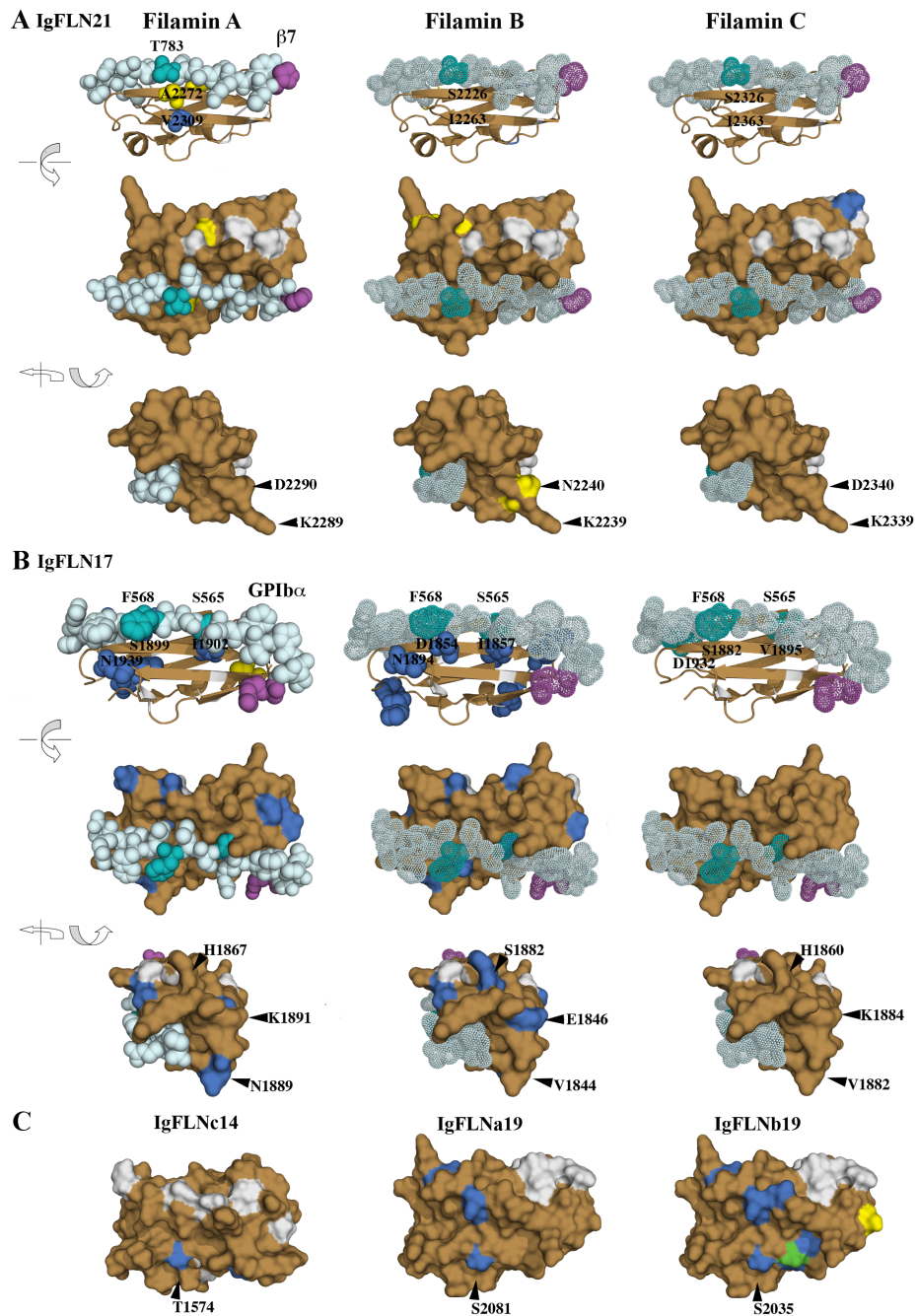
Figure 2.6. **KEY:**Ancestral**, Distinctive (**Teleostei**,** Amphibian**,** Mammalian**) Distinctive divergence outside the binding interface of Ig-like domains 17 and 21.** Coloring as in Figure 2.4. (**A**) IgFLN 21 in complex with β7 integrin (PDB ID 2BRQ). Top row has the same orientation as Figure 2.4. Row 2, 180° rotation, views the CD face of the domain. β strand D is opposite ligand. All isoforms have class-distinctive residues (yellow) surrounded by highly ancestral regions outside the binding pocket. (**B**) IgFLN 17 in complex with GPIbα ligand residues (PDB ID 2BP3). Similarly, domain 17 has distinctive residues outside the binding pocket surrounded by ancestral residues (Filamin A and B only) (**C**) Domains 14 and 19 have a profile of distinctive residues along β strand C that is similar to both IgFLN 17 and IgFLN 21.
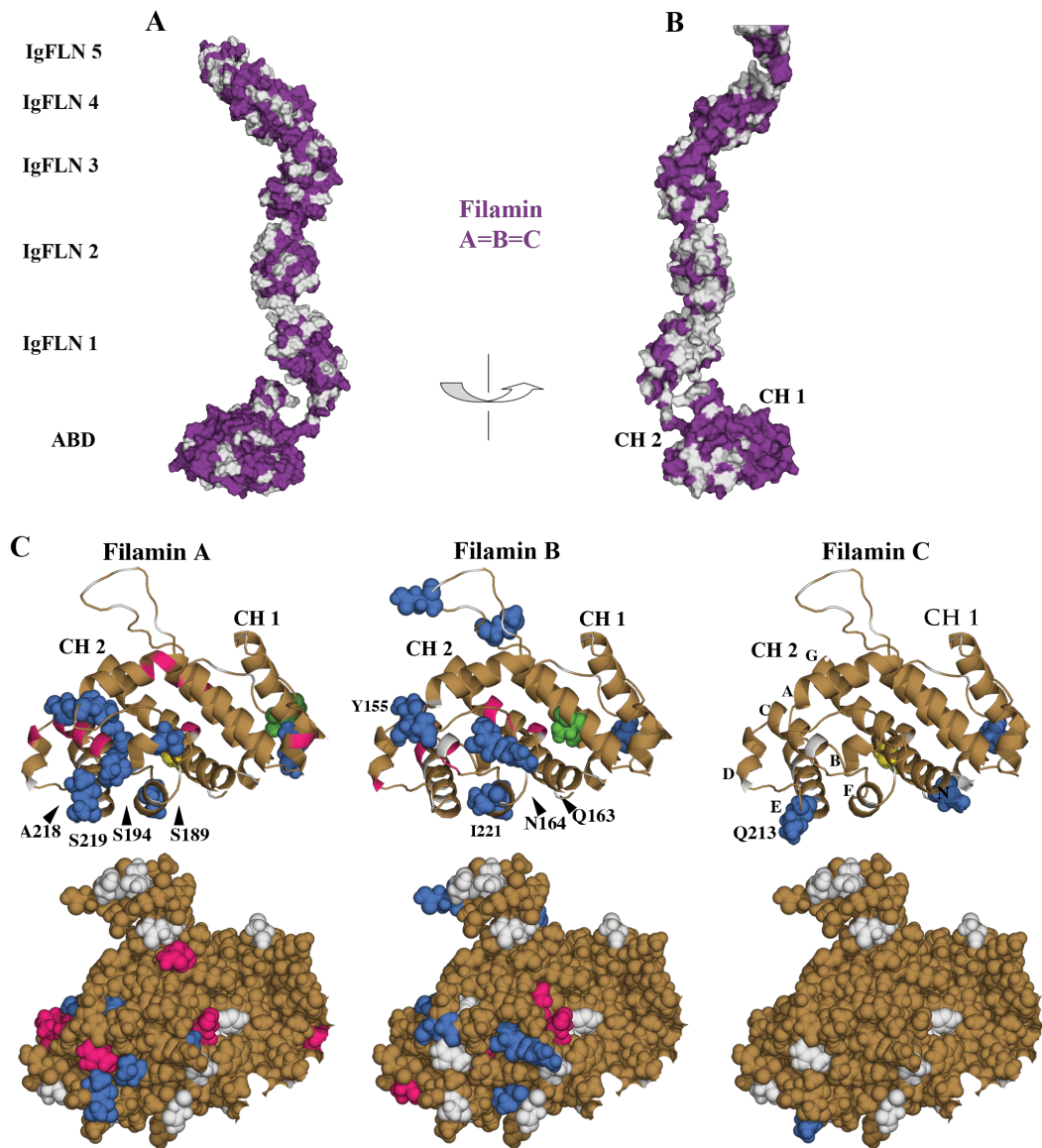
KEY: Ancestral, Distinctive (Teleostei, Amphibian, Mammalian), Disease

Figure 2.7. Spatial and temporal distribution of divergence in the actin-binding domain.
Coloring is the same as in Figure 2.4. In addition (top row) violet is ancestral and similar among isoforms; bottom row (pink) shows mutations leading to disease. (**A**) Surface representation of the actin-binding domain (ABD) and filamin Ig domains 1-5. Surface is highly ancestral. (**B**) 180° rotation of A about the vertical axis. Domains 1 and 2 have an asymmetric conservation pattern. (**C**) The ABD subdomains, calponin homology (CH) domains 1 and 2. CH domains contain seven helices labeled A-G on filamin C. Top is cartoon, bottom is space-filled model. Most of the distinctive residues in the ABD are localized to one region of CH 2, composed of helices B, C, E and F. Class-distinctive residues also co-localize with disease-causing mutations. Filamin A class-distinctive substitutions are mostly serine residues, while filamin B favors large bulky residues. Filamin C has fewer class-distinctive residues and is more ancestral than filamin A and B.

45
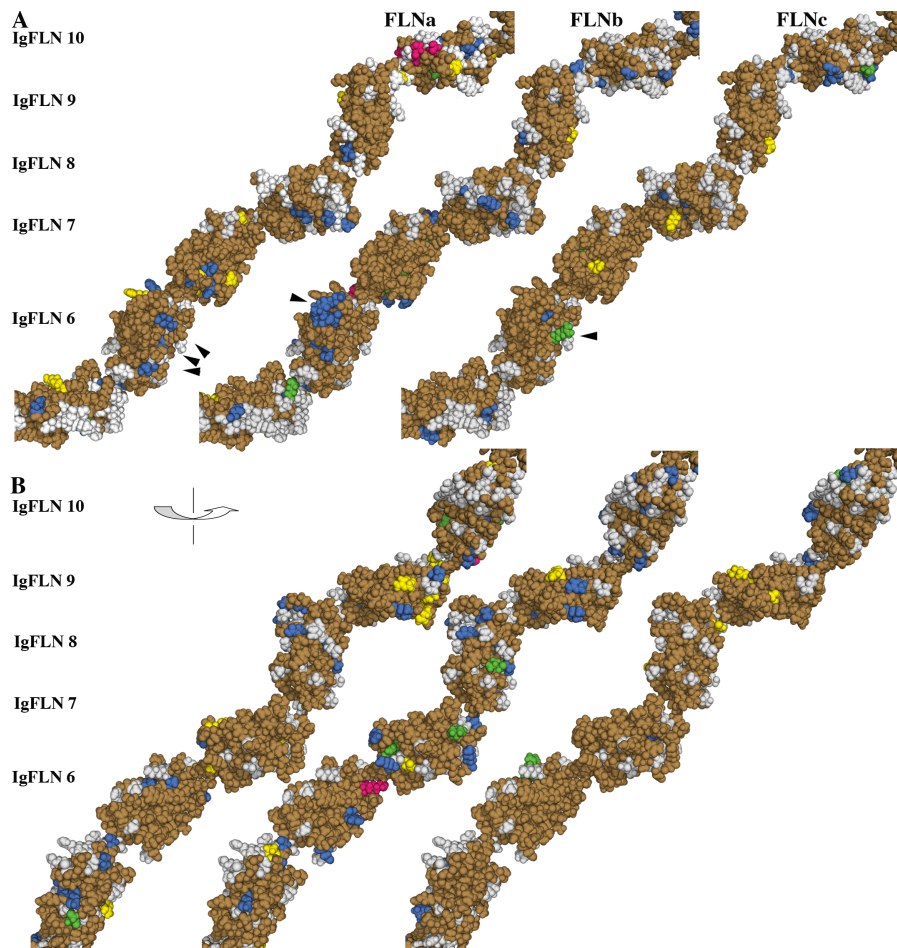
KEY: Ancestral, **Distinctive** (Teleostei, Amphibian, Mammalian), **Disease**

Figure 2.8. Distinctive divergence of Ig-like filamin domains 5-10. Coloring same as Figure 2.4. (**A**) Space-filling model of Ig-like domains 5-10 of filamins A, B, and C. Some domains are highly conserved (e.g., IgFLN 7), and some are highly divergent in each isoform (e.g., IgFLN 6).   Arrows in IgFLN 6 point to clustering of distinctive residues, including to an entire surface on IgFLNb6. Filamin C is more ancestral but nearly as variable. Unitary mammalian distinctive residues appear on most domains. (**B**) This side of the domains appears to have more distinctive residues than the other side, especially in filamin B. Disease mutations are surface-exposed but closer to the ends of the domain.
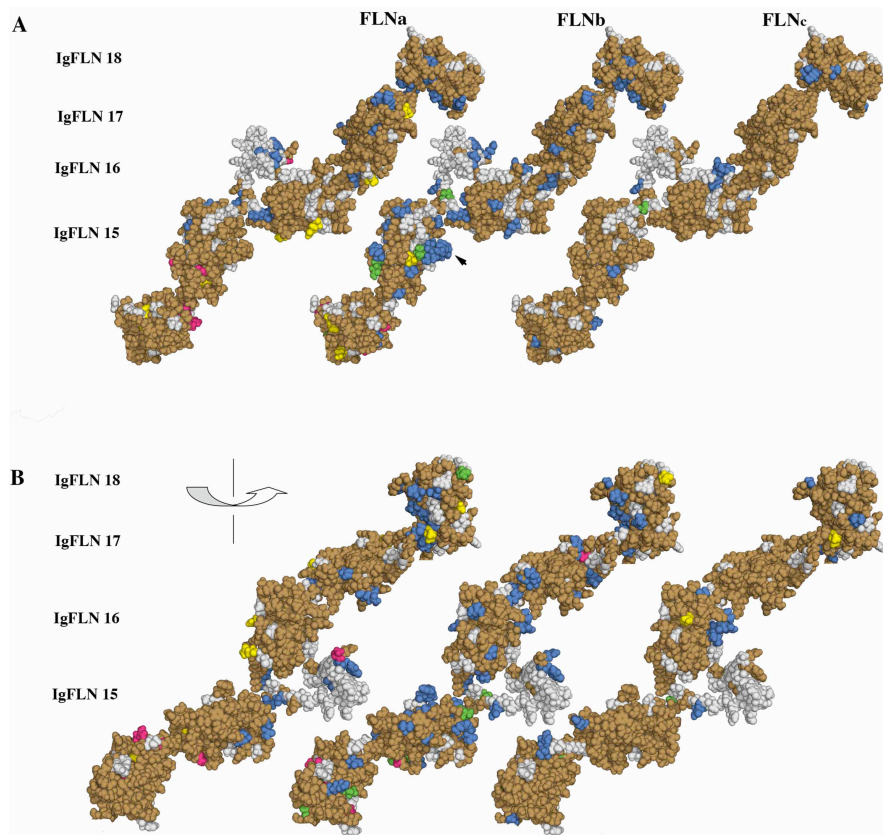
KEY: **Ancestral**, **Distinctive (Teleostei, Amphibian, Mammalian), Disease**

Figure 2.9. Distinctive divergence of Ig-like filamin domains 14-18. Coloring same as Figure 2.4. (**A**)Space-filling model of Ig-like domains 14-18 of filamin A, B, and C. The hinge between 15 and 16 is highly variable. Filamin C is highly conserved in each domain and has isolated Mammalian period distinctive residues. Filamin B is most divergent, especially domain 15 (arrow). The loop between β-strand A and B on IgFLN 15 contains a distinctive deletion in filamin B of 8 amino acids. Filamin A and C are similarly conserved. Filamin A and B have clusters of distinctive residues in domain 17. Domain 18 appears to diverge equally in the same region of each isoform. This region includes the unique feature of this domain, lacking an intact β strand A. (**B**)180° rotation around the vertical axis of A. Domains are similar as described in A.

Other domains

Results from the quantitative analysis of filamin class-distinctive residues, implies that filamin isoforms diversified in central domains. Here we examine class-distinctive residues mapped onto Ig-like domains 5-10 (Figure 2.8) and 14-18 (Figure 2.9).

Most class-distinctive residues in IgFLN 5-10 (Figure 2.8) are surface exposed and in loops between β strands of each domain. The sites surrounding class-distinctive residues on

most domains are also highly ancestral. Filamin C diverged during the Mammalian period, whereas filamin A and B started diverging in the Teleostei period and continued into the Amphibian period. IgFLNb6 is noted for large numbers of class-distinctive residues that cluster to various regions of the domain (arrows in Figure 2.8 A point to class-distinctive residue clusters: filamin A. surrounding β strand G (3 arrows); filamin B, loop CD (one arrow); filamin C, on β strand G (one arrow)). The largest cluster maps to a surface exposed loop between β strands C and D. This multi-residue substitution encompassed a negatively charged loop (^{784}DARVLSEDEEDV^{795}) in filamin B replacing an ancestral hydrophobic loop (APGVVGPAEADI) of filamin A and C. A motif scan (http://www.elm.eu.org/) suggests that filamin B potentially contains either a 14-3-3 or a casein kinase II (CK2) phosphorylation site and filamin A and C potentially contains an SH3 binding motif. Another domain with an interesting pattern of class-distinctive residues is IgFLN 9. The age of the class-distinctive residues in IgFLN 9 illustrates that this domain diverged both early in vertebrate and later during mammal evolution. One particular region (Figure 2.8 B IgFLNa9 and IgFLNb9, near center of the domain) contained two class-distinctive sites that became fixed in filamin A during the Teleostei (Q^{1115}L, along β strand D) and Mammalian (P^{1087}R, along β strand B) period and in filamin B during the Teleostei period (P^{1087}E, Q^{1115}S). In each case, although both isoforms utilized the same site, divergence is to a different residue type. Close to these residues is a cluster of class-distinctive residues on filamin A that striped across β sheet ABED. These residues are mostly from the Mammalian period and identified a rare cluster of mammalian divergence in filamin A. In a previous section we noted that IgFLNc9 is absent of Teleostei period class-distinctive residues. IgFLNc9 however, contained several Mammalian period class-distinctive residues, thus all three isoforms

48

diverged in the Mammalian period. It is also worth noting that domain 10 has a cluster residues associated with disease in close proximity to class-distinctive residues just described (Figure 2.8 A). Thus, in general domains in the range 5-10 are highly conserved, however we find many class-distinctive sites in which filamin A and B co-localize and filamin C is ancestral (e.g. IgFLNa8 and IgFLNb8 and IgFLNa9 and IgFLNb9). This pattern suggests that filamin A and B are diverging from ancestral function that filamin C possesses.

Ig-like domains in the range 14-18 are noted for their large quantities of class-distinctive residues relative to other domains. Class-distinctive divergence in this region occurs mostly in loop regions towards the terminal ends of each domain (Figure 2.9). Divergent residues are also in highly ancestral environments. There is a broader spatial distribution of class-distinctive residues in domains in the range 14-18 than domains in the range 5-10. An extreme example of this is class-distinctive residues in domains IgFLNb15, b16 and b17 compared to IgFLNb7 and b9. In domain IgFLNb15 class-distinctive residues clustered in 2-3 large groups, each surrounded by ancestral residues. One of these clusters localizes to eight residues in the turn between β strand A and B that is encoded by a single exon (171). This region has been previously categorized as an alternative splice variation in filamin A (172) and a deletion in FLNb (171). These previous classifications are consistent with our findings that filamin B distinctively diverged from ancestral filamin A and C by a deletion in this region. A motif scan suggests that this region may contain a GSK-3 phosphorylation site. As noted earlier, IgFLNa18 has a large number of class-distinctive residues. There is a similar number of class-distinctive residues in both filamin A and B, and almost every class-distinctive site co-localizes between isoforms. This highly class-distinctive region overlapped with β strands A, B and G. When we derived a homology

model for a filamin monomer, we modeled a turn in the linear extension of domains due to

our observation that β strand A in this region lacked residues bound to β strand B. Since

isoform divergence localize with this unique feature, each isoform may have a distinct intra-

domain architecture between domains 17, 18, and 19. The other domain we modeled lacking

β strand A (21) is highly ancestral in the analogous region (not shown) suggesting that the

architecture of 19, 20, and 21 is preserved from the common ancestor in each isoform.
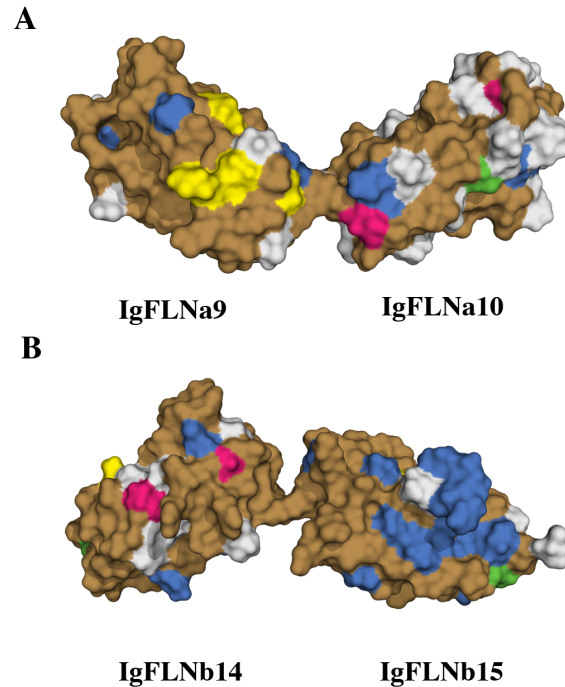
KEY:Ancestral, Distinctive (Teleostei, Amphibian, Mammalian), Disease

Figure 2.10. Distinctive residues associated with mutations causing disease. Coloring same as Figure 2.4. (**A**) IgFLNa9 and IgFLNa10 are surface models showing localization of distinctive and ancestral residues and residues with disease-causing mutations. (**B**) IgFLNb14 also shows similar localization.

Finally, we note that some class-distinctive residues co-localize with the sites of mutations causing disease (Figures 2.8, 2.9, and 2.10 A and B). While direct co-localization on a particular isoform is observed (e.g. IgFLNa10 and IgFLNb15 Figure 2.10 A and B), we also observed inter-domain localization (Figure 2.8 B, intersection between IgFLNb7 and IgFLNb8) and localization to homologous sites between domains (e.g. IgFLNa10 and IgFLNb10). Sites where mutations exist are predominantly surface exposed on loops between β strands, or near linkers between domains. Furthermore, almost all these mutations are on residues labeled ancestral. These data suggest that there is a correlation with class-distinctive changes and etiology of disease. One possibility is that disease mutations interfere with protein-protein interactions, another possibility is that disease mutations interfere with structural aspects of domains.

2.4 Discussion

Using phylogenetic analysis, an evolutionary trace, and an all-atom structural model of filamin A, we analyze spatial and temporal patterns of filamin family divergence to give us insights into structural elements involved in isoform class-distinctive and common function. The three components to this analysis consist of development of a structural model of filamin A; a phylogenetic analysis of invertebrate, invertebrate chordate, and vertebrate filamin paralogs; and an evolutionary trace analysis, which we use to observe the divergences of individual filamin residues in a spatial and temporal context.

The predicted structural model of filamin A is consistent with recent electron

micrographs and experimental data (62) that suggests that filamin A has two distinct architectural regions, a region consisting of domains 1-15 having a linear extension of domains, followed by a region (domains 15-24) having a compact conformation. The linear region is similar to the linear conformation of filamin domains that are proposed to exist based upon the high-resolution structure of ddfilamin domains 4-6 (155). The compact region results from two domains, 18 and 20 lacking β strand A. Interestingly, early electron micrographs of filamin A (18, 161) have an overall dimension and appearance quite similar to our model.

Our phylogenetic study of vertebrate filamin suggests that the three isogenes of filamin (A, B, and C) originated from a single ancestral gene that existed prior to the advent of vertebrates and that it is also the progenitor to extant urochordate (tunicates) filamin. The tunicate *Ciona intestinalis* has been extensively studied as having a simplified body plan that mimics the ancestor to vertebrates (162). Tunicates have organs and systems in which filamin isoforms have been shown to play a developmentally critical role. These include a well-developed muscle of their one chambered heart (163); vasculature (163); and immune system, that represents the prototype for the innate vertebrate immune system (164). Filamin binding partner and associated pathway orthologs have also been identified in ascidians, including specialized von Willebrand factor type A -like (165) and β1 integrin family –like members (1). The interaction between GPIbα (von Willebrand receptor) and filamin A is critical to platelet activation and adhesion as part of wound healing in vertebrates (2, 166). A recent report suggests that analogs of the von Willebrand factor and many other proteins involved in the vertebrate intrinsic wound healing pathway, do not appear to have a role in ascidian wound healing (167), but play a role in ascidian clot formation during allo-rejection

52

response (167). Thus, it is possible that ascidian filamin have similar protein-protein interactions, but may function in pathways that do not necessarily lead to the same overall biological function.

Evidence for the significance of our results includes observations of class-distinctive residues that co-localize with disease mutations, localize to critical sites within ligand binding pockets, or have the potential to be involved in F-actin binding regulation. Other evidence for the significance of our results comes from the placement of class-distinctive sites within regions of the ABD that are functional in homologous ABDs. For example, we find that a region in ABD CH 2 of α-actinin, reported to be associated with phospholipid binding (173) contains a high concentration of class-distinctive sites in filamin. This observation is consistent with the report that filamin function is also regulated by phopholipids *in vitro* (174) and we hypothesize some of our sites contribute to sensitivity to lipid regulation.

Our evolutionary trace analysis reveals ancestral conservation and isoform class-distinctive divergence down to the spatial granularity of an individual residue and temporal granularity to three periods – Teleostei, Amphibian, and Mammalian. We observe a non-uniform distribution of class-distinctive residues that varies over periods, and across domains and isoforms. Most protein interactions have been isolated to domains 15 and above. However, we do not see a clear delineation in divergence pattern between this and the preceding region, suggesting potential limitations in reported interactions resulting from the method used to initially identify interactions with filamin, the yeast two-hybrid method. Our methods, however, are unbiased in this way, treating each domain independently and of equal significance. We also find that conserved residues are not uniformly distributed (Figure 2,

Appendix A), implying potential conservation of interfaces in domains also outside the most extensively studied domains 15-24. Additionally, the temporal distribution of class-distinctive sites is not uniformly distributed. For example, we find it intriguing that some domains have Mammalian class-distinctive residues that clustered both in time and space (e.g. domain 9 with many Mammalian class-distinctive residues). Since filamin is involved in many biological processes critical for mammalian development (heart involvement for instance) we feel that these few clusters can hold the keys to important changes in mammalian evolution, and thus should be the object of future investigations.

We also hypothesize that the class-distinctive sites we identify may contribute to the evolution of isoform class-distinctive behavior in several ways, including variable specificity in binding; differential regulation of binding of adhesion proteins, effector proteins, and F-actin; and potentially class-distinctive modulation of inter-domain structure. For example, over 35% of class-distinctive residue changes in filamin A and C involve changes to or from S/T residues while only 11% of class-distinctive changes in B involve S/T residues. Thus, filamin isoforms are likely to have different roles in signaling as phosphoproteins. As an example of identifying isoform variable specificity in binding, we identify a class-distinctive residue on each of filamin B and filamin C, that subsequently was identified as critical for distinct binding of filamin A to FILGAP (175).

Since filamin plays a critical role as a structural protein (18) we also hypothesize that some of the class-distinctive residues we observe may mediate the structural characteristics of isoforms filamin, which may translate into distinct structural roles of isoforms in cells. Recently, Kesner et al. showed that filamin Ig domains have a heterogeneous set of stable conformations when forced to unfold under biologically significant levels of force (176). The simplest explanation for the heterogeneity of unfolding is primary sequence differences

between Ig domains. Since isoforms also differ in primary sequence, their Ig-like domains could have different forced unfolding properties. These properties include the stress required to change conformation and the stable intermediate conformation induced by force. It is possible that class-distinctive residues mediate response to stress. In support of this hypothesis, we find many filamin A class-distinctive residues in domain 18 that localize to the region we associate with having an aberrant first strand, suggesting that these distinctive residues play a role in the six-stranded conformation. In addition, filamin A has been reported to play a critical role in both exogenous and intracellular stress sensing (57, 58, 177). Therefore, isoform-distinct stress sensing could potentially play a role in how different cells respond to stress.

Chapter 3
N-terminal strands of filamin Ig domains act as a conformational switch under biological forces

This chapter was published as a part of a manuscript to *Protein: structure, function, and bioinformatics structure, Function, and Bioinformatics*; and is formatted in the journal style.

3.1 Introduction

Filamin family members (A, B, and C) are composed of an actin-binding domain and subsequent 24 immunoglobulin(Ig)-like domains. Filamins have been shown to play an important role in maintaining the rheological properties of the actin cytoskeleton (17). Filamin molecules can cross-link F-actin fibrils into actin-networks. As mechanical proteins, filamins localize to the cortex, stress fibers, and muscle Z-line. Besides their critical roles as mechanical proteins, filamins are also involved in cellular response to stress. Filamin A has been postulated to respond to cues from the stress environment of cells, as reported from its involvement in cell-fate determination (56), activation of platelets (44, 45, 178), and mechanoprotection (58, 61). How could filamin A transduce the stress state of a cell? Some studies report that filamin A turnover and ability to bind F-actin can be regulated through its interactions with calmodulin and calpain, proteins that are activated by changes in cellular calcium level (15, 56, 179). Calcium levels in turn fluctuate with cellular stress levels (99). Other studies suggest that filamin A's role as a scaffold protein, binding over 70 proteins (69), is modulated by its quaternary structure, which in turn could be regulated by either physical or thermal stress (93, 180). Based on these studies, we hypothesize that stress-induced conformational changes of filamin A play a direct role in signaling either by

disrupting existing interactions or by introducing new interactions. Therefore, it is important to investigate possible stress-induced conformational changes that could play a functional role.

To study force-induced unfolding of filamin A using DMD we need an atomistic structure of filamin A. However, structures of only five of the 24 Ig-like domains of filamin A have been experimentally determined (2, 5, 93, 181). Hence, in this study we use the all-atom structure of filamin A derived from comparative modeling in an earlier study (Kesner et al. *submitted*). The average sequence similarity of all the Ig-like domains is > 40%. Therefore, we expect the homology-derived structures to be close to their native structures (182). In that study we observed that there are two classes of Ig-like domains in filamin: six-stranded (lacking the N-terminal strand) and canonical seven-stranded. Domains 16, 18, and 20 are six stranded. In a recent crystal structure of domains 19-21 it was observed that the putative first strand of domain 20 interacts with domain 21 instead of forming a β sheet with the second strand. This interaction has been proposed to inhibit domain 21's interaction with other binding partners (93). Based on this quaternary structure it is possible that β strand A of an Ig-like domain could form complementary strands with adjacent Ig-like domains. Hence, we hypothesize that unraveling of the N-terminal strand of Ig-like domains could initiate inter-domain interactions through that strand. We study the six-stranded Ig-like domains from two perspectives. First, we compare domains 16 and 18 to that of domain 20 (whose structure has been determined) (93) to study if the six-stranded domains have similar physical characteristics. Second, we compare these domains to the seven-stranded domains to understand the role of the missing strand.

For each of the 24 Ig-like domains that make up filamin, we perform constant-force simulations by applying a set of physiologically relevant ranges of forces (0-315 pN) (16). Based on these systematic DMD simulations, we estimate the critical force of unfolding of each individual domain and characterized the conformational changes under mechanical stress. We find that the critical forces are heterogeneous among the 24 domains due to the sequence differences. We also find that there is no significant difference between the critical unfolding forces of domains with and without N-terminal strands. We find that under small forces up to ~35 pN, which correspond to the physiological forces (145), most Ig-like domains remain in the native states. As the force increases up to the intermediate force levels of ~70 pN, large conformational changes takes place. Interestingly, Ig-like domains feature a common initial conformational change, where the first β strand unfurls. Domains that lack their N-terminal β strands appear similar to one another, maintaining their native-like conformation under low forces. At intermediate forces, they unfold to a heterogeneous population of intermediate states similar to seven-stranded domains.

Additionally, the effect of temperature versus mechanical stress on unfolding pathways has been the subject of much debate (146). The general hypothesis is that thermally induced and force-induced unfolding pathways are independent from one another since the thermal fluctuations are exerted on the protein globally, while the effect of mechanical force is localized and non-homogeneous. To examine whether the thermally and force-induced unfolding of filamin Ig-like domains are different from each other, we also perform thermal unfolding of three Ig-like domains, 14, 21, and 24. Interestingly, we find that thermally induced unfolding shares the same initial unfolding intermediate state as forced unfolding, where the first strand unfolds. This observation suggests that filamin evolved to feature the

intermediate state of unfolded N-terminal strand(s). Given the observation that the putative

first strand of domain 20 interacts with domain 21 instead of forming a β sheet with the

second strand, we propose that an N-terminal strand may act as a conformational switch that

unfolds under stressed physiological conditions leading to exposure of cryptic binding sites,

removing of native binding sites, and modulating the quaternary structure of domains.

3.2 Materials and methods

All-atom DMD

A detailed description of the DMD algorithm can be found elsewhere (149, 183).

Briefly, interatomic interactions in DMD are governed by square-well potential functions.

Neighboring interactions along the sequence (such as bonds, bond angles, and dihedrals) are

modeled by infinitely deep square-well potentials to model the geometry of residues and

peptide bonds. During a simulation, an atom's velocity remains constant until a potential step

is encountered, during which it changes instantaneously according to the laws of

conservation of energy, momentum and angular momentum. Simulations proceed as a series

of such collisions, with a rapid sorting algorithm employed at each step to determine the next

collision.

The all-atom DMD method employs a united atom protein model, where heavy atoms

and polar hydrogen atoms are explicitly modeled (149). We include van der Waals, solvation,

and environment-dependent hydrogen bond interactions. For solvation, we adopt the

Lazaridis-Karplus solvation model (184, 185) and use the fully-solvated conformation as the

reference state. Due to the strong screening effect of solvent, distant charges have weak polar

interactions. For salt-bridges, we expect the hydrogen bonds to partially account for their polar interactions.

Protein models

As described in the introduction filamin A is composed of an ABD followed by 24 Ig-like domains. Ig-like domains are seven-stranded β sandwich folds (Figure 3.1 A). Loops with conserved residues GPG and SPF bisect the first and last strands. We refer the resulting subdomains as A, A', G', G. Subdomain A forms a β sheet with B (A-B region) and in some domains A' and G (A-G' region) form a β sheet. We use the homology models of human filamin A Ig-like domains (Figure 3.1 A) that were derived from an earlier work (Kesner et al., *submitted*). Briefly, closest matching homology templates were derived from a structural alignment of homologous filamin domains in the protein data bank (PDB) (186). *Insight II* (Accelrys, CA) was then used to perform homology modeling of each filamin A Ig-like domain. Domain 16, 18 and 20; have only six complete strands since the best-match template lacks all (16), or sub-strand A (18, 20) of β strand A.

Each model is validated using the sequence-structure compatibility score from the *Verify-3d* module of *InsightII* (Accelrys Inc.). This analysis measures the compatibility of modeled residues with their structural environment (154). Quality scores of greater than 0.1 indicates valid structural models (154) and typical biochemical structures have a score of 1. The mean *Verify-3d* score obtained for Ig-like domain models was 0.89 with standard deviation 0.30. Overall, Ig-like domains have high quality scores indicative of their high accuracy.

Following homology modeling, each structure is relaxed using 1000 steps of steepest

descent energy minimization as in AMBER(187) 9.0. In addition, each model is minimized

using the Medusa modeling suite (150, 151, 188).
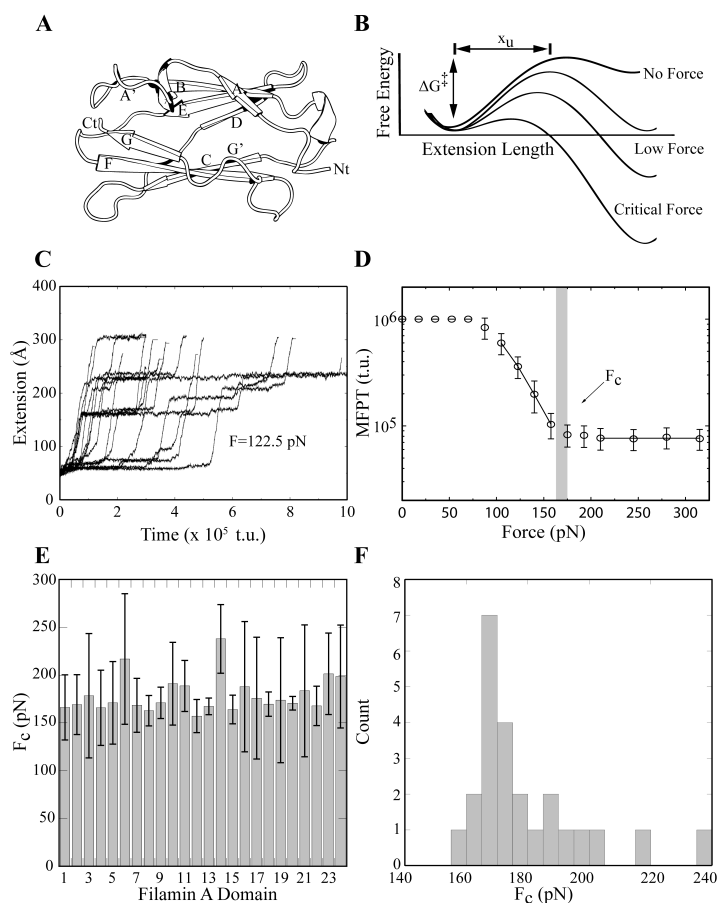
Figure 3.1. Critical force-induced unfolding parameters of filamin Ig-like domains. (**A**) Seven-stranded Ig-like domain with subdomain termini and β strands A-G labeled. (**B**) Graph depicting the free energy landscape along the extension length as the reaction coordinate at different levels of constant-force pulling. F_c is the critical unfolding force and x_u is the distance to the center of the unfolding energy barrier. (**C**) Unfolding behavior of a domain pulled apart with a constant force (122.5 pN) is described by plotting the extension length as a function of time. Replicates with differing initial velocities vary in their unfolding behavior. (**D**) A linear natural log plot of the pulling force versus mean first passage time. F_c is the critical unfolding force. A linear fit of forces is shown for the barrier (left of F_c) and barrierless (right of F_c) regimes. (**E**) Critical unfolding forces observed for each domain. The error bars are determined as described in Methods. (**F**) Histogram of mean F_c.

Constant-force unfolding

For each domain, we start the constant-force simulations from the homology-derived native state. The dimension of a cubic simulation box is 500 Å and the periodic boundary condition is used. We use an Anderson thermostat(189) to maintain constant temperature at 270 Kelvin. Constant-force pulling is achieved by applying a discretized step-function with a constant energy jump, dE, at the distance step of dR (1 Å) between the N- and C-termini. The pulling force equals to dE/dR. For example, dE of 0.5 kcal/mol is equivalent to approximately 35 pN. The applied forces range from zero to 315 pN. From zero to 210 pN, we sample the force values with an interval of 17.5 pN. From 245 to 315 pN, the interval is 35 pN. For each force, we perform 10-20 realizations of pulling simulations, each of which start with randomized velocities. We perform 10 simulations for pulling simulations at low forces (70 pN or less), since there is a low probability of unfolding during the simulation time. For higher forces, we perform 20 realizations.

Derivation of critical unfolding forces

Under external force F, the unfolding time under pulling forces can be approximated as $\tau = \mathrm{A}e^{(\Delta \mathrm{G}^{\ddagger} - Fx_u)*\beta}$. Here, A is the Arrhenius prefactor, $\Delta \mathrm{G}^{\ddagger}$ is the zero force activation energy barrier, x_u is the distance to the center of the activation energy barrier, and β is the inverse of $k_b T$, where k_b is the Boltzmann constant and T is the temperature. We use the mean first passage time (MFPT) of unfolding as the approximation of the unfolding time. We use 200 Å as the cutoff distance to assess whether a protein is unfolded. This cutoff was determined base upon observations that after an extension of 200 Å most domains lack metastable states, although the minimum full extent for an unfolded domain is 240 Å (Figure

3.2). If the protein has not unfolded within the maximum simulation time (one million time units), we use this time as the first passage time.

We plot the ln(MFPT) versus force and obtain linear fits in the barrier and barrierless unfolding regions. The intersection of the two linear fits determines the critical force of unfolding (F_c). We use the error bar of the linear regression to estimate the error bar of the estimated critical forces with a simple Monte Carlo method. In each Monte Carlo step, linear lines constructed based on the Gaussian distribution of error bars (derived form the linear regression) is used to determine a new intersection. The mean and standard deviation of the critical unfolding force is derived from the set of intersections.

Forced unfolding pathways analysis

For each domain and each force, we combine trajectories of all realizations and compute the histogram of extension lengths. Gaussian curve fitting of the histogram is used to identify metastable states. Only the states that are present for more than 1% of the total simulation time are considered stable. To determine the representative structure of each metastable state, we first extract all conformations having extension length within the mean ± 2 SD of the corresponding Gaussian fit. Then, we compute a contact frequency map for each extracted ensemble. Finally, we select the representative structure as the one that has the least contact distance from the average contact frequency of the ensemble.

We characterize the unfolding pathway using a directed graph. The nodes of the graph are formed by the states identified from the extension length histogram and the edges are the transitions between the different states. Occupancy of each node corresponds to the number of conformations associated with that state normalized by the total number of conformations.

Edges have an associated transition probability determined by counting the number of transitions between a node and its neighbor and then dividing by total number of transitions to all its neighbors in the same direction. The graph representation of the unfolding pathway is visualized using Cytoscape (190).

Thermal unfolding

Equilibrium simulations are performed at a wide range of temperatures: 300, 330, 360, 390, and 420 Kelvin. We use the percentage of native contacts, Q, to approximate the reaction coordinate. We define a contact when the distance between two $C\beta$ atoms ($C\alpha$ for Glycine) is less than 8.5 Å. We use a similar approach to identify the representative structure of each state, where the Q-value rather than the extension distance is used as the approximation of the reaction coordinate. To determine the localization of loss of contacts during thermal unfolding simulations we calculate a subdomain average Q-value. Subdomains are delineated based upon a strand-by-strand analysis of each Ig-like domain homology model. Average Q-values for an ensemble are obtained by dividing the marginal sum of native contacts from the residue-by-residue contact frequency map by the marginal native contacts from the reference structure contact map.

3.3 Results

We model constant-force unfolding by applying a pulling force between the $C\alpha$ atoms of the N- and C-terminal residues (Materials and methods). The force is modeled by assigning a corresponding potential between the corresponding atom pair. For each force value, we perform 10-20 independent simulations with randomized initial velocities. Each

simulation lasts a million DMD time units.

Computationally identified critical unfolding forces of filamin Ig-like domains agree with AFM experiments

Most single domain proteins, such as the Ig-like domains in filamin, feature a two-state folding/unfolding dynamics without external forces. At 270 K, a protein stays folded and its native state corresponds to the lowest energy state in the free energy landscape (Figure 3.1 B). Unfolding requires overcoming a free energy barrier (the folding/unfolding transition state) separating the folded and unfolded states, $\Delta G^{\ddagger}$. The unfolding time is proportional to the exponential of $\Delta G^{\ddagger}/k_B T$. As the force (F) applied to the N- and C-termini of a protein increases, the unfolded states with large end-to-end extensions are favored (Figure 3.1 C), and the corresponding free energy barrier decreases ($\Delta G^{\ddagger}$-Fx_u (Figure 3.1 B)). Here, x_u denotes the difference of end-to-end distances between the transition and native states. Assuming that x_u does not change for different external forces, the natural logarithm of the unfolding time is linearly dependent on the external force. As the force reaches the critical value (F_c), the barrier will disappear and activation driven unfolding dynamics is reduced to barrierless unfolding. As a result, the dependence of the unfolding time with respect to the external force changes near the F_c (Figure 3.1 D). We estimate the critical unfolding forces based on the change of the scaling of the unfolding time with respect to external forces (see Methods). For each applied force, we compute the mean first passage time (MFPT) over 20 trajectories to estimate the unfolding time. We use a cutoff distance of 200 Å to assess whether a protein is unfolded. We estimate the critical force as the crossover point in the plot of ln(MFPT) as a function of external forces (Figure 3.1 D). We compute the

critical unfolding forces for each domain (Figure 3.1 E). The mean critical unfolding force over all domains is 179±19 pN. We find that the distribution of F_c values features a peak around 155-180 pN with a long tail from 185-205 pN (Figure 3.1 F). Therefore, most of the Ig-like domains of filamin A can withstand the stretching forces up to 150 pN with a few domains being able to resist higher forces. For example, domains 6 (217±68 pN) and 14 (238±36 pN) have F_c values greater than 205 pN. Our calculated critical forces are consistent with the experimental measurements (50-220 pN as identified by AFM pulling study of Filamin A) (16).

Sequence alignment during the homology modeling of domains 16, 18, and 20 suggested that these domains lack the N-terminal β strand of the canonical Ig fold. Such a deviation from the canonical Ig-like domain suggests a unique biological function for these domains. Interestingly, the absence of these structural features did not significantly change the mechanical stability of these domains with respect to other domains. The two domains, 18 and 20, have F_c of 170±18 pN, and domain 16, has an F_c of 188±68 pN. These data suggest that the mechanical stability of Ig domains is independent of the N-terminal β strand and likely to be caused by sequence difference in the core of the protein.

Recent *in vitro* pulling experiments on filamin-actin networks suggest that filamin A Ig-like domains are subjected to equilibrium forces up to 80 pN (145). Hence, we analyze the conformational changes of each of the 24 Ig-like domains under 35 and 70 pN of constant pulling force next (Forced unfolding pathways analysis materials and methods).

Under the 35 pN pulling force, most filamin Ig-like domains remain in a native-like state with a few domains featuring intermediate conformations that could potentially

play a role in stress signaling

       We analyze the population of conformations of each Ig-like domain under the 35 pN

pulling force, which we expect to be physiologically relevant. We find that most of the

domains stay in their near-native conformations as shown for domain 21 (Figure 3.2 B). The

average end-to-end extension of these conformations is increased by ~10 Å, compared to the

extension length seen in zero force extension (Figure 3 A and B, Appendix A). This is the

result of pulling out leading N-terminal residues (Figure 3.3), which make little contacts in

the native state. The rest of the structure remains intact.

Figure 3.2. Equilibrium stable intermediate states of domains at 35 pN of constant-force pulling. Histogram of extension lengths for domains 13 (**A**) and 21 (**B**). A representative conformation for each state is shown. The solid line represents the distribution from simulations at 0 pN pulling and the dotted line represents the distribution from simulations at 35 pN pulling.
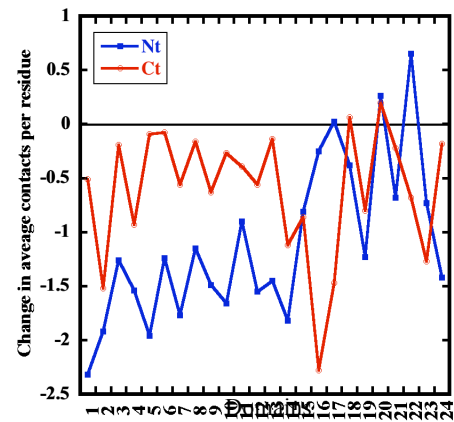
Figure 3.3. Loss of N-terminal contacts of domains pulled apart at 35 pN. For each domains we measured the average change in contacts relative to native, for N-t β strands A and C-t β strand G of each domain forced to unfold at 35 pN of force.

67

Interestingly, we find that at the 35 pN pulling force a few domains feature a partially unfolded intermediate state, in addition to a near-native conformation. This is exemplified by partially unfolded state of domain 13 (Figure 3.2 A), which has the interaction between strands A and B lost while maintaining A'-G region contacts (Figure 3.2 A; the 55 Å conformation). This conformation is analogous to changes hypothesized for TNI27 when it was pulled at much higher forces (124, 191). The loss of β-strand A contacts while other β strands are left intact is common to most of the intermediate states of these domains. We do observe one exception: domain 15 loses its entire β strand A (Figure 3.4). This major conformational shift occurs in a small number of ensembles and produces an extension length of ~86 Å (Figure 2 B, Appendix A). Thus, while several domains have a modest conformational shift pulled at the 35 pN force, at least one domain has a major conformational shift that could serve as a potential signal of relatively low levels of cellular stress.
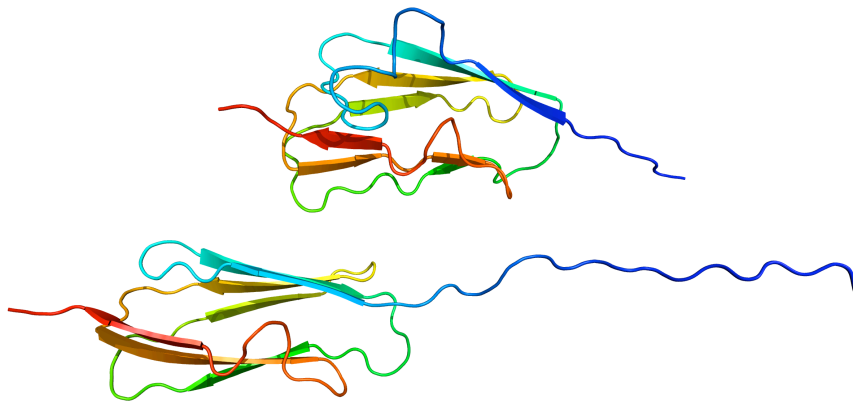
Figure 3.4. Domain 15 force-induced unfolding conformations at 35 pN.
We observe two stable ensembles when domain 15 is forced to unfold at 35 pN.
A highly populated ensemble with mean extension of 53 Å (top) (representative conformation), and a smaller population with mean extension of 75 Å (bottom).
In this population, β strand A is formed from loop AB.

At 70 pN, filamin domains have a diverse set of metastable states. The unfolding pathway of all domains is initiated by the loss of contacts in the N-terminal β strands

Under 70 pN pulling force, all Ig-like domains undergo large conformational changes (Figure 3.5 and Figure 3 C, Appendix A). The most populated states are those with end-to-end extension within 20 Å of native state, where β strands A and/or G has been pulled out slightly, leaving the remaining structure native-like (Figure 3.5). The stretched conformations with longer extensions (>20 Å) have one β strand sliding past another (Figure 3.5 A) or lose contacts between strand A' and G (63 Å extension in Figure 3.5 B; and 72 Å extension in Figure 3.5 C). Interestingly, TNI27 is reported to have similar initial unfolding pathways which features loss A-B contacts, loss of contacts in the A'-G region, and then quickly unfolding (124, 191). However, at 70 pN stretching force filamin Ig-like domains do not traverse directly to the unfolded state, instead they get trapped in several metastable states (Figure 3 C, Appendix A). These states include domains in which strand A (130 Å extension in Figure 3.5 B) or strands A and B have been unfolded (150 Å extension in Figure 3.5 A, and 149 Å extension in 3.5 B) from the rest of the domain resulting in conformations such as a single β sheet (150 Å in Figure 3.5 A) or a β sandwich (149 Å in Figure 3.5 B). These conformations are analogous to those observed in the metastable intermediate of ddFLN4 when unfolded using forces in the order of 60 pN(133, 134). Also, FN-III$_{10}$ has been observed to lose N-terminal β strands when forced to unfold using ~100 pN (131, 137). Domain 21 has a proclivity to also lose β strand G entirely (Figure 3.5 B 149Å), which could potentially be related to its role in the previously mentioned interdomain conformation, involving domain 20. Most other domains only lose β-strands at the N-terminus.
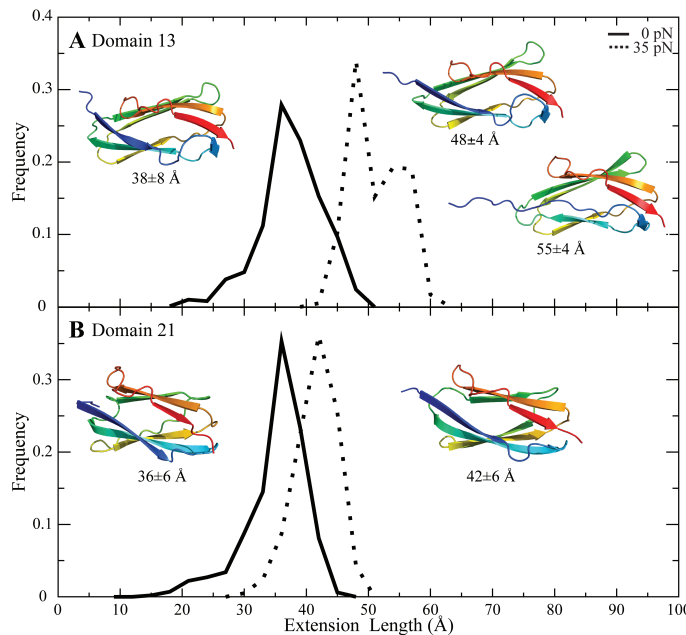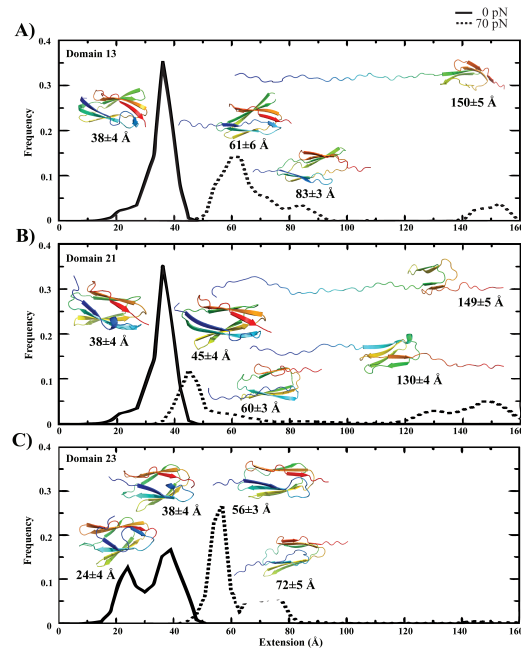
Figure 3.5. Equilibrium stable intermediate states of domains at 70 pN of constant-force pulling. Histogram of extension lengths for domains 13 (**A**), 21 (**B**), and 23 (**C**) A representative conformation for each state is shown. The solid line represents the distribution from simulations at 0 pN pulling and the dotted line represents the distribution from simulations at 70 pN pulling.

The most stable domains at 70 pN unfolding force are 3, 9, and 22, which have metastable states close to that of their native state (extension length < 86Å). The frequent interconversion between intermediate states of these domains suggests a low energy barrier between successive extension length states (domain 3). Domains 1, 6, 12, 13, and 14 also feature the near-native state being highly populated. However, these domains also feature an intermediate state missing the N-terminal β strands. Interestingly, we find other domains (e.g. 21) are fully unfolded using a force ~70 pN, suggesting a lesser stability of these domains.

70

Domains in the range 9 to 15 are considered secondary sites for binding actin, and having unusually stable domains in this range could be critical to the structural role of filamin in cells (62).

We also find that the mechanical stability of some Ig-like domains does not depend entirely on N-terminal β strands. We modeled domains 16 (N-terminal residues up to and including β strand A'), 18 and 20 (N-terminal residues up to and including β A) lacking N-terminal β strands (Kesner et al., *submitted*) as their templates lacked β strand integrity in this region. We find that at both 35 and 70 pN, these domains have similar unfolding pathways and metastable conformations as seven-stranded domains. Domain 20 maintains a highly populated native-like conformation featuring an A'-G region β sheet at 35 pN (Figure 3.6) and an unfolding pathway where the most populated state is native-like at 70 pN similar to seven-stranded domains. Domain 18 at 35 pN features two metastable states and does not unfold entirely. The most populated state is a β barrel with few of the native contacts lost (Figure 3.6). However, domain 18 also features a lowly populated state with a conformation that has an unfurled N-terminal β strand B at 35 pN, thus making it somewhat distinct from seven-stranded domains in which we observe the unfolding of β strand B only at 70 pN. At 70 pN domain 18 has a similar unfolding pathway as that of domain 24, featuring highly populated intermediate states that have extension lengths much greater than that of their native states (Figure 3 C, Appendix A). Domain 16 features the greatest differences in the unfolding pathway compared to seven-stranded domains: it has an unfolding pathway that is distinct from that of domains 18 and 20 and most seven-stranded Ig-like domains unfolding from the C-terminus at 35 pN (Figure 3.6). However, domain 16 is

very stable because at 70 pN, its most populated state is a metastable and not the fully unfolded state. Thus, all three domains lacking N-terminal β strands are highly stable at 35 and 70 pN of force, although domain 18 and 20 are more similar to each other than to domain 16, which unfolds from the C-terminus.

This systematic analysis of our *in silico* pulling studies of filamin Ig-like domains reveals metastable intermediate states as observed in AFM experiments. Our simulations are able to reproduce experimentally derived critical unfolding forces. The conformational changes of all the Ig-like domains feature a common intermediate state where the N-terminal strand is unfolded, which could play a role in the signaling pathway of cell responses to mechanical stress. Next, we study the thermally induced unfolding of the filamin domains to test whether the intermediate states observed in forced pulling are also observed in thermally induced unfolding pathways. We restrict our study to three Ig-like domains, 14, 21 and 24. These domains are selected due to distinct kinetic properties and biological significance: domain 14 has the highest critical force, domain 21 has well-characterized thermodynamics(180), and domain 24 is involved in dimerization (181).

Extension length 37 Å

Extension length 75 Å

Domain 16

Extension length 20 Å

Extension length 41 Å

Domain 18

Extension length 20 Å
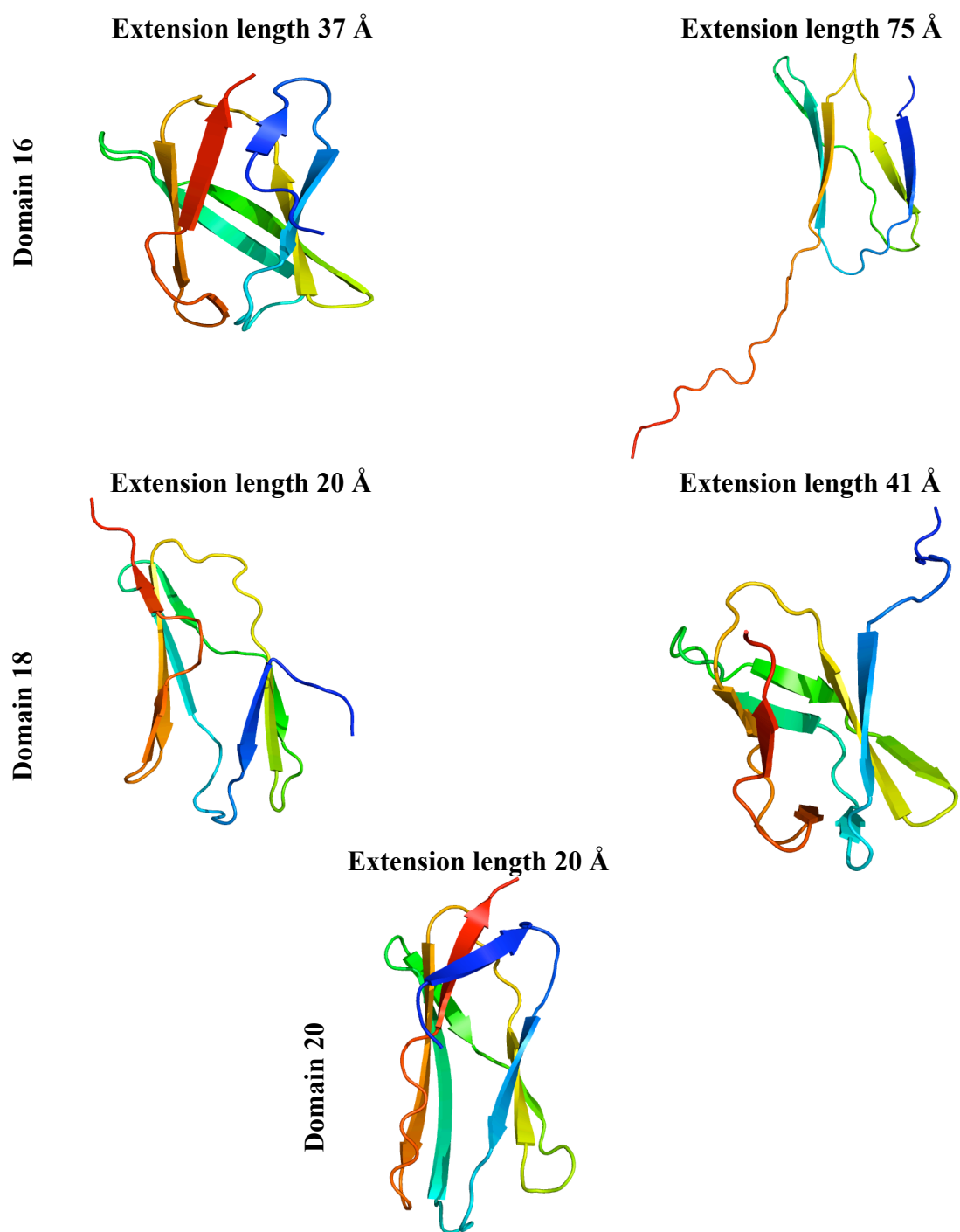
Domain 20

Figure 3.6. Domains without N-terminal β strands force-induced unfolding conformations at 35 pN. Domain 16 lacks β strand A and A'. Domains 18, and 20 lack β strand A. Shown are stable conformations of these domains pulled apart at 35 pN. Domain 16 has two stable conformations, a native-like fold and a conformation where β strand G is unfolded. Domain 18 has two stable conformations, a native-like fold and a second conformation where β strand A' unfolds from β strand B. Domain 20 has a highly stable native-like fold.

73

Thermal unfolding of filamin Ig-like domains features a common initial unfolding of the N-terminal β strand

We perform equilibrium DMD simulations of Ig-like domains 14, 21 and 24 over a wide range of temperatures (300-420 K). We use the weighted histogram analysis method (WHAM)(149) to compute the specific heat (Figure 3.7 A). The highest peak in the specific heat corresponds to the unfolding transition. The shoulders in the specific heat plot suggest the existence of an intermediate state. Our results indicate that domain 14 and 21 have similar unfolding temperature of ~340 K. Domain 24 is the most stable of these domains with unfolding transition temperature ~10 degrees higher than either 14 or 21 (Figure 3.7 A). The increased stability of domain 24 is possibly the result of its well-characterized role in dimerization of filamin monomers (181). Domain 14 has a prominent shoulder at temperature ~370 K that is absent in domain 21, suggesting the existence of a thermally stable intermediate state.

To further investigate the conformational states of these domains, we use the fraction of native contacts (Q) as the approximation of the reaction coordination to analyze the conformational state at different temperatures (Figure 3.7 B). We also compute the Q-value of each of segments in the native state (Figure 3.1 A, 3.7 C). At the low temperature of 300 K, all domains have near-native conformation with high Q-values (Figure 3.7 B). As the temperature increases to 330 K, we find domain 14 features two states, I1 and I2, while domains 21 and 24 have only one. We examine the Q-values of each secondary structure segment at 330 K (Figure 3.7 C). For domain 14 both states I1 and I2 have large losses in A-B region contacts, I2 however has more contacts lost in A', B and G than does I2.

Furthermore, the conformation in I1 is a closed β sandwich whereas the conformation of I2 is an open β sandwich (Supplementary Figure 5A). Thus it is likely that A'-G region contacts are involved in maintaining the closed conformation of I1. Domain 21 has lost significant amount of native contacts in both the N- and C-terminal β strands. Domain 24 loses contacts of mainly in β strand A. The existence of two states of domain 14 is consistent with the shoulder in the specific heat. At high temperature of 360 K, the most populated state of each domain is the unfolded conformation. Consistent with the high thermal stability, the β strands CDFG of domain 24 still retain partial native interaction at 360 K (Figure 3.7 C).
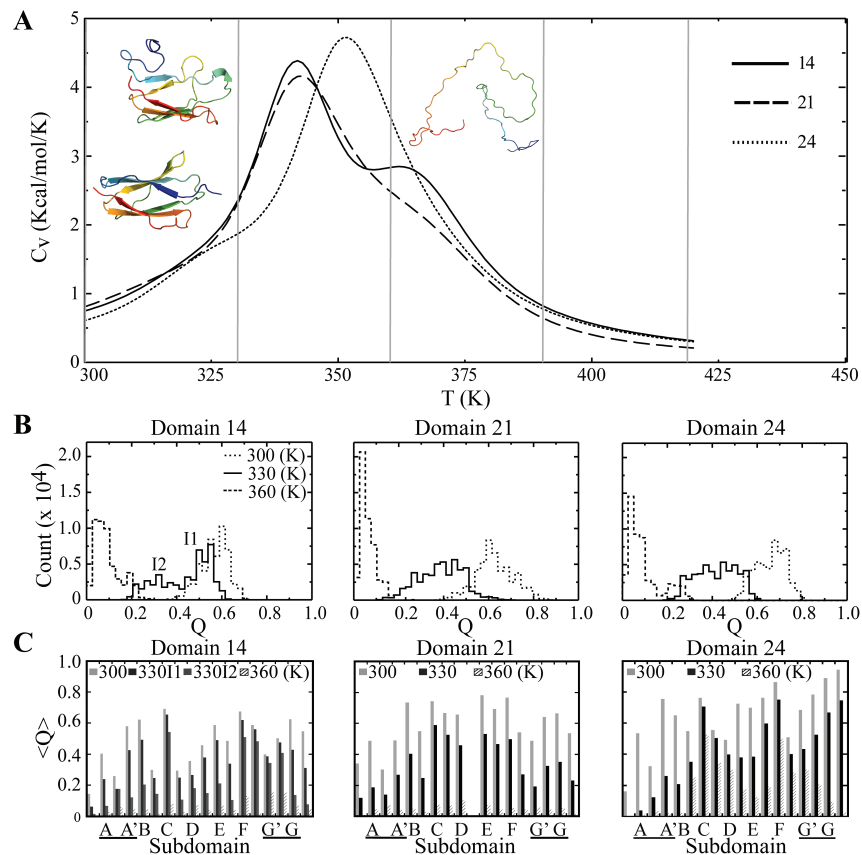
Figure 3.7. Thermal unfolding of filamin domains. (**A**) Plot of the specific heat as a function of temperature for domains 14, 21, and 24 calculated using the WHAM. Representative conformations of the metastable states of domain 21 at temperatures 300, 330 and 360 are also shown. (**B**) Histogram of Q-values for compiled replicated trajectories at each temperature (300, 330, 360 K) for domains 14, 21 and 24. Metastable states of domain 21 at 330 K are labeled I1 and I2. (**C**) Average Q-value for metastable state for each subdomain (Figure 3.1 A) at each temperature (300, 330, 360 K).

Jiang et al., using an NMR based approach determined the unfolding temperature of domain 21 to be 319 K. They also reported that different filamin Ig-like domains have different thermal stabilities. Domain 19 and 17 are fully folded at 333 K while domain 21 is unfolded (180). They have also reported a partially unfolded fraction of domain 21 under mild conditions (298 K). It is possible that the differences between domains observed were the result of β strand A's displacement that we observed in domain 21 in our simulations at 330 K since the residue they used as a folding marker is adjacent to β strand A. Our results are quite distinct from those of other reported simulations of thermal unfolding of Ig-like domains. For example, Paci et al., simulated the unfolding of TNI27 at 450 K and found unfolding pathways in which the last strands to unfold were the terminal β strands (192). Given the high temperatures used, it is possible that these observations were the result of non-equilibrium simulations.

3.4 Discussion

Mounting experimental evidence suggests that filamin A molecules act not only as structural and mechanical proteins that cross-links F-actin fibrils, but also as scaffold proteins that play a critical role in cellular response to external mechanical stress (44, 45, 56, 58, 61, 178). We hypothesize that conformational changes of filamin A due to mechanical stress are important for signaling, where stress-induced conformational changes can either disrupt existing interactions or introduce new interactions. We systematically studied the conformational dynamics of forced unfolding of filamin Ig-like domains using DMD. We developed a novel technique to derive the critical unfolding force from constant-force trajectories that is based upon finding the intersection between two unfolding regimes with respect to pulling forces. As a validation of our methodology, our computationally estimated forced unfolding dynamics of filamin A Ig-like domains agree with AFM experiments. For example, our simulation suggests the average critical unfolding force is 150-240 pN, which is consistent with the experimental measurements of 50-220 pN. Our analysis of the force-induced unfolding trajectories suggests the existence of multiple intermediate states corresponding to partially unfolded β sandwich of Ig-like domains. Additionally, an intermediate state observed in the simulations common to most filamin A Ig-like domains is characterized by the unfolding of the first N-terminal β strand. This intermediate state is consistent with the intermediate state observed in the force-induced unfolding of *D. discoideum* filamin domain 4, 27 Ig-like domain of titin, and fibronectin type three domain 10 (122-124, 131, 133, 135, 137, 191). Thus, our stretching simulations using all-atom DMD is able to recapitulate the conformational dynamics of Ig-like domains under force.

Our simulation results suggest that at low force (35 pN) most of the Ig-like domains remain in the native state with only a few domains featuring a weak population of a partially unfolded N-terminal strand. As the force increases (up to ~70 pN, which is close to the maximal forces observed *in vivo*) (193), we observe an increased population of intermediate states. Most of the domains still remain in their native or near-native conformations since these forces are still below the F_c. We find a large heterogeneity in both the population of the force-induced intermediate states and the transitions between them. However, we do find a common initial unfolding intermediate common to all Ig-like domains, where the N-terminal strand unfolds, similar to those observed in AFM experiments on TNI27 (123, 191).

Studies with TNI27 have suggested that contacts within A'-G region are important for maintaining the stability of its metastable state (121). We observe a similar phenomenon in domain 20's metastable states. However, the stability of other six-stranded domains appears to be independent of this region. Domain 18 loses A'-G region contacts during unfolding, and yet does not completely unfold. Also, domain 16 which lacks subdomain A', has a similar critical unfolding force as other domains. These three domains might not be representative of all domains since in their native-like state they are stable absent N-terminal β strands. On the other hand both thermally and forced unfolding pathways of most seven-stranded Ig-like domains feature metastable states that lack any A' contacts. Thus, it is likely the A'-G region plays a minor role in filamin Ig-like domain stability.

Our results from thermal unfolding of Ig-like domains suggest that there is no correlation between thermal and mechanical stability. For example, domain 14, which has the highest critical unfolding force, is thermally less stable than domain 24. However, there does appear to be a correlation between thermal unfolding and force-induced unfolding pathways.

In the case of both low force unfolding and low temperature thermal unfolding, unfolding is mostly initiated with the loss of contacts in N-terminal β strands. In the few cases where we observe initiation of forced unfolding at both the N- and C- terminus (domain 21) we also observe loss of contacts at both termini due to thermal denaturation. Thus, as evidenced by both thermal and force-induced unfolding pathways, Ig domains of filamin have a similar initial unfolding intermediate state. This similarity may be the result of evolutionary optimization of filamin due to its role as a structural and scaffold protein.

Filamins interact with a wide range of cell signaling proteins, and the conformational changes upon stress can either disrupt or promote these interactions. Additionally, the exposure of partially unfolded strands can promote modifications such as phosphorylation, which in turn can activate downstream signaling pathways. Therefore, we postulate that intermediates with their N-terminal strands unfolded could be involved in the cellular response to external mechanical stress. Further experimental studies are required to verify this hypothesis.

3.5 Conclusion

Stress levels in cells can be very large, especially in the cortex where the amount of stress depends on its compliance to the surrounding surface (194). In this study, we find that under forces between 150–240 pN filamin domains readily unfold. However, under lower forces, domains unfold in stages producing a heterogeneous population of metastable states lacking N-terminal β strands. Our simulations of thermal unfolding reveal similar metastable states. Furthermore, we observe that the domains that were modeled lacking all or part of the N-terminal β strand (16, 18 and 20) have similar forced unfolding characteristics as canonical

Ig-like domains. Thus, it appears that the stability of Ig-like domains is independent of N-terminal β strands. Domain 20 has been shown to have an N-terminal β strand playing an inhibitory role for protein-protein interactions involving filamin A (93). Since domain 18 has a similar forced unfolding pathway as domain 20, it is possible that β strand A of domain 18 has a similar biological function. Moreover, it is possible that unraveling of N-terminal β strands leads to interdomain interactions, that may produce the inhibitory behavior, as observed by Lad *et al.*(93). Other possible ramifications to having exposed β strands include exposure of cryptic binding sites or post translational modification sites; higher degree of connectivity between filamin monomers as more β strand edges are exposed; a higher turnover rate as more residues are exposed to potential cleavage; forming of new F-actin structures (cross-links vs. bundles); attraction of heat shock proteins to refold domains as stress increases; allowing Ig-like domains to more quickly refold back to their native fold when stresses and then are relaxed; or allowing multiple domains to fold properly(195). For example, proteomic screens revealed two phospho-serines (S^{1081} and S^{1084}) (91, 92) in the loop between β strands A and B of domain 9 of filamin A. Thus, we postulate that the stress-induced unfurling of the N-terminal β strands of filamin domains could play an important cellular role as a conformational switch.

Chapter 4
Discussion

4.1 Summary

In Chapters 2 and 3 we analyzed the functional aspects of filamin from two different perspectives. In Chapter 2 we analyzed the functional divergence of filamin isoforms. During this analysis we develop a structural model of filamin A; a phylogenetic analysis of invertebrate, invertebrate chordate, and vertebrate filamin; and an evolutionary trace (ET) of the vertebrate filamin family, which we use to observe the divergence of individual filamin residues in a spatial and temporal context. In Chapter 3 we analyzed the mechanical function of filamin A. During this analysis we used molecular dynamics to perform single-molecule force-induced unfolding experiments. For each Ig-like domain, we observe critical unfolding forces and low force (35-70 pN) unfolding pathways. We also examined the effect of thermal unfolding on Ig domains.

In each analysis we used novel techniques. In our ET analysis, we create an all-atom homology model of a filamin A monomer. We used this model to observe spatial and temporal divergence of family members in the context of a quaternary structure. We also used a highly constrained version of ET to restrict our results to functionally informative sites. In our analysis of the mechanical properties of filamin Ig domains, we use Discrete Molecular Dynamics (DMD) as the molecular dynamics technique. To our knowledge, this technique has not been applied previously to study the mechanical properties of proteins. The efficiency of DMD allows us to perform multiple replicated simulations instead of a single

simulation using traditional molecular dynamics. In addition, we developed a new method of determining critical unfolding force from constant force trajectories. This method is based upon finding the intersection between two unfolding regimes with respect to pulling forces. Our results are in agreement with published data and suggest the validity of our techniques.

Validation of results

Our all-atom model has dimensions and architecture of filamin that are similar to those observed in electron micrographs (18, 62, 161). Recent electron micrographs and experimental data (62) suggest that filamin A has two distinct architectural regions, one consisting a linear sequence of domains (rod 1) and the other a compact group of domains (rod 2). This linear region is similar to the high-resolution crystal structure of ddFilamin domains 4-6 (PBD ID 1WLH) (155). The compact region results from two domains, 18 and 20, which lack β strand A. As further validation of the modeling of the compact region, our pulling data suggest that Ig domains 18 and 20 are stable, even though they lack β strand A.

Our phylogenetic study of vertebrate filamin suggests that the three isogenes of filamin (A, B, and C) originated from a single ancestral gene that existed prior to the advent of vertebrates, and that this gene is also the progenitor of extant urochordate filamin. If this model is correct, then some functions shared among vertebrate isoforms of filamin should also be common to extant urochordate filamin. All three isoforms of filamin interact with all β1 integrin family members through a highly ancestral binding surface on domain 21 (21, 47, 86). In addition, each isoform has a low number of class-distinctive sites in domain 21, an indication of lack of isoform-specific functional divergence. Therefore, we posit that the interaction of filamin with integrins predates vertebrate evolution. Support for this hypothesis

comes from humans and *Ciona* transcripts of integrins (Figure 4.1) that have high sequence similarity in the filamin-binding region. The highest sequence similarity is between human β3 integrin and *Ciona* β1 integrin.

Validation of our ability to identify sites critical for filamin functional divergence comes from a recent study of the interaction of FILGAP with filamin domain 23. We identify a class-distinctive residue on each of filamin B and filamin C that subsequently was shown to be critical for the distinct binding of filamin A to FILGAP(175). We also identify sites that suggest isoforms have diverged in their ability to regulate their interactions with binding partners. *In vitro* studies confirm this possibility (5, 49). These binding partners include migfilin and integrins. Furthermore, we find many instances in which functionally divergent


```
hB1A   NPIYKSAVTT  VVNP
hB2    NPIYKSAVTT  VVNP
hB3    NPLYKEATST  FTNI
hB7    NPLYKSAITT  TINP
cB1    NPIFKKATST  FKNP
cB2    NPIFEEASTR  FENP
cB3    NPIFEEASTR  FENP
```

Figure 4.1. Alignment of β1 integrin family members in the filamin A binding region. Top four sequences are human (h), rest are from *Ciona* (c) (1, 2).

sites co-localize with disease-causing mutations, indicating biochemical functional significance.

Our pulling results are also commensurate with several reports. The average critical unfolding force for all Ig domains overlaps with reported atomic force microscopy (AFM) results — 150-240 pN (DMD) vs. 50-220 pN (AFM) (16). In addition, the stable intermediates we observe in filamin Ig-like are similar to the intermediates ddFilamin4, TNI27, and FN-III$_{10}$. Also similar are the N-terminal unfoldings of β strands. In addition, the

range of forces we observe in the unfolding of filamin Ig-like domains (35-70 pN) is very similar to the range observed in FN-III$_{10}$ and ddFilamin4 (Table 3). Our thermal unfolding results for domain 21 (330-360 K) are also similar to reported results (319 K) (180).

Functional divergence of isoforms as scaffold proteins

We also hypothesize that our distinctive residues may contribute to the evolution of isoform-distinctive behavior in several ways. These include variable specificity in binding; differential regulation of binding of adhesion proteins, effector proteins, and F-actin; and potentially distinctive modulation of inter-domain structure. For example, over 35% of distinctive residue changes in filamin A and C involve changes to or from S/T residues, while only 11% of distinctive changes in B involve S/T residues. Thus, filamin isoforms are likely to have different roles in signaling as phosphoproteins. From our analysis of the mechanical properties of Ig domains, it is possible that some of S/T are exposed only on force-induced unfolding. For example, in filamin A, proteomic screens revealed two phospho-serines in the loop between β strands A and B of domain 9 (91, 92). Thus, force-induced conformational changes may lead to differential scaffolding activity.

Functional divergence of isoforms as mechanical proteins

Since filamin isoforms play critical roles as mechanical proteins (18), we also hypothesize that some of the distinctive residues observed may mediate isoform-specific structural characteristics. These characteristics may translate into isoform-distinct structural roles in cells. For example, in Chapter 2 we identify a cluster of distinctive sites in ABD CH 2 that co-localize with mutations causing skeletal disorders. The function of CH 2 is not

currently understood; however, the ABD CH 2 has been reported to modulate the affinity of filamin for F-actin (169). Thus, we posit that distinctive changes in the ABD may also modulate the affinity of filamin isoforms for F-actin. The resulting ramifications of these distinctive changes could potentially be related to isoform-distinct roles in differentiation (55, 57). Further experimental testing of the distinctive residues is required to understand their isoform-specific role.

Several missense mutations along rod 1 of filamin A and B also cause skeletal disorders, whose etiology is unknown. One hypothesis is that mutations interfere with the structure of filamin. Since these missense mutations are not buried, it is difficult to understand how they could pertain to structure. However, ET analysis can identify residues that are involved in the structural divergence of isoforms. The structural divergence of mechanical proteins may be related to how their domains respond to force-induced unfolding. Since some distinctive residues in rod 1 co-localize near disease mutations, it is possible that these mutations could interfere with force-induced unfolding. In a model presented by Gehler et al., (57) filamin acts as a stress sensor in extracellular matrix (ECM) compliance matching (matching of ECM stiffness to intracellular stress), leading to cellular differentiation. The failure of filamin Ig domains to unfold properly under force may cause filamin to fail as a stress sensor, leading to disease.

Importance of dual evolutionary constraints on filamin function

We hypothesize that filamin Ig-like domains hold unique structural properties because in both thermal- and forced-unfolding pathways, unfolding is initiated by the unraveling of N-terminal β strands. This is an unexpected result (192) since thermal

denaturation can affect all bonds of a domain, whereas denaturation is localized during forced unfolding. Therefore, we speculate that filamin domains have evolved in this particular way because of dual constraints on filamin, as both a scaffold protein and a mechanical protein. Under stress, N-terminal β strand unfolding may protect binding of ligands to the CD face. Currently, the only structurally characterized binding method is by anti-parallel β sheet extension of β strand C (5). Alternatively, unfolding of N-terminal β strands may allow filamin to act as a, mechanosensory scaffold protein, by disrupting or creating binding surfaces. For example, the HIV-1 receptor was identified as binding to filamin A Ig domain 10, β strand G (196), and its interaction with filamin would likely be disrupted by β strand A unfolding.

Importance of heterogeneity of domain characteristics

In Chapter 3, we observe heterogeneity in domain mechanical characteristics. We hypothesize that this heterogeneity can allow filamin to act as a multiport switch that is tuned to the forced unfolding of other domains. Gehler et al. reported, that Filamin A is a component of the acromyosin contractility pathway (Figure 1.5), which has a role in ECM remodeling (57). Critically, the ECM is composed of fibronectin (FN), which can unravel the N-terminus of its domains under forces (Table 3) similar to those used to unfold filamin (137, 197). Since filamin and FN are in the same stress pathway, we postulate that they can partially unfold simultaneously, leading to coordinated efficient inside-out and outside-in signaling. One way to test this hypothesis would be to simulate unfolding of pairs of individual fibronectin and filamin Ig domains to see if they cooperatively unfold. Identified pairs can then be tested biochemically, either by over-expression or deletion studies. The

observation by Lu et al. (55), that differentiation of proliferating chondrocytes in the growth plate is delayed in filamin B-null mutant chondrocytes, could potentially be explained by inefficient signal transduction due to loss of filamin B.

4.2 Future directions and conclusions

We have analyzed the functional divergence of filamin isoforms, and determined sites we feel are important for their divergences. Thus, our work provides the basis for site-directed mutagenesis studies. In addition, our analysis of force-induced unfolding of Ig-like domains suggests that the filamin molecules act as a mechanosensory scaffold proteins when cells are placed under stress. One way to test this hypothesis is to use an immunogenic approach, by creating an antibody to a force-induced conformation. We have also suggested that disease mutations may disrupt switch-like function. To test this hypothesis, we can perform forced-unfolding unfolding simulation on mutant Ig-like domains. Our model also suggests a reason for splice variants lacking hinge 1. We posit that lacking hinge 1 would make filamin more sensitive to stress. Thus hinges may control stress tolerance. A cell may require this tolerance to protect itself from premature differentiation(21).

Thus, viewing filamins as both a structural protein and a stress-induced dynamic scaffold can potentially shed light on the commonality among disease mutations as well as the roles of distinct isoforms in acromyosin contractility, mechanoprotection, and cell differentiation. Thus, our findings may lead to our understanding of the etiology of many diseases (Table 2) including prolapsed valve (198), muscular dystrophy, and Alzheimer's disease (199).

The work contained within this dissertation extended our knowledge of the function of filamin. While most experimentalists have analyzed filamin from the perspective of their

particular protein, we have taken a top-down approach, allowing us to identify, without bias, residues critical to the function of filamin. We have also identified mechanical properties of filamin Ig domains which we postulate enable filamin to sense stress. The methods we employ in this dissertation can be used to understand the function of many protein families, from coronins to viral coat proteins. Since studying the biochemistry of proteins under stress is a difficult task, our work will forge the way for others to use the tools we have developed.

A
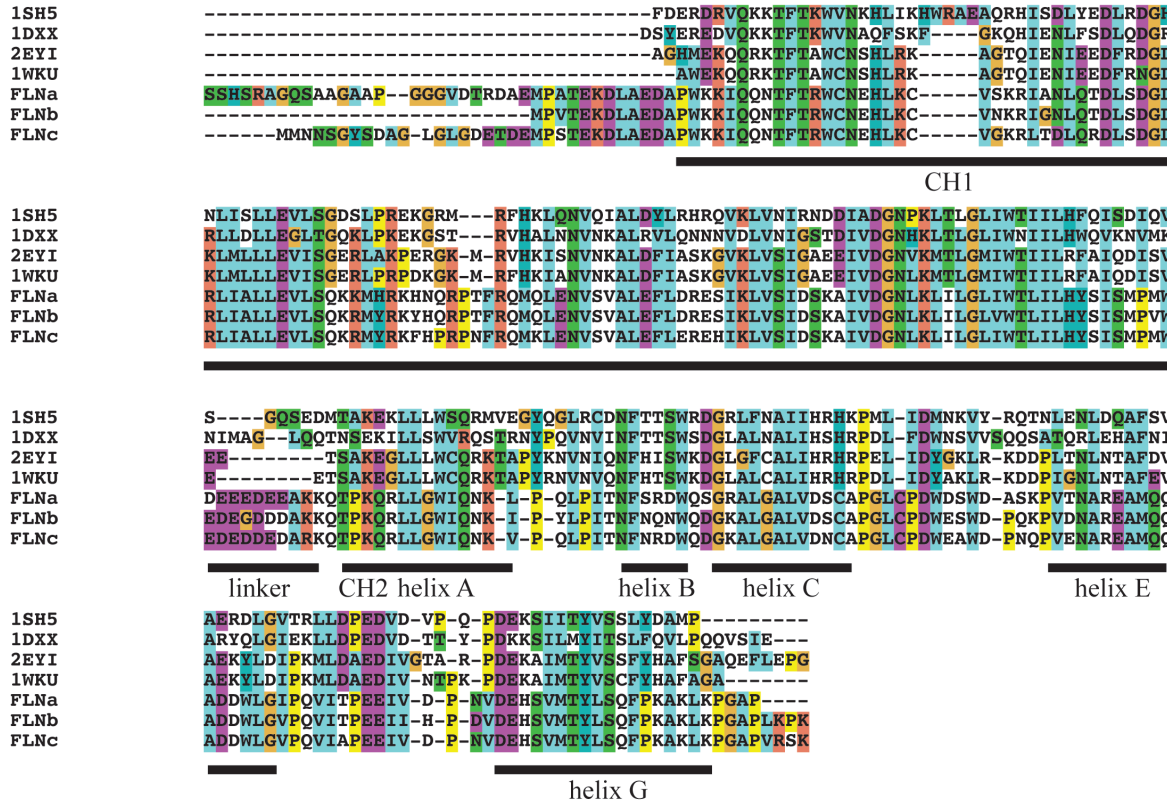
Figure 1. Multiple sequence alignment used to obtain homology model templates of filamin domains. (**A**) Alignment of homologous actin-binding domains (ABD) from plectin (PDB ID 1SH5), dystrophin (PDB ID 1DXX), α-actinin 1 (PDB ID 2EYI), α-actinin 3 (PDB ID 1WKU), human filamin A, human filamin B, and human filamin C. Muscle specific α-actinin 3 is used as a template to model filamin A's ABD. (**B**) Alignment of target and template (asterisk (*)) filamin Ig-like domains. Excluded are predicted Ig-like domains that have close homologs in the PDB (e.g., we do not include the sequence IgFLNa12 along with IgFLNb12 (PDB ID 2DIC)). We show all isoforms of domain 16 for comparison.

B

```
1a      KLNPKKARAYGPGIEPTGNMVKKRAEFTVETRS----------AGQGEVLVIYVEDPA----------GHQEEAKVTANNDKNRIFSVWYVPEVTGTHKVTVLFAGQ-HIAKSPFEVVY--
2a      QGDASKVTAQGPGLEPSGNIANKTTYFEIFTAG----------AGIGEVEVVIQDPM----------GQKGTVEPQLEARGDSIYRCSYQPTMEGVHTVHVTFAGV-PIPRSPYTVYV--
3a      ACNPSACRAVGRGLQPKGVRVKETADFKVYTKG----------AGSGELKVTVKGPK----------GEERVKQKDLGDGVYGFEYYPMVPGTIIVTITWGGQ-NIGRSPFEVKV--
4a      ECGNQKVRAWGPGLE---GGVGKSADFVVEAIG----------DDVGTLGFSVEGPS----------QAKIECDDKGDGSCDVRYWPQEAGEYAVHVLCNSE-DIRLSPFMADI--
5a      DFHPDRVKARGPGLEKTGVAVNKPAEFTVDAKH----------GGKAPLRVQVQDNE----------GCPVEALVKDNGNGTYSCSYVPRKPVKHTAMVSWGGV-SIPNSPFRVNV--
6a      GSHPNKVKYGPGVYAKTGLKAHEPTYFTVDCAE----------AGQGDVSIGIKCAPGVVGPAEAIDFDIIRNDNDGTYAVTYTPRGAGSYTIMVLFADQ-ATTSPIRVKV--
7a      SHDASKVRAEGPGLSRTGVEEGKPTHFTVNAKA----------AGKGKLDVQFSGLIK----------GDAVRDVDIIDHHDNTYTVKYTPVQQGPVGVNVTYGGD-PIPKSPFSVAV--
8a      SLDLSKIKVSGLGEK---VDVGKDQEFTVKSKG----------AGGQGKVASKIVGPS----------GAAVPCKVEPGLGADNSVVRFLPREEGPYEVEVTYDGV-PVPGSPFPLEA--
9b*     PPPPSKVKAHGPGLE---GGLVGKPAEFTIDTKG----------AGTGGLGLTIVEGPC----------EAKIECSDNGDGTCSVSYLPTKPGEYFVNILEEV-HIPGSPFKADI--
110b*   PFDPSKVVASGPGLE---HGKVGEAGLLSVDCSE----------AGPGALGLEAVSDS----------GTKAEVSIQNNKDGTYAVTYVPLTAGMYTLTMKYGGE-LVPHFPARVKV--
11a     AVDTSGVQCYGPGIEQQGVFREATTEFSVDARA----------LTQTGGPHVKARVANPS----------ESKINCRDNKDGSCSAEIIPFAPGDIDVNITYGGA-HIPGSPFRVPV--
12b*    GCQPSRVQAQGPGLK---EAFINKPNVFTVVTRG----------AGIGGLGITVEGPS----------GLVEPVNVVDNGDGTHTVTYTPSQEGPYMVSVKYADE-EIRSPFKVKV--
13b*    VVDPSKVKLAGPGLGS-GVRARVLQSFTVDSSK----------AGLAPLEVRLGPR----------GRPKKANIRDNGDGTVTVSYLPDMSGRYTITIKYGGD-EIYSPFRIR--
14c*    AHDASKVRASGPGLNASGIPASLPVEFTIDARD----------AGEGLTVQILDPE----------GTHEAADVIENEDGTHDIFYTAAKPGTYVTYVRFGGV-DIPNSPFTVMA--
15b*    TGDASKCLATGPGIAS-TVKIQGEEVGFVVDAKT----------AGKGKVTCTVLTPD----------GTEAEADVIENEDGTYDIFYTAAKPGTYVITYVRFGGV-DIPNSPFTVMA--
16a     TWAPERPLVGVNGLD---VTSLRPFDLVIP--TLKK--------GEITGEVRMPS----------GKVAQPTIIDNKDGTVTVRYAPSEAGL-HEMDIRYDNM--HIPGSPLQFYV--
16b     TEEAYVPYSDMNGLG---PKPFDLVIP---FAVRK--------GEITGEVHMPS----------GKTAPTEIVDNKDGTVTVRYAPTEVGL-HEMHIKYMGS--HIPGSPLQFYV--
16c*    TEEPVVPVEPMESM----LRPFNLVIP---FAVQK--------GEITGEVRMPS----------GKTARPNITDNKDGTITVRYAPTEKGL-HQMGIKYDGN--HIPGSPLQFYV--
17a*    VVNCGHVTAYGPGIT---HGVVNKPATFTVNTKD----------AGEGGLSLAIEGPS----------KAEISCTDNQGICSVSLPVLPGDVSILVKYNEQ--HVPGSPFTARV--
17c*    AINSRHVSAYGPGLS---HGMVNKPATFTIVTKD----------AGEGGLSLAVEGPS----------KAEITCKDNKDGGTCTVSYLPTAPGDISIIVRFDDK--HIPGSPFTAKI--
18b*    SFTAKITDDSRRCS---QVKLGSAADFLDISE----------TDLSSLTASIKAPS----------GRDEPCLLKRLPNNHIGISFIPREVGEHLVSIKKNGN--HVANSPVSIMV--
19b*    IGDARRAKYGRGLS---EGRIFEMSDFIVDTRD----------AGYGGISLAVEGPS----------KVDIQTEDLEDGICKVSFPFVPGVIVSTKFADE--HVPGSPFTVKI--
20a     ---------------VGSHCDLSLKIPEISI----------QDMTAQ-VT--SPS----------GKTHEAEIVEGENHIYCIRFVPAEMGTHTVSVKYKGQ--HVPGSPFQFYV--
21a     EGGAHKVRAGGPGLE---RAEAGVPAEFSIWTRE----------AGAGGLAIAVEGPS----------KAEISFEDRKDSCGGVAIVVQEFPDIEVSKFNEE--HIPDSPFVVPV--
22c*    SDDARRLTVTS---LQETGLKVNQPASFAVQLNG----------A-RGVIDARVHTPS----------GAVEECYVSELDSDKHITRFIPHENGVHSIDVKFNGA--HIPGSPFKIRV--
23c*    AGDPGLVSAYGPGLE---GGTTGVSSEFIVNTLN----------ASSGALSVTIDGPS----------KVQLDCRECPEG-HVVTYTPMAPGNYLIAIKYGGPQHIVGSPFKAKVT-
24c*    -SDASKVTRGPGLS---QAFVGQKNSFTVDCSK----------AGTNMMMVGVHGPK----------TPCEEVYVKHMGNRVNNVTYTVKEKGDYILIVKWGDE-SVPGSPFKVRV--
ddFLN4  AADPEKSYAEGPGLDGG---ECFQPSKFKIHAVD----------PDGV-DGGDGFVVT----------IEGPAPVDPV--MVDNGDGTYDVEFEPKEAGDYVINLTLDGDNVNGFPKIVVV-
```

90

Figure 2. Conservation within domains is bi-modal. The number of residues in each domain that are conserved to the common ancestor across filamin isoforms. N-terminal domains 1-16 alternate between the two conservation modes and have a mean of 62±9%. C-terminal domains 17-24, have a mean are more highly conserved overall (mean 71±8%). The histogram, right is bimodal, with a mode at ~70% and a mode at ~55%. Individual highly conserved domains are CH 1 (88%), IgFLNa21 (84%) and IgFLNa22 (79%). Domains with the least conservation include 5, 8, 10, 11 (~50%).
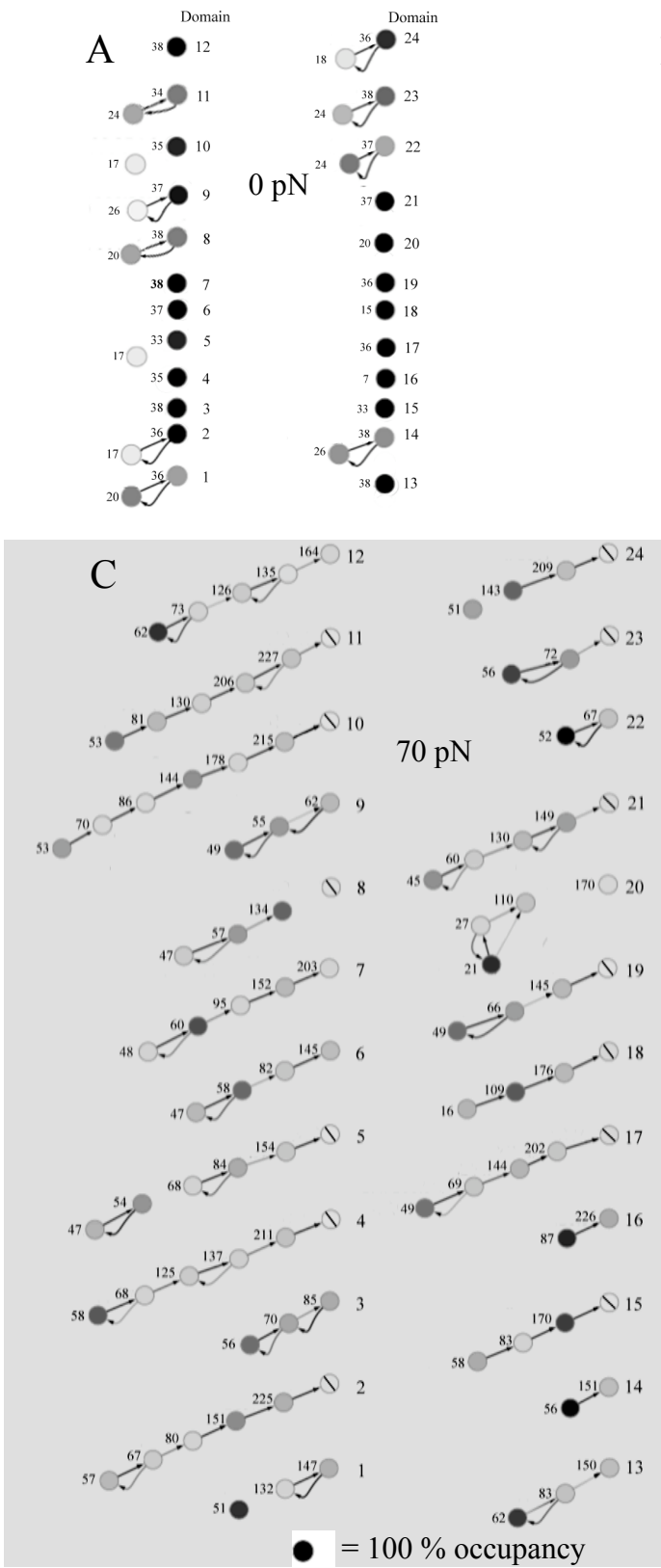
Figure 3. Force-induced unfolding pathways of filamin Ig-like domains. Each filamin domain is subjected to constant opposing force acting on the Nt and Ct Cα atoms for 1 x 10⁶ time units (t.u.). The last 0.5 x 10⁶ t.u., are examined for stable intermediates (circles) using a histogram of extension lengths (x). Minimum occupancy of a state is ~ 5%. Directed graphs represent unfolding pathways. Circle density represents trajectory occupancy, arrow density represents transition probability, circle with bar represents unfolded. Numbers to left of state represents mean x of stable intermediate. (**A**) At 0 pN pulling force, most domains are native (~ 40 Å), shorter states are from floppy N-terminus. (**B**) At 35 pN pulling force most domains have a native pulled state is ~ 47 Å x. However, some domains have a less populated state at ~ +10 Å. Domains lacking β strand A (16, 18, 20) and A' (16) are less stable, having secondary intermediates with larger jumps (20-30 Å), or unfold. (**C**) At 70 pN pulling force, domains have varying unfolding pathways. Some domains are stable close to their native-like x (e.g., 3, 9, 22), while others fully unfolded. Stable pathways tend to cycle between the first two metastable states. Pathways having a third state, with x in the range 140-170 Å; have partially unfolded ensembles, which lack N-terminal β strands. Domains 16, 18, and 20 appear similar to 7 β stranded counterparts.

Appendix B
Additional tables

Table 1. Filamin domain homology templates

Domain (target)	Template(s)	PDB ID	Method	Verify score
ABD	Human a actinin 3 closed conformation	1wku	X-ray	0.75
Ig-like				
1	hsIgFLNb9	2di9	NMR	0.72
	hsIgFLNc14	2e9j	NMR	
2	hsIgFLNc14	2e9j	NMR	1.12
3	hsIgFLNb19	2di9	NMR	1.42
4	hsIgFLNa17	2bp3	X-ray	0.87
5	hsIgFLNc14	2e9j	NMR	0.88
	ddFLN5 from ddFLN4-6	1wlh	X-ray	
6	hsIgFLNc14	2d9j	NMR	0.88
	ddFLN5 from ddFLN4-6	1wlh	X-ray	
7	hsIgFLNc14,	2d9j	NMR	0.97
	ddFLN5 from ddFLN4-6	1wlh	X-ray	
8	hsIgFLNb10	2dia	NMR	0.73
	hsIgFLNb19	2di8	NMR	
	ddFLN5 from ddFLN4-6	1wlh	X-ray	
9	hsIgFLNb9	2di9	NMR	1.14
10	hsIgFLNb10	2dia	NMR	1.01
11	hsIgFLNb13	2dj4	NMR	0.51
12	hsIgFLNc12	2dic	NMR	0.89
13	hsIgFLNb13	2dj4	NMR	0.58
14	hsIgFLNc14	2d9j	NMR	0.85
15	hsIgFLNb15	2dmb	NMR	0.72
16	hsIgFLNc10	2dia	NMR	0.64
	hsIgFLNc16	2d7n	NMR	
17	hsIgFLNa17	2bp3	X-ray	0.98
18	hsIgFLNb18	2dmc	NMR	0.11*
19	hsIgFLNb19	2di8	NMR	1.17
20	hsIgFLNb18	2dmc	NMR	0.65
21	hsIgFLNb19	2di8	NMR	1.13
22	hsIgFLNc22	2dlg	NMR	1.25
23	hsIgFLNc23	2d7q	NMR	1.45
24	hsIgFLNc24	1v05	X-ray	1.04

Table 2. Linkers between Ig-like domains

Ig domain pairs	Linker length	Short linker
1-2	4	^{375}NKSQ^{379}
2-3	2	^{476}GA^{477}
3-4	2	^{572}TE^{573}
4-5	6	^{664}RDAPQD^{669}
5-6	3	^{764}GAG^{766}
6-7	3	^{867}EPS^{869}
7-8	3	^{966}SPS^{969}
8-9	3	^{1062}VAP^{1064}
9-10	3	^{1155}VPC^{1157}
10-11	3	^{1250}EPA^{1253}
11-12	3	^{1350}TEG^{1352}
12-13	3	^{1443}HDV^{1445}
13-14	3	^{1540}LPT^{1542}
14-15	3	^{1637}VPT^{1638}
16-17	2	^{1861}DY^{1862}
18-19	4	^{2039}SQSE^{2042}
20-21	5	^{2231}GPLGE^{2235}
21-22	5	^{2226}ASPSG^{2330}

Table 3. Filamin protein sequences

Protein name	ID
Human filamin A	NP_001447.2
Canis familiaris filamin A	ENSCAFT00000031079
Mus musculus filamin A	ENSMUST00000052678
Rattus norvegicus filamin A	ENSRNOT00000012796
Xenopus tropicalis filamin A	ENSXETT00000031243
Danio rerio filamin A	ENSDART00000046261
Tetraodon nigroviridis filamin A	GSTENT00008691001
Takifugu rubripes filamin A	NEWSINFRUT00000159596
Human filamin B	NP_001448.2
Canis familiaris filamin B	ENSCAFT00000012236
Mus musculus filamin B	ENSMUST00000052678
Rattus norvegicus filamin B	ENSRNOT00000012796
Xenopus tropicalis filamin B	ENSXETT00000018970
Danio rerio filamin B	ENSDART00000086028
Tetraodon nigroviridis filamin B	GSTENT00012499001
Takifugu rubripes filamin B	NEWSINFRUT00000147431
Human filamin C	NP_001449.3
Canis familiaris filamin C	ENSCAFT00000002504
Mus musculus filamin C	ENSMUST00000065090
Rattus norvegicus filamin C	XP_342654.2
Xenopus tropicalis filamin C	ENSXETT00000032738
Danio rerio filamin C	ENSDART00000073442
Tetraodon nigroviridis filamin C	GSTENT00029754001
Takifugu rubripes filamin C	NEWSINFRUT00000181322
Ciona intestinalis, Ciona savignyi	ENSCINT00000006732 (42,48)/ ENSCING00000003272 (gene), ENSCSAVT00000017772/ ENSCSAVG00000010354(gene)
Strongylocentrotus purpuratu	XP_784715.2
Anopheles gambiae	AGAP004335-PA

Table 4. Filamin domain ancestral conservation

FLN domain	Isoform common ancestral residues	Range (filamin A)	Residues
ABD			
CH 1	97	40-151	112
CH 2	75	168-264	97
Ig-like			
1	62	279-374	96
2	52	379-475	97
3	67	478-71	94
4	90	574-663	90
5	45	670-763	94
6	67	767-866	100
7	64	870-965	96
8	46	969-1061	93
9	67	1065-1154	90
10	46	1158-1249	92
11	48	1253-1349	97
12	62	1352-1442	91
13	55	1446-1539	94
14	67	1543-1636	94
15	56	1640-1744	105
16	53	1771-1860	90
17	66	1863-1952	90
18	44	1968-2039	72
19	64	2044-2134	91
20	48	2159-2230	72
21	77	2236-2325	90
22	72	2331-2420	90
23	62	2427-2516	90
24	62	2553-2646	94

Table 5. Filamin domain isoform-distinctive content

Isoform	Domain	Ancestral Conservation[1]	Distinctive Modifications[2] (Teleostei period)
A	CH 1	High *c*	Low
	CH 2	High	High
	12	High	High
	14	High	No -
	17	High	High
	18	Low	High +
	21	High *i*	Low
B	CH 1	High *c*	Low -
	CH 2	High	High
	6	High	High
	12	High	High
	15	Low	High
	16	Low	High
	17	High	High
	18	Low	High +
	19	High	High
	20	High	Low -
	21	High *i*	Low
	22	High	Low
	23	High	Low
C	CH 1	High *c*	Low
	CH 2	High	Low
	6	High	Low
	9	High	No -
	11	High	High +
	12	High	Low
	16	High	High +
	17	High	No -
	19	High	Low
	21	High *i*	No -

[1] High/Low ancestral conservation implies greater than or less than 0.6 percent ancestral.
[2] High/Low distinctive residue implies per domain (A: > 6 / < 2 B: > 8 / < 4 C: > 6 / < 3).
c, i Most Ancestral Residues for Isoform (*c* = CH domain, *I* = Ig-like domain)
+/- Most/Least Distinctive Residues.
consecutive distinctive/ancestral

97

References

1.	Ewan R, Huxley-Jones J, Mould AP, Humphries MJ, Robertson DL, & Boot-Handford RP (2005) *BMC evolutionary biology* 5, 31.

2.	Nakamura F, Pudas R, Heikkinen O, Permi P, Kilpelainen I, Munday AD, Hartwig JH, Stossel TP, & Ylanne J (2006) *Blood* 107, 1925-1932.

3.	Vogel V & Sheetz MP (2009) *Current opinion in cell biology* 21, 38-46.

4.	Feng Y & Walsh CA (2004) *Nature cell biology* 6, 1034-1038.

5.	Kiema T, Lad Y, Jiang P, Oxley CL, Baldassarre M, Wegener KL, Campbell ID, Ylanne J, & Calderwood DA (2006) *Molecular cell* 21, 337-347.

6.	Sokol NS & Cooley L (1999) *Curr Biol* 9, 1221-1230.

7.	Cox D, Condeelis J, Wessels D, Soll D, Kern H, & Knecht DA (1992) *The Journal of cell biology* 116, 943-955.

8.	Sokol NS & Cooley L (2003) *Developmental biology* 260, 260-272.

9.	Cox D, Ridsdale JA, Condeelis J, & Hartwig J (1995) *The Journal of cell biology* 128, 819-835.

10.	Iwadate Y & Yumura S (2008) *Journal of cell science* 121, 1314-1324.

11.	Sabass B, Gardel ML, Waterman CM, & Schwarz US (2008) *Biophysical journal* 94, 207-220.

12.	Stossel TP, Condeelis J, Cooley L, Hartwig JH, Noegel A, Schleicher M, & Shapiro SS (2001) *Nature reviews* 2, 138-145.

13.	Nagano T, Morikubo S, & Sato M (2004) *J Neurosci* 24, 9648-9657.

14. Nagano T, Yoneda T, Hatanaka Y, Kubota C, Murakami F, & Sato M (2002) *Nature cell biology* 4, 495-501.

15. Nakamura F, Hartwig JH, Stossel TP, & Szymanski PT (2005) *The Journal of biological chemistry* 280, 32426-32433.

16. Furuike S, Ito T, & Yamazaki M (2001) *FEBS letters* 498, 72-75.

17. Gardel ML, Nakamura F, Hartwig JH, Crocker JC, Stossel TP, & Weitz DA (2006) *Proceedings of the National Academy of Sciences of the United States of America* 103, 1762-1767.

18. Hartwig JH & Stossel TP (1981) *Journal of molecular biology* 145, 563-581.

19. Gorlin JB, Yamin R, Egan S, Stewart M, Stossel TP, Kwiatkowski DJ, & Hartwig JH (1990) *The Journal of cell biology* 111, 1089-1105.

20. Sheen VL, Feng Y, Graham D, Takafuta T, Shapiro SS, & Walsh CA (2002) *Hum Mol Genet* 11, 2845-2854.

21. van der Flier A, Kuikman I, Kramer D, Geerts D, Kreft M, Takafuta T, Shapiro SS, & Sonnenberg A (2002) *The Journal of cell biology* 156, 361-376.

22. Li M, Bermak JC, Wang ZW, & Zhou QY (2000) *Molecular pharmacology* 57, 446-452.

23. Lin R, Canfield V, & Levenson R (2002) *Pharmacology* 66, 173-181.

24. Li M, Li C, Weingarten P, Bunzow JR, Grandy DK, & Zhou QY (2002) *Biochemical pharmacology* 63, 859-863.

25. Lin R, Karpa K, Kabbani N, Goldman-Rakic P, & Levenson R (2001) *Proceedings of the National Academy of Sciences of the United States of America* 98, 5258-5263.

26. Hjalm G, MacLeod RJ, Kifor O, Chattopadhyay N, & Brown EM (2001) *The Journal of biological chemistry* 276, 34880-34887.

27.   Thelin WR, Chen Y, Gentzsch M, Kreda SM, Sallee JL, Scarlett CO, Borchers CH, Jacobson K, Stutts MJ, & Milgram SL (2007) *The Journal of clinical investigation* 117, 364-374.

28.   Lu J, Tiao G, Folkerth R, Hecht J, Walsh C, & Sheen V (2006) *The Journal of comparative neurology* 494, 476-484.

29.   Behrendt B (2002) *Trends in cell biology* 12, 213.

30.   Nikki M, Merilainen J, & Lehto VP (2002) *The Journal of biological chemistry* 277, 11432-11440.

31.   Kessels MM & Qualmann B (2002) *EMBO J* 21, 6083-6094.

32.   Loy CJ, Sim KS, & Yong EL (2003) *Proceedings of the National Academy of Sciences of the United States of America* 100, 4562-4567.

33.   Ozanne DM, Brady ME, Cook S, Gaughan L, Neal DE, & Robson CN (2000) *Molecular endocrinology (Baltimore, Md* 14, 1618-1626.

34.   Wang Y, Kreisberg JI, Bedolla RG, Mikhailova M, deVere White RW, & Ghosh PM (2007) *Oncogene* 26, 6061-6070.

35.   Sasaki A, Masuda Y, Ohta Y, Ikeda K, & Watanabe K (2001) *The Journal of biological chemistry* 276, 17871-17877.

36.   Zheng L, Baek HJ, Karsenty G, & Justice MJ (2007) *The Journal of cell biology* 178, 121-128.

37.   Nunnally MH, D'Angelo JM, & Craig SW (1980) *The Journal of cell biology* 87, 219-226.

38.   Playford MP, Lyons PD, Sastry SK, & Schaller MD (2006) *The Journal of biological chemistry* 281, 34104-34112.

39.   van der Ven PF, Obermann WM, Lemke B, Gautel M, Weber K, & Furst DO (2000) *Cell motility and the cytoskeleton* 45, 149-162.

40. van der Ven PF, Ehler E, Vakeel P, Eulitz S, Schenk JA, Milting H, Micheel B, & Furst DO (2006) *Experimental cell research* 312, 2154-2167.

41. van der Ven PF, Wiesner S, Salmikangas P, Auerbach D, Himmel M, Kempa S, Hayess K, Pacholsky D, Taivainen A, Schroder R*, et al.* (2000) *The Journal of cell biology* 151, 235-248.

42. Beatham J, Romero R, Townsend SK, Hacker T, van der Ven PF, & Blanco G (2004) *Hum Mol Genet* 13, 2863-2874.

43. Lowe T, Kley RA, van der Ven PF, Himmel M, Huebner A, Vorgerd M, & Furst DO (2007) *Hum Mol Genet* 16, 1351-1358.

44. Andrews RK & Fox JE (1991) *The Journal of biological chemistry* 266, 7144-7147.

45. Feng S, Resendiz JC, Lu X, & Kroll MH (2003) *Blood* 102, 2122-2129.

46. Sharma CP, Ezzell RM, & Arnaout MA (1995) *J Immunol* 154, 3461-3470.

47. Loo DT, Kanner SB, & Aruffo A (1998) *The Journal of biological chemistry* 273, 23304-23312.

48. Takala H, Nurminen E, Nurmi SM, Aatonen M, Strandin T, Takatalo M, Kiema T, Gahmberg CG, Ylanne J, & Fagerholm SC (2008) *Blood*.

49. Lad Y, Jiang P, Ruskamo S, Harburger DS, Ylanne J, Campbell ID, & Calderwood DA (2008) *The Journal of biological chemistry* 283, 35154-35163.

50. Kim H, Sengupta A, Glogauer M, & McCulloch CA (2008) *Experimental cell research* 314, 834-846.

51. Ohta Y, Suzuki N, Nakamura S, Hartwig JH, & Stossel TP (1999) *Proceedings of the National Academy of Sciences of the United States of America* 96, 2122-2128.

52. Bellanger JM, Astier C, Sardet C, Ohta Y, Stossel TP, & Debant A (2000) *Nature cell biology* 2, 888-892.

53.	Ohta Y, Hartwig JH, & Stossel TP (2006) *Nature cell biology* 8, 803-814.

54.	Krakow D, Robertson SP, King LM, Morgan T, Sebald ET, Bertolotto C, Wachsmann-Hogiu S, Acuna D, Shapiro SS, Takafuta T*, et al.* (2004) *Nature genetics* 36, 405-410.

55.	Lu J, Lian G, Lenkinski R, De Grand A, Vaid RR, Bryce T, Stasenko M, Boskey A, Walsh C, & Sheen V (2007) *Hum Mol Genet* 16, 1661-1675.

56.	Mammoto A, Huang S, & Ingber DE (2007) *Journal of cell science* 120, 456-467.

57.	Gehler S, Baldassarre M, Lad Y, Leight JL, Wozniak MA, Riching KM, Eliceiri KW, Weaver VM, Calderwood DA, & Keely PJ (2009) *Mol Biol Cell*.

58.	Glogauer M, Arora P, Chou D, Janmey PA, Downey GP, & McCulloch CA (1998) *The Journal of biological chemistry* 273, 1689-1698.

59.	D'Addario M, Arora PD, Fan J, Ganss B, Ellen RP, & McCulloch CA (2001) *The Journal of biological chemistry* 276, 31969-31977.

60.	Kainulainen T, Pender A, D'Addario M, Feng Y, Lekic P, & McCulloch CA (2002) *The Journal of biological chemistry* 277, 21998-22009.

61.	Shifrin Y, Arora PD, Ohta Y, Calderwood DA, & McCulloch CA (2009) *Mol Biol Cell* 20, 1269-1279.

62.	Nakamura F, Osborn TM, Hartemink CA, Hartwig JH, & Stossel TP (2007) *The Journal of cell biology* 179, 1011-1025.

63.	Kanaji T, Russell S, & Ware J (2002) *Blood* 100, 2102-2107.

64.	Zhou X, Tian F, Sandzen J, Cao R, Flaberg E, Szekely L, Cao Y, Ohlsson C, Bergo MO, Boren J*, et al.* (2007) *Proceedings of the National Academy of Sciences of the United States of America* 104, 3919-3924.

65.	Dalkilic I, Schienda J, Thompson TG, & Kunkel LM (2006) *Molecular and cellular biology* 26, 6522-6534.

66.   Matsudaira P (1994) *Seminars in cell biology* 5, 165-174.

67.   Pavalko FM, Otey CA, & Burridge K (1989) *Journal of cell science* 94 (Pt 1), 109-118.

68.   Takafuta T, Saeki M, Fujimoto TT, Fujimura K, & Shapiro SS (2003) *The Journal of biological chemistry* 278, 12175-12181.

69.   Zhou X, Boren J, & Akyurek LM (2007) *Trends in cardiovascular medicine* 17, 222-229.

70.   Marti A, Luo Z, Cunningham C, Ohta Y, Hartwig J, Stossel TP, Kyriakis JM, & Avruch J (1997) *The Journal of biological chemistry* 272, 2620-2628.

71.   Hall A (1998) *Science (New York, N.Y* 279, 509-514.

72.   Takai Y, Sasaki T, & Matozaki T (2001) *Physiol Rev* 81, 153-208.

73.   Wennerberg K, Rossman KL, & Der CJ (2005) *Journal of cell science* 118, 843-846.

74.   Tzima E (2006) *Circulation research* 98, 176-185.

75.   Mammoto A, Mammoto T, & Ingber DE (2008) *Current opinion in hematology* 15, 228-234.

76.   Jeon YJ, Choi JS, Lee JY, Yu KR, Kim SM, Ka SH, Oh KH, Kim KI, Zhang DE, Bang OS*, et al.* (2009) *EMBO reports* 10, 374-380.

77.   Goldmann WH (2002) *Cell biology international* 26, 567-571.

78.   Tigges U, Koch B, Wissing J, Jockusch BM, & Ziegler WH (2003) *The Journal of biological chemistry* 278, 23561-23569.

79.   Woo MS, Ohta Y, Rabinovitz I, Stossel TP, & Blenis J (2004) *Molecular and cellular biology* 24, 3025-3035.

80. Sarkisian MR, Bartley CM, Chi H, Nakamura F, Hashimoto-Torii K, Torii M, Flavell RA, & Rakic P (2006) *Neuron* 52, 789-801.

81. Garcia E, Stracher A, & Jay D (2006) *Archives of biochemistry and biophysics* 446, 140-150.

82. Liu G, Thomas L, Warren RA, Enns CA, Cunningham CC, Hartwig JH, & Thomas G (1997) *The Journal of cell biology* 139, 1719-1733.

83. West DK, Brockwell DJ, Olmsted PD, Radford SE, & Paci E (2006) *Biophysical journal* 90, 287-297.

84. Kim C, Cho Y, Kang CH, Kim MG, Lee H, Cho EG, & Park D (2005) *EMBO reports* 6, 1045-1051.

85. Chen Y, Thelin WR, Yang B, Milgram SL, & Jacobson K (2006) *The Journal of cell biology* 175, 169-178.

86. Pfaff M, Liu S, Erle DJ, & Ginsberg MH (1998) *The Journal of biological chemistry* 273, 6104-6109.

87. Takafuta T, Wu G, Murphy GF, & Shapiro SS (1998) *The Journal of biological chemistry* 273, 17531-17538.

88. Xu W, Xie Z, Chung DW, & Davie EW (1998) *Blood* 92, 1268-1276.

89. Fox JE (1985) *The Journal of clinical investigation* 76, 1673-1683.

90. Hartwig JH (2006) *Semin Hematol* 43, S94-100.

91. Daub H, Olsen JV, Bairlein M, Gnad F, Oppermann FS, Korner R, Greff Z, Keri G, Stemmann O, & Mann M (2008) *Molecular cell* 31, 438-448.

92. Dephoure N, Zhou C, Villen J, Beausoleil SA, Bakalarski CE, Elledge SJ, & Gygi SP (2008) *Proceedings of the National Academy of Sciences of the United States of America* 105, 10762-10767.

93.	Lad Y, Kiema T, Jiang P, Pentikainen OT, Coles CH, Campbell ID, Calderwood DA, & Ylanne J (2007) *EMBO J* 26, 3993-4004.

94.	Kubler MD, Jordan PW, O'Neill CH, & Watt FM (1991) *Journal of cell science* 100 (Pt 1), 153-165.

95.	D'Addario M, Arora PD, Ellen RP, & McCulloch CA (2002) *The Journal of biological chemistry* 277, 47541-47550.

96.	Sosinski J, Szpacenko A, & Dabrowska R (1984) *FEBS letters* 178, 311-314.

97.	Janson LW, Kolega J, & Taylor DL (1991) *The Journal of cell biology* 114, 1005-1015.

98.	Discher DE, Janmey P, & Wang YL (2005) *Science (New York, N.Y* 310, 1139-1143.

99.	Formigli L, Meacci E, Sassoli C, Squecco R, Nosi D, Chellini F, Naro F, Francini F, & Zecchi-Orlandini S (2007) *Journal of cellular physiology* 211, 296-306.

100.	Jay D, Garcia EJ, Lara JE, Medina MA, & de la Luz Ibarra M (2000) *Archives of biochemistry and biophysics* 377, 80-84.

101.	Vadlamudi RK, Li F, Adam L, Nguyen D, Ohta Y, Stossel TP, & Kumar R (2002) *Nature cell biology* 4, 681-690.

102.	Cunningham CC, Gorlin JB, Kwiatkowski DJ, Hartwig JH, Janmey PA, Byers HR, & Stossel TP (1992) *Science (New York, N.Y* 255, 325-327.

103.	Feng Y, Chen MH, Moskowitz IP, Mendonza AM, Vidali L, Nakamura F, Kwiatkowski DJ, & Walsh CA (2006) *Proceedings of the National Academy of Sciences of the United States of America* 103, 19836-19841.

104.	Hart AW, Morgan JE, Schneider J, West K, McKie L, Bhattacharya S, Jackson IJ, & Cross SH (2006) *Hum Mol Genet* 15, 2457-2467.

105.	Farrington-Rock C, Kirilova V, Dillard-Telm L, Borowsky AD, Chalk S, Rock MJ, Cohn DH, & Krakow D (2007) *Hum Mol Genet*.

106.   Robertson SP (2005) *Current opinion in genetics & development* 15, 301-307.

107.   Robertson SP, Twigg SR, Sutherland-Smith AJ, Biancalana V, Gorlin RJ, Horn D, Kenwrick SJ, Kim CA, Morava E, Newbury-Ecob R*, et al.* (2003) *Nature genetics* 33, 487-491.

108.   Sheen VL, Jansen A, Chen MH, Parrini E, Morgan T, Ravenscroft R, Ganesh V, Underwood T, Wiley J, Leventer R*, et al.* (2005) *Neurology* 64, 254-262.

109.   Goetsch SC, Martin CM, Embree LJ, & Garry DJ (2005) *Stem cells and development* 14, 181-187.

110.   Ferland RJ, Batiz LF, Neal J, Lian G, Bundock E, Lu J, Hsiao YC, Diamond R, Mei D, Banham AH*, et al.* (2009) *Hum Mol Genet* 18, 497-516.

111.   Smith SC, Oxford G, Baras AS, Owens C, Havaleshko D, Brautigan DL, Safo MK, & Theodorescu D (2007) *Clin Cancer Res* 13, 3803-3813.

112.   Gu X (1999) *Molecular biology and evolution* 16, 1664-1674.

113.   Cicchetti P, Mayer BJ, Thiel G, & Baltimore D (1992) *Science (New York, N.Y* 257, 803-806.

114.   Sadowski I, Stone JC, & Pawson T (1986) *Molecular and cellular biology* 6, 4396-4408.

115.   Mayer BJ, Hamaguchi M, & Hanafusa H (1988) *Nature* 332, 272-275.

116.   Innis CA, Shi J, & Blundell TL (2000) *Protein Eng* 13, 839-847.

117.   Madabushi S, Gross AK, Philippi A, Meng EC, Wensel TG, & Lichtarge O (2004) *J Biol Chem* 279, 8126-8132.

118.   Chakravarty S, Hutson AM, Estes MK, & Prasad BV (2005) *J Virol* 79, 554-568.

119.   Go N (1983) *Annual review of biophysics and bioengineering* 12, 183-210.

120. Klimov DK & Thirumalai D (2000) *Proceedings of the National Academy of Sciences of the United States of America* 97, 7254-7259.

121. Fowler SB & Clarke J (2001) *Structure* 9, 355-366.

122. Rief M, Gautel M, Oesterhelt F, Fernandez JM, & Gaub HE (1997) *Science (New York, N.Y* 276, 1109-1112.

123. Marszalek PE, Lu H, Li H, Carrion-Vazquez M, Oberhauser AF, Schulten K, & Fernandez JM (1999) *Nature* 402, 100-103.

124. Fowler SB, Best RB, Toca Herrera JL, Rutherford TJ, Steward A, Paci E, Karplus M, & Clarke J (2002) *Journal of molecular biology* 322, 841-849.

125. Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, & Fernandez JM (2000) *Prog Biophys Mol Biol* 74, 63-91.

126. Li H, Carrion-Vazquez M, Oberhauser AF, Marszalek PE, & Fernandez JM (2000) *Nature structural biology* 7, 1117-1120.

127. Fisher TE, Carrion-Vazquez M, Oberhauser AF, Li H, Marszalek PE, & Fernandez JM (2000) *Neuron* 27, 435-446.

128. Li H, Linke WA, Oberhauser AF, Carrion-Vazquez M, Kerkvliet JG, Lu H, Marszalek PE, & Fernandez JM (2002) *Nature* 418, 998-1002.

129. Carrion-Vazquez M, Marszalek PE, Oberhauser AF, & Fernandez JM (1999) *Proceedings of the National Academy of Sciences of the United States of America* 96, 11288-11292.

130. Taniguchi Y, Brockwell DJ, & Kawakami M (2008) *Biophysical journal* 95, 5296-5305.

131. Li L, Huang HH, Badilla CL, & Fernandez JM (2005) *Journal of molecular biology* 345, 817-826.

132. Gao M, Craig D, Lequin O, Campbell ID, Vogel V, & Schulten K (2003) *Proceedings of the National Academy of Sciences of the United States of America* 100, 14784-14789.

133. Schwaiger I, Kardinal A, Schleicher M, Noegel AA, & Rief M (2004) *Nature structural & molecular biology* 11, 81-85.

134. Schwaiger I, Schleicher M, Noegel AA, & Rief M (2005) *EMBO reports* 6, 46-51.

135. Yew ZT, Krivov S, & Paci E (2008) *J Phys Chem B* 112, 16902-16907.

136. Baneyx G, Baugh L, & Vogel V (2002) *Proceedings of the National Academy of Sciences of the United States of America* 99, 5139-5143.

137. Gao M, Craig D, Vogel V, & Schulten K (2002) *Journal of molecular biology* 323, 939-950.

138. Ding F & Dokholyan NV (2005) *Trends in biotechnology* 23, 450-455.

139. Chen Y & Dokholyan NV (2006) *The Journal of biological chemistry* 281, 29148-29154.

140. Serohijos AW, Chen Y, Ding F, Elston TC, & Dokholyan NV (2006) *Proceedings of the National Academy of Sciences of the United States of America* 103, 18540-18545.

141. Sharma S & Dokholyan NV (2008) *Biophysical journal* 94, 1-3.

142. Sharma S, Ding F, & Dokholyan NV (2007) *Biophysical journal* 92, 1457-1470.

143. Serohijos AW, Hegedus T, Aleksandrov AA, He L, Cui L, Dokholyan NV, & Riordan JR (2008) *Proceedings of the National Academy of Sciences of the United States of America* 105, 3256-3261.

144. Hegedus T, Serohijos AW, Dokholyan NV, He L, & Riordan JR (2008) *Journal of molecular biology* 378, 1052-1063.

145.	Ferrer JM, Lee H, Chen J, Pelz B, Nakamura F, Kamm RD, & Lang MJ (2008) *Proceedings of the National Academy of Sciences of the United States of America* 105, 9221-9226.

146.	Evans E & Ritchie K (1997) *Biophysical journal* 72, 1541-1555.

147.	van der Flier A & Sonnenberg A (2001) *Biochimica et biophysica acta* 1538, 99-117.

148.	Travis MA, van der Flier A, Kammerer RA, Mould AP, Sonnenberg A, & Humphries MJ (2004) *FEBS letters* 569, 185-190.

149.	Ding F, Tsao D, Nie H, & Dokholyan NV (2008) *Structure* 16, 1010-1018.

150.	Yin S, Ding F, & Dokholyan NV (2007) *Nature methods* 4, 466-467.

151.	Yin S, Ding F, & Dokholyan NV (2007) *Structure* 15, 1567-1576.

152.	Franzot G, Sjoblom B, Gautel M, & Djinovic Carugo K (2005) *Journal of molecular biology* 348, 151-165.

153.	Soding J, Biegert A, & Lupas AN (2005) *Nucleic acids research* 33, W244-248.

154.	Luthy R, Bowie JU, & Eisenberg D (1992) *Nature* 356, 83-85.

155.	Popowicz GM, Muller R, Noegel AA, Schleicher M, Huber R, & Holak TA (2004) *Journal of molecular biology* 342, 1637-1646.

156.	Sjekloca L, Pudas R, Sjoblom B, Konarev P, Carugo O, Rybin V, Kiema TR, Svergun D, Ylanne J, & Djinovic Carugo K (2007) *Journal of molecular biology* 368, 1011-1023.

157.	Clamp M, Cuff J, Searle SM, & Barton GJ (2004) *Bioinformatics* 20, 426-427.

158.	Galtier N, Gouy M, & Gautier C (1996) *Comput Appl Biosci* 12, 543-548.

159.	Huelsenbeck JP & Ronquist F (2001) *Bioinformatics* 17, 754-755.

160.   Perriere GaG, M (1996) *Biochimie* 78, 364-369.

161.   Tyler JM, Anderson JM, & Branton D (1980) *The Journal of cell biology* 85, 489-495.

162.   Holland LZ (2007) *Nature* 447, 153-155.

163.   Davidson B (2007) *Seminars in cell & developmental biology* 18, 16-26.

164.   Khalturin K, Panzer Z, Cooper MD, & Bosch TC (2004) *Molecular immunology* 41, 1077-1087.

165.   Wakoh T, Ikeda M, Uchino R, Azumi K, Nonaka M, Kohara Y, Metoki H, Satou Y, Satoh N, & Satake M (2004) *DNA Res* 11, 345-352.

166.   Hartwig J (2007) in *Platelets*, ed. Michelson AD (Elsevier, Amsterdam), pp. 75-98.

167.   Oren M, Escande ML, Paz G, Fishelson Z, & Rinkevich B (2008) *PLoS ONE* 3, e3123.

168.   Lichtarge O, Bourne HR, & Cohen FE (1996) *Journal of molecular biology* 257, 342-358.

169.   Sawyer GM, Clark AR, Robertson SP, & Sutherland-Smith AJ (2009) *Journal of molecular biology*.

170.   Cranmer SL, Pikovski I, Mangin P, Thompson PE, Domagala T, Frazzetto M, Salem HH, & Jackson SP (2005) *The Biochemical journal* 387, 849-858.

171.   Chakarova C, Wehnert MS, Uhl K, Sakthivel S, Vosberg HP, van der Ven PF, & Furst DO (2000) *Human genetics* 107, 597-611.

172.   Maestrini E, Patrosso C, Mancini M, Rivella S, Rocchi M, Repetto M, Villa A, Frattini A, Zoppe M, Vezzoni P, *et al.* (1993) *Hum Mol Genet* 2, 761-766.

173.   Full SJ, Deinzer ML, Ho PS, & Greenwood JA (2007) *Protein Sci* 16, 2597-2604.

174. Furuhashi K, Inagaki M, Hatano S, Fukami K, & Takenawa T (1992) *Biochemical and biophysical research communications* 184, 1261-1265.

175. Nakamura F, Heikkinen O, Pentikainen OT, Osborn TM, Kasza KE, Weitz DA, Kupiainen O, Permi P, Kilpelainen I, Ylanne J, *et al.* (2009) *PLoS ONE* 4, e4928.

176. Kesner BA, Ding F, Temple BR, & Dokholyan NV (2009) *Proteins*.

177. D'Addario M, Arora PD, Ellen RP, & McCulloch CA (2003) *The Journal of biological chemistry* 278, 53090-53097.

178. Christodoulides N, Feng S, Resendiz JC, Berndt MC, & Kroll MH (2001) *Thrombosis research* 102, 133-142.

179. O'Connell MP, Fiori JL, Baugher KM, Indig FE, French AD, Camilli TC, Frank BP, Earley R, Hoek KS, Hasskamp JH, *et al.* (2009) *The Journal of investigative dermatology*.

180. Jiang P & Campbell ID (2008) *Biochemistry* 47, 11055-11061.

181. Pudas R, Kiema TR, Butler PJ, Stewart M, & Ylanne J (2005) *Structure* 13, 111-119.

182. Baker D & Sali A (2001) *Science (New York, N.Y* 294, 93-96.

183. Zhou Y & Karplus M (1997) *Proceedings of the National Academy of Sciences of the United States of America* 94, 14429-14432.

184. Lazaridis T & Karplus M (1999) *Journal of molecular biology* 288, 477-487.

185. Lazaridis T & Karplus M (1999) *Proteins* 35, 133-152.

186. Helen M. Berman , Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov lN, & Bourne PE (2000) *Nucleic acids research* 28, 235-242.

187. Case DA, Darden TA, Cheatham TE, III CLS, Wang J, Duke RE, Luo R, Crowley M, Walker RC, Zhang W, *et al.* (2008) (University of California, San Francisco).

188.   Ding F & Dokholyan NV (2006) *PLoS computational biology* 2, e85.

189.   Andersen HC (1980) *The Journal of chemical physics* 72, 2384-2393.

190.   Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, & Ideker T (2003) *Genome Res* 13, 2498-2504.

191.   Lu H, Krammer A, Isralewitz B, Vogel V, & Schulten K (2000) *Advances in experimental medicine and biology* 481, 143-160; discussion 161-142.

192.   Paci E & Karplus M (2000) *Proceedings of the National Academy of Sciences of the United States of America* 97, 6521-6526.

193.   Bendix PM, Koenderink GH, Cuvelier D, Dogic Z, Koeleman BN, Brieher WM, Field CM, Mahadevan L, & Weitz DA (2008) *Biophysical journal* 94, 3126-3136.

194.   Kumar S & Weaver VM (2009) *Cancer Metastasis Rev* 28, 113-127.

195.   Raman EP, Barsegov V, & Klimov DK (2007) *Proteins* 67, 795-810.

196.   Jimenez-Baranda S, Gomez-Mouton C, Rojas A, Martinez-Prats L, Mira E, Ana Lacalle R, Valencia A, Dimitrov DS, Viola A, Delgado R*, et al.* (2007) *Nature cell biology* 9, 838-846.

197.   Oberhauser AF, Badilla-Fernandez C, Carrion-Vazquez M, & Fernandez JM (2002) *Journal of molecular biology* 319, 433-447.

198.   Kyndt F, Gueffet JP, Probst V, Jaafar P, Legendre A, Le Bouffant F, Toquet C, Roy E, McGregor L, Lynch SA*, et al.* (2007) *Circulation* 115, 40-49.

199.   Zhang W, Han SW, McKeel DW, Goate A, & Wu JY (1998) *J Neurosci* 18, 914-922.