The Structural and Functional Characterization of the Periplakin Linker Domain

by

Graeden B. Winkelaar

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department of Biochemistry
University of Alberta

Abstract:

Desmosomes are large, complex protein assemblies which, at the most fundamental level, facilitate cellular adhesion. On a grander scale, desmosomes, and more specifically desmosomal networks, are critical for the maintenance of tissue integrity. They link the cytoskeletons of adjacent cells through a vast array of dynamic, multivalent interactions which ultimately equip tissues with the ability to withstand mechanical stress. The interactions of desmosomal proteins between cells in the extracellular space are quite well characterized. On the intracellular space, the story becomes much more complex as a plethora protein families and their constitutive members assemble in a tissue specific, multivalent fashion to create a mature desmosome.

The plakin proteins facilitate the final connection between the desmosome and the cytoskeleton of each cell, thereby establishing the network of cytoskeletal connectivity throughout a given tissue. This connections between the desmosome and the cytoskeleton have been historically mapped to C-terminal plakin repeat domains (PRDs) of the plakin proteins. However, each member of the plakin family has a linker domain while not all plakins have PRDs. This implies some distinct functionality of these linker domains. Moreover, the linker domain of periplakin has been identified as a site for phosphorylation, which has been previously linked to the dynamic assembly and disassembly of desmosomes.

This work seeks to elucidate the functional role of the periplakin linker domain by combining structural nuclear magnetic resonance spectroscopy (NMR) and small-angle x-ray scattering (SAXS) analysis, with quantitative binding analysis using MicroScale Thermophoresis (MST). By performing phosphomimic mutagenesis, we also gain insight into the effects of the aforementioned phosphorylation on the function of this domain.

Acknowledgements:

Table of Contents:

List of Figures:

List of Tables:

List of Abbreviations:

Amp	Ampicillin
B_1	Vitamin B_1: thiamine
CV	Column volume
DNA	Deoxyribonucleic Acid
DP	Desmoplakin
DSC	Desmocollin
DSG	Desmoglein
E. coli	*Escherischia coli*
FPLC	Fast protein liquid chromatography
GST	Glutathione S-Transferase
HEPES	(4-(2-hydoxythyl)-1-peperazineethanesulphonic acid)
HSQC	Heteronuclear Single Quantum Coherence
IF	Intermediate Filament
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
kDa	Kilodalton
LB broth	Luria Bertani broth
LD	Linker Domain
MST	MicroScale Thermophoresis
MW	Molecular Weight
OD	Optical Density
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PG	Plakoglobin
PKP	Plakophilin
PPL	Periplakin

PR	Plakin Repeat
PRD	Plakin Repeat Domain
PTS	Protein Thermal Shift
SDS	Sodium Docecyl Sulfate
SEC	Size Exclusion Chromatography
SOC	Super Optimal broth with Catabolite repression
SR	Spectrin Repeat
SAXS	Small Angle X-ray Scattering
SPR	Surface Plasmon Resonance
TB	Terrific Broth
TCEP	Tris(2-carboxyethyl)phosphine
T$_m$	Melting temperature
Tris	2-Amino-2-(hydroxymethyl)propane-1,3-diol
WT	Wild Type

CHAPTER 1

INTRODUCTION

Chapter One: Introduction

1.1 Cellular Adhesion and Junctions

Multicellular organisms rely on the cooperativity of specialized tissue types to carry out the required mechanisms for life. On a large scale, different organs execute a multitude of specific functions and work synergistically with other organs. Each of these individual organs is comprised of sets of specialized tissue types. At a more fundamental level, these tissues are composed of differentiated cell types which adhere to each other to form the tissue itself. For these tissues to perform their functions in an efficient capacity, the cell network that constitutes them must be linked in a robust manner. Cell-cell and cell-extracellular matrix adhesion is achieved by several types of specialized cell junctions which have a particularly high level of expression in connective tissue, cardiomyocytes, and epithelial tissues including the linings of the gastrointestinal tract, uterus, bladder, dermis, and epidermis[1]. These adhesive junction types can be divided into three distinct categories: communicating junctions, occluding junctions, and anchoring junctions[1]. While communicating and occluding junctions exist between cells of similar tissue type, their functionality relies on the establishment and maintenance of those tissue. These tasks fall largely on desmosomal anchoring junctions, especially in epithelial tissues.

1.1.1 Communicating junctions

Communicating junctions, such as gap junctions, are critical for metabolic and signaling synchronicity and coupling throughout a tissue. This is achieved through the formation of protein tunnels between cells. These tunnels are comprised of proteins called connexins. Various isoforms of connexins form homo- and hetero-hexamers in the membranes of cells termed connexons. The connexons of two adjacent cells in turn come together to form functional homo- and heterotypic

channels through which metabolites, ions, and solutes up to ~1 kDa in size can diffuse freely from cell to cell (Figure 1.1). These channels are dynamically opened and closed via phosphorylation and pH gating mechanisms to suit the needs of the cell at any given time[2].

Connexin   Connexon

| Connexon: | Homomeric | Heteromeric | Homomeric | Heteromeric |
| Gap Junction: | Homotypic | Homotypic | Heterotypic | Heterotypic |
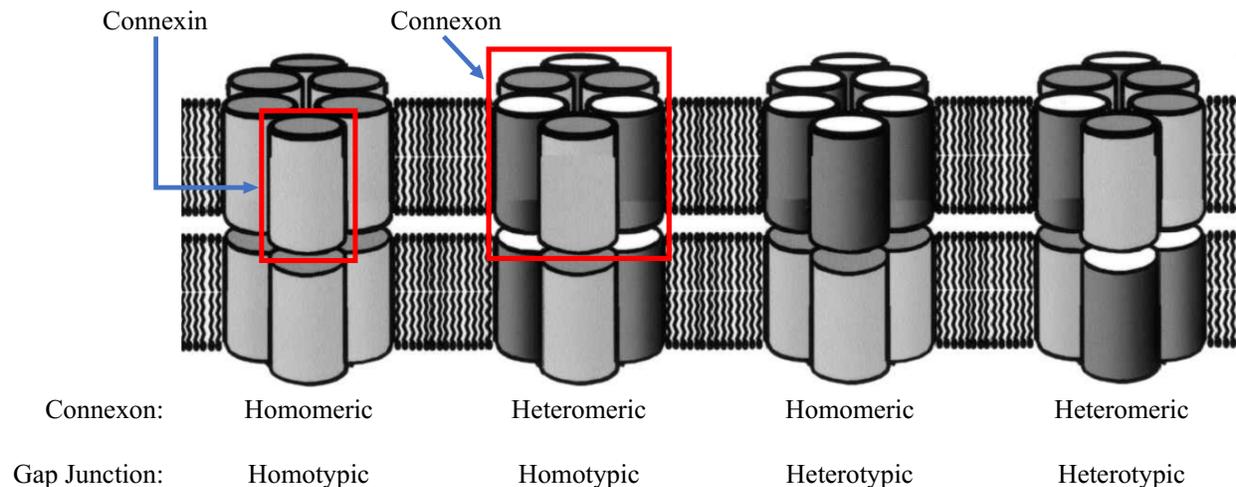
Figure 1.1 Schematic architecture of gap junctions.
A schematic diagram illustrating the possible arrangements of connexins and connexons used to form functional gap junctions. Adapted from Evans & Martin, 2002.

1.1.2 Occluding junctions

The quintessential occluding junction is the tight junction. Tight junctions are comprised of diverse families of proteins that act as rivets between the plasma membranes of adjacent cells (Figure 1.2). They are critical for maintaining tissue and cell polarity by precluding the diffusion of solutes between cells and the diffusion of membrane proteins from one area of the surface of a cell to another[3]. Instead, solutes must rely on cellular transport proteins for their uptake and movement in a regulated fashion. This allows cells of the gastrointestinal epithelium to establish apical and basolateral surfaces for specific nutrient absorption and delivery into the bloodstream, for example. However, tight junctions may also participate in complex gating mechanisms to preferentially allow paracellular diffusion of specific solutes as well and provide signaling information to other components of the cell[4].
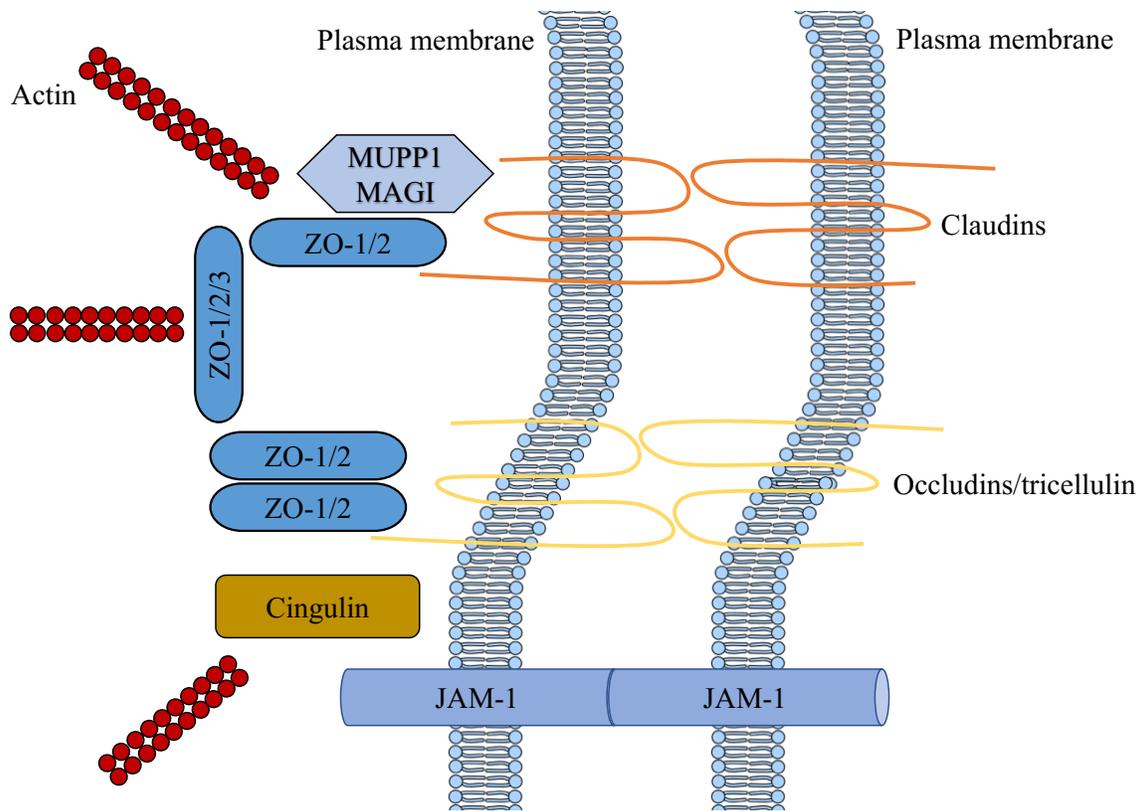
Figure 1.2 Schematic of a typical tight junction.
The schematic diagram reveals the diversity in assembly of tight junctions as well as several of the scaffolding proteins involved in joining the junctions themselves to the actin cytoskeleton of adjacent cells. Adapted from Niessen, 2007.

1.1.3 Anchoring junctions

Anchoring junctions include adherens junctions, hemidesmosomes, and desmosomes. These junctions share similar architecture to that of occluding junctions in that they consist of extracellular proteins which contain C-terminal transmembrane domains. In turn, various protein assemblies create dynamic attachment sites for intermediate filaments of the cytoskeleton intracellularly. Together, the attachment of intermediate filaments between cells of a tissue creates a robust network capable of dealing with substantial mechanical stress. The tissue integrity established by anchoring junctions is critical for the functioning of healthy tissue. Adherens junctions and desmosomes overlap considerably in their functionality: they both facilitate adhesion

between identical cell types by linking adjacent cytoskeletons. However, they differ substantially in their protein composition, cytoskeletal target, and their position in the intercellular cleft; typically, adherens junctions are found closer to the apical surface, whereas desmosomes are found proximal to the basolateral surface of cells. Hemidesmosomes typically facilitate cytoskeletal attachment of differing cell types and, more often, attachment of cells to the extracellular matrix[5]. Each of these anchoring junctions are highly dynamic complexes and experience varying degrees of turnover which plays a role in wound healing[6], cellular differentiation and division[7], and maintenance of tissue integrity.

1.2 Desmosomal Architecture

Desmosomes link the cytoskeletons of adjacent cells through a complex, dynamic, and diverse assembly of proteins. In a broad sense, the desmosome can be categorized into several protein families: cadherins, armadillo proteins, plakins, and intermediate filaments[8]. Each of these protein families work cooperatively to establish a robust network of connectivity throughout a tissue, enabling the management of mechanical stress. The assembly and regulation of these proteins into functional, mature desmosomes is a diverse, non-sequential process involving a vast network of accessory kinases, phosphatases, and proteases[9]. Given the critical role of these protein complexes in multicellular organisms, it is unsurprising that loss in functionality of any combination of desmosomal components leads to diseases of varying consequence[10].
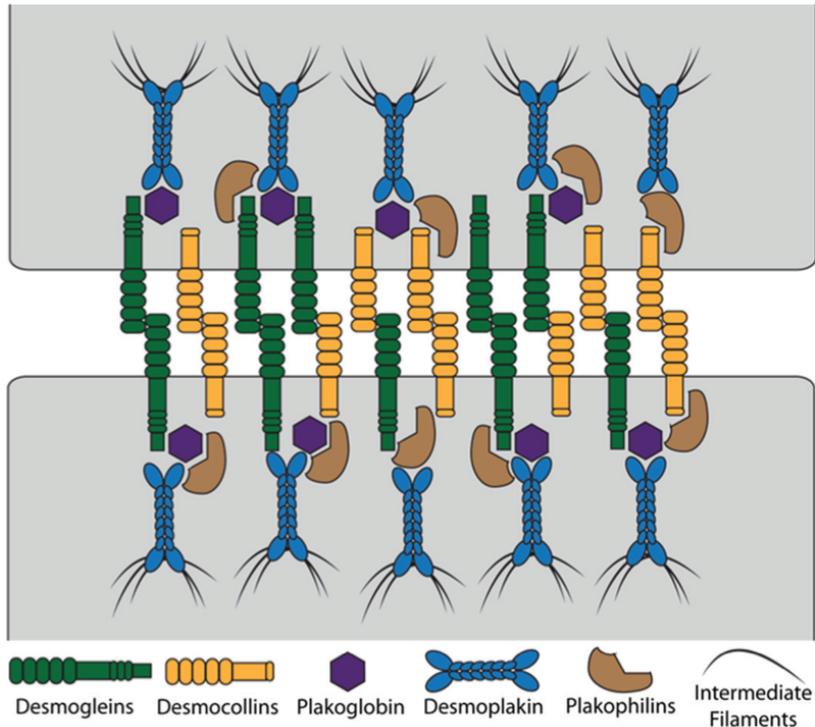
Figure 1.3 Schematic diagram of a desmosome.
A schematic showing the overall architecture of a mature desmosome. The function of each component is broadly displayed. The cadherins, desmogleins and desmocollins, facilitate cell-cell contact. The armadillo plakoglobin and plakophilins make contact with the C-termini of the cadherins and to the N-termini of the plakin proteins. The plakins may bind directly to the cadherins, or via an armadillo adapter. Plakins mediate the final interaction of the desmosome to the cytoskeleton. Adapted from (Brooke et al., 2012).

1.2.1 The cadherin protein family

Desmosomal cadherins include desmogleins (Dsg) 1-4 and desmocollins (Dsc) 1-3. The N-terminal regions of these proteins reside in the extracellular space and contain four repeating extracellular domains (ECs). It is these EC domains of the desmogleins and desmocollins which facilitate the physical attachment of one cell to the next through Ca^{2+}-dependent homo- and hetero-dimerization interactions (Figure 1.3)[11,12]. C-terminal to the EC is the extracellular anchor (EA) domain, sometimes termed the fifth EC domain. The EA domain resides directly adjacent to the

extracellular surface of the plasma membrane[13]. This domain is followed directly by a single, type I transmembrane domain which leads to the cytosolic C-terminal tails of the cadherins. The C-terminal tails of the cadherins facilitate attachment to the intracellular armadillo[14] and plakin[15] proteins of the desmosome.

1.2.2 The armadillo protein family

The armadillo family of desmosomal proteins includes plakophilins (PKP) 1-3 and plakoglobin (PG), also called γ-catenin. The plakophilins consist of globular N-terminal domains followed by 9 three-helix bundles termed armadillo repeats of 42 amino acids each[8]. The N- and C-terminal domains of plakoglobin are of unknown structure[16]. The central region consists 12 armadillo repeats which form a positively charged groove 95Å in length[17]. The groove created by the extended armadillo repeats facilitates adhesion to the cadherins of the desmosome[17]. Armadillo proteins are critical for the assembly and stability of mature desmosomes as they act largely as adapter proteins and provide a means of connection between the cadherins and the plakins[15].

1.2.3 The plakin protein family

The plakin family consists of bullous pemphigoid antigen-1 (BPAG1), microtubule-actin crosslinking factor-1 (MACF1), epiplakin, plectin, desmoplakin, envoplakin, and periplakin. Of these, only desmoplakin I and II, envoplakin, and periplakin are found to associate with desmosomes[10,18]. These proteins are comprised of N-terminal plakin domains followed by central filamentous rod domains and finally the C-terminal tails, which are quite divergent (Figure 1.4).

The N-terminal plakin domains are composed of left-handed antiparallel triple-helix motifs termed spectrin repeats (SRs) which follow a heptad repeat structure[19]. The heptad repeat allows for stabilization of the three-helix bundle as well as packing with adjacent SRs. Spectrin repeats

have been shown to mediate binding interactions with both armadillo proteins and the cytosolic C-terminal tails of cadherins[20,21,22]. The central rod domains of these proteins provides flexibility and enables homo- and heterodimer formation with the rod domains of other plakin proteins[23]. Together, this series of interactions mediates formation of the outer dense plaque of desmosomes[24] (Figure 1.5).
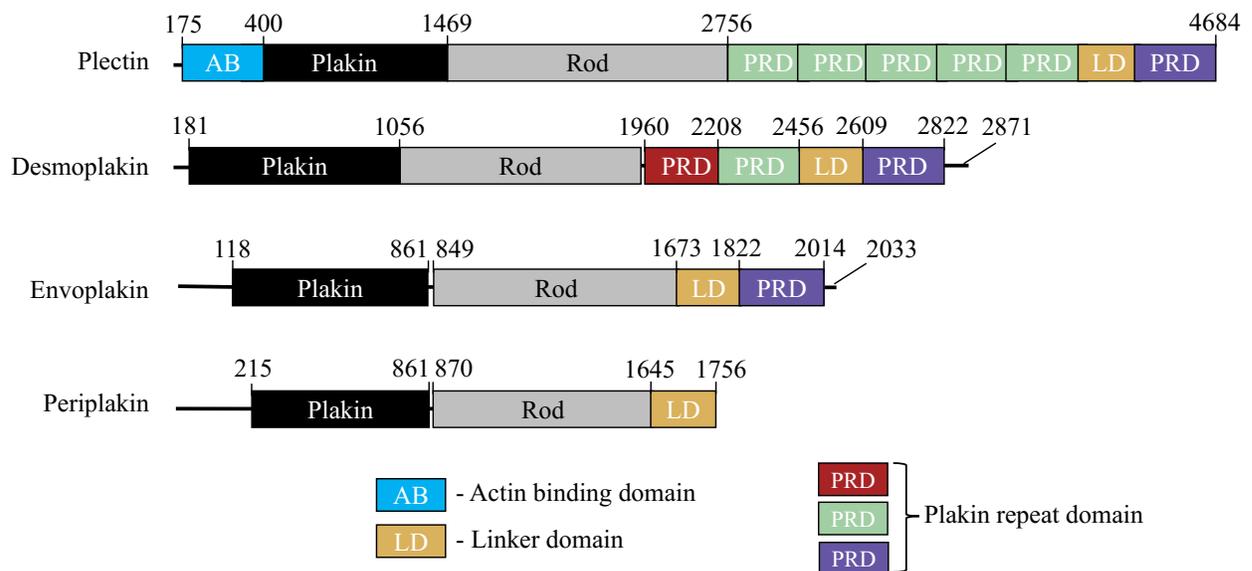
Figure 1.4 Schematic diagram of several members of the plakin protein family.
The plakins share conserved plakin domains and central rod domains. Their architecture differs in the C-terminal regions with the presence and frequency of PRDs. As a result, the plakins differ substantially in their length. Periplakin contains no PRDs. Adapted from (Bouameur et al., 2014; Fogl et al., 2016; Künzli et al., 2016).

The C-terminal regions of plakin proteins diverge in the presence, position, and frequency of Plakin Repeat Domains (PRDs) and Linker Domains (LD), which have been shown to bind to intermediate filaments of the cytoskeleton such as keratin and vimentin[25–29]. This interaction between the plakins and IFs of the cytoskeleton create the inner dense plaque (IDP) as seen in electron micrographs (Figure 1.5). It is unclear as to exactly why the members of the plakin family differ so greatly in their PRD arrangement, although it is thought to have an impact on the avidity of these C-terminal regions for IFs[27].

As with the rest of the desmosomal components, phosphorylation of plakin proteins has been established as a common means of desmosomal regulation. In desmoplakin, the phosphorylation of serine residues in the C-terminal region following PRD-C has been shown to disrupt desmosomal assembly by modulating the ability of the protein to bind to keratin intermediate filaments[30]. Desmoplakin and periplakin have also been found to be substrates of specific kinases such as protein kinase A (PKA) and AKT1/PKB for functional modulation and signaling purposes[9]. Mass spectrometry has also identified phosphorylation of Tyr/Ser/Thr residues throughout the plakin proteins which likely affect their interactions with other desmosomal proteins and signaling pathways necessary for dynamic desmosome assembly and disassembly[31,32].
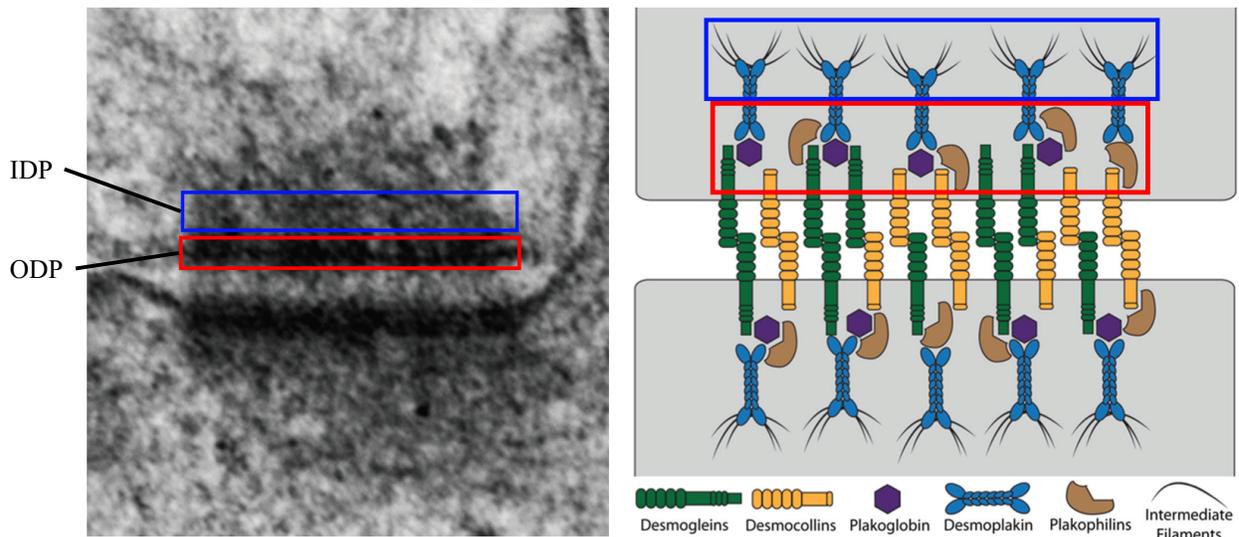
Figure 1.5 Electron micrograph of a desmosome and its corresponding schematic.
The electron micrograph of a single desmosome (left) and the schematic diagram (right) of the corresponding protein arrangement. Outlined in red is the outer dense plaque (ODP), which comprises the network of interactions between cadherins, armadillo proteins, and plakins. Outlined in blue is the inner dense plaque (IDP) which assembles from the interaction of IFs with the C-terminal regions of plakin proteins. Adapted from (Brooke et al., 2012).
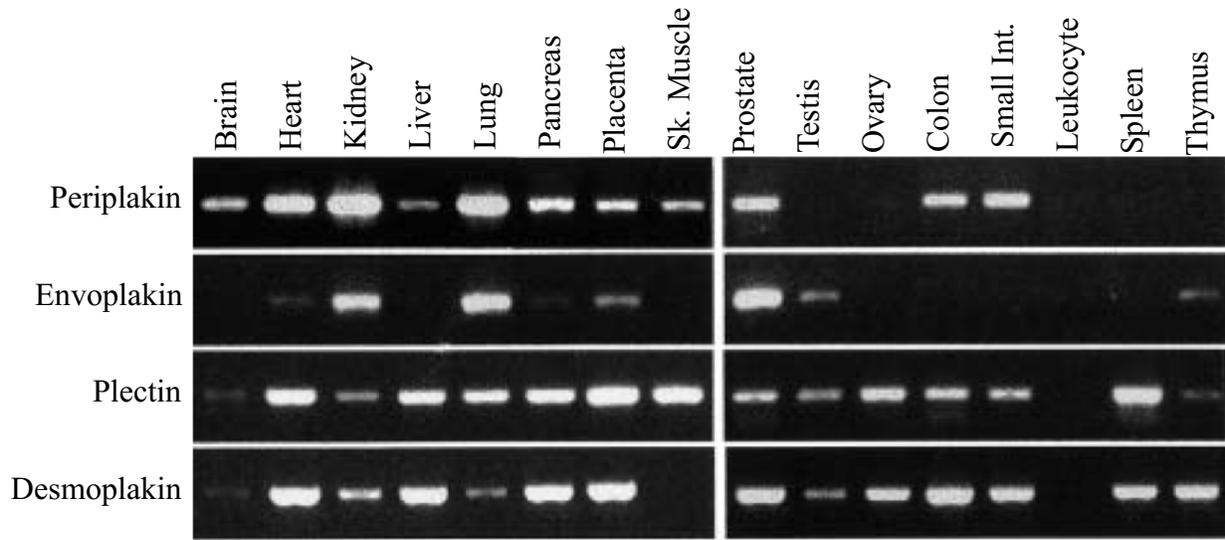
Figure 1.6 cDNA library showing the expression pattern of plakin proteins.
Agarose gel showing the results of a cDNA panel testing for plakin protein expression. Desmoplakin is quite ubiquitously expressed along with plectin. Periplakin and envoplakin show a more specific tissue expression pattern, suggesting some functional preference for those proteins in the desmosomes of given tissues. Adapted from (Kazerounian, Uitto, & Aho, 2002).

1.2.5 The plakin repeat domains (PRDs)

The plakin repeat domains of the plakin proteins can be characterized into PRD-A, PRD-B, and PRD-C. The difference between these three types of PRDs is mainly their primary sequence and resultant charge in their binding grooves[28]. All PRDs characterized to date have been found to adopt a globular conformation and comprise 4½ so-called 'plakin repeats' (PRs). The first four plakin repeats consist of approximately 38 amino acids which form a conserved 11 residue β-hairpin followed by a series of two antiparallel α-helices, H1 and H2, 8 amino acids and 14 amino acids in length, respectively. The last plakin repeat contains only the first 20 amino acids of a regular plakin repeat and lacks the entirety of H2. Individually, these plakin repeats are stabilized by ionic and hydrophobic interactions. As the plakin repeats fold together to form a PRD, the effects from the same stabilizing forces are relied upon, primarily through hydrophobic packing of

residues in the β-hairpin and H2 regions[26]. An interesting difference arises between PRDs-A and B and PRD-C of desmoplakin: PRDs-A and B have been found to have non-canonical PRs in their N-termini[33]. This region spans the first 50 residues of the PRD and resembles a typical PR followed by a third α-helix (Figure 1.7). This non-canonical PR has been thought to be unique to PRDs A and B, as it is not present in the PRD-C of desmoplakin or the PRD-C of envoplakin. However, this unique, extended PR motif has since been observed outside the context of PRDs as discussed in following sections.

The PRD-B and PRD-C of desmoplakin and have been shown to be points of interaction with vimentin and keratin intermediate filaments of the cytoskeleton, likely through electrostatic complementarity between a conserved basic groove in the PRD and acidic residues on the IFs themselves[26–28]. Rigorous proof of this electrostatic mechanism has been shown using the PRD-C of envoplakin: conserved Arg and Lys residues in its basic groove are critical for its interaction with vimentin cytoskeletal filaments[28]. As members of the plakin family dimerize through their central rod domains, it is thought that the C-terminal PRDs act as multiple points of contact for intermediate filaments, establishing a dense network of connectivity and ultimately leading to the formation of the IDP.
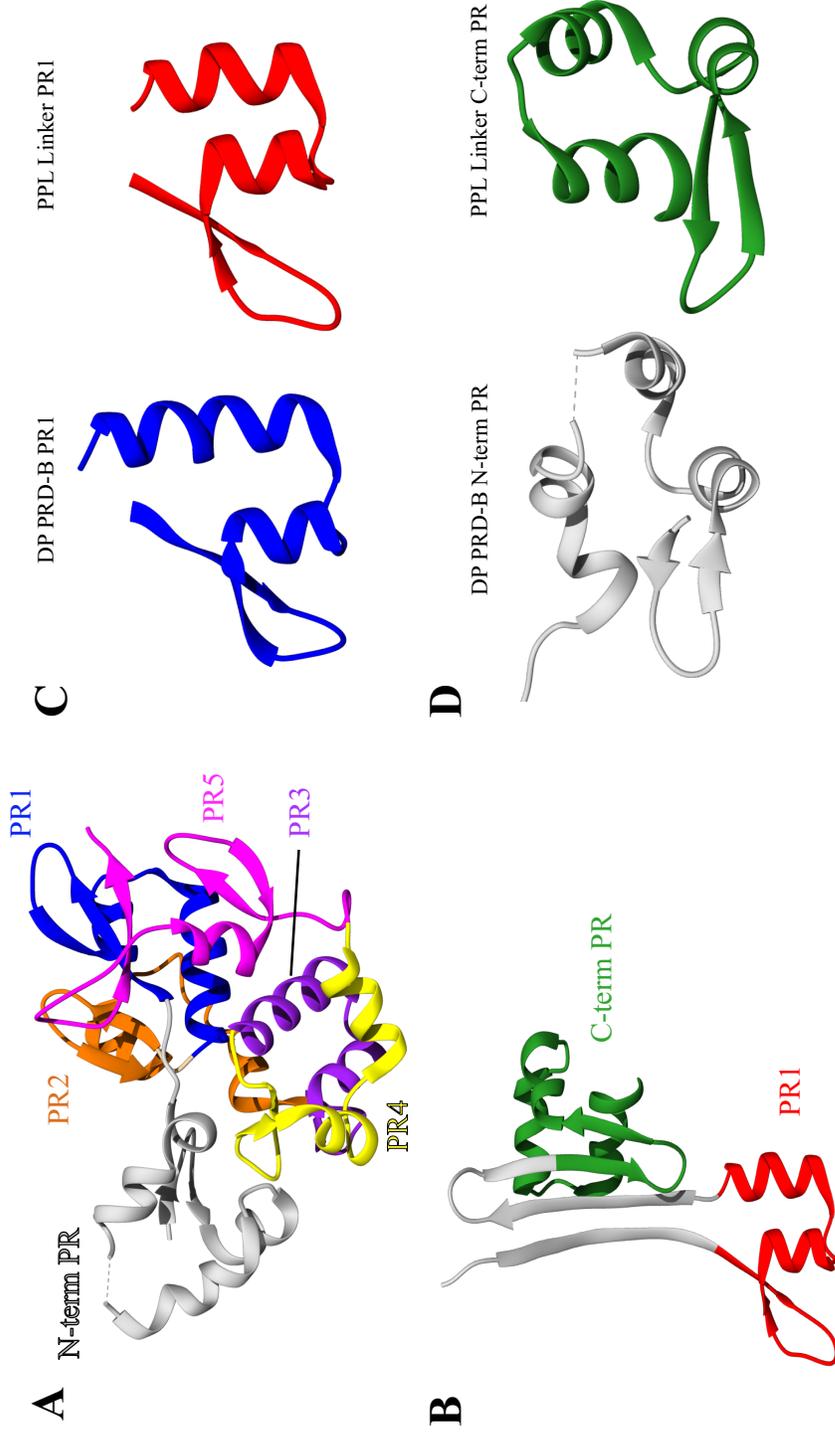
Figure 1.7 Comparison of the desmoplakin PRD-B and PPL₁₆₅₅₋₁₇₅₄ structures.
A) The X-ray crystal structure of the desmoplakin (DP) PRD-B (PDB 1LM7) at 3.0Å resolution. The non-canonical plakin repeat is designated in gray/white followed by each of the canonical 4½ PRs that compose typical PRDs. **B)** The existing crystal structure of the periplakin linker domain at 2.64Å resolution (PDB 4Q28). Like the PRD-B of DP, it contains a single PR (shown in red), and a non-canonical PR (shown in green). This structure lacks the first 10 native residues. Instead, there is an N-terminal 6xHis-tag which participates in an extended β-sheet, which likely compromises elements of the native structure. **C)** A comparison between the PR1 of DP and PR1 of the periplakin linker from the existing structure. Both show typical PR features: a β-hairpin followed by antiparallel α-helices. **D)** Comparison between the non-canonical PRs observed in the PRD-B of DP and the C-terminus of the PPL linker domain. Both exhibit typical PR β-hairpins and antiparallel α-helices, but also include an atypical third α-helix C-terminally.

1.2.4 The plakin linker domains (LDs)

As seen in Figure 1.4, the linker domains of plakin proteins are ubiquitous, implying that they impart some functionality that is distinct from the other domains of the plakin proteins. Historically, the linker domains have been identified simply as "flexible linkers" between PRDs. However, the linker domains themselves likely serve some function other than only joining adjacent PRDs. The sequence of these linker domains is highly conserved among the plakin family as well as through phylogenetic space. Furthermore, the linker domain of periplakin has been observed to interact with both keratin and vimentin filaments in a yeast two-hybrid assay[29], although the details of this interaction remain unclear. The structure of periplakin linker domain has been crystalized, although there are some obvious issues with the current structure which are discussed in detail in section 1.4.

1.3 Intermediate Filaments of the Cytoskeleton

Cytoskeletal intermediate filaments (IFs) are composed of long, rod like proteins that assemble into oligomers to form a scaffold inside the cell. This scaffold helps the cell maintain its structural rigidity and deal with mechanical stress. Typical intermediate filaments include keratin and vimentin, both of which are alpha helical and have been found to adhere to the C-terminal regions of plakin proteins in mature desmosomes[28]. IF sequences also tend to consist of heptad repeats. These repeating sequences facilitate dimerization between individual filaments while exposing patches of charged residues which offer binding functionality.

1.3.1 Keratin

Keratins have been classically divided into Class I/Type A (acidic) and Class II/Type B (basic/neutral)[34]. However, a new nomenclature for keratins has emerged and the current number

of identified human keratins has climbed to fifty-four. Three categories now exist for keratins: epithelial keratins, hair keratins, and psuedokeratins. The epithelial keratins form the cytoskeletons of epithelial cells. Twenty of the fifty-four identified keratins fall into this category[35]. The expression patterns for the specific types and combinations of cytoskeletal keratin is largely tissue and organ specific, implying some sort of functional diversity among the family[36]. Oligomerization typically occurs between pairs of acidic and basic filaments which form stable, electrostatically and hydrophobically zippered dimers which then pair themselves with additional dimers in an antiparallel fashion to form elongated filaments[35,37].

1.3.2 Vimentin

Vimentin is another member of the intermediate filament family. Similar to keratin filaments, it is predominantly α-helical and forms homo-tetrameric complexes as the basis for vimentin filament formation. It's core structure consists of a globular head, central helical coil regions (coils 1A, 1B, and 2) divided by short linker segments, and a globular tail. Around 90% of the central coiled domain structure has been elucidated by x-ray crystallography[38–42] and it is this region of vimentin that facilitates the dimerization and tetramerization necessary for filament assembly and polymerization[43].

These vimentin filaments form a vast framework throughout the cytosol and have important implications for cytosolic processes such as cellular proliferation, migration, organelle positional maintenance, and overall cellular rigidity[44]. As widely contributing elements of cellular structure and function, vimentin filaments require dynamic assembly and disassembly, which are regulated largely through site-specific phosphorylation by various kinases such as Cdk1, PKA, and PKC[38,43,45].

Vimentin, similarly to keratin, has been identified as an intermediate filament which forms assemblies with plakin proteins to establish the IDP of desmosomes[46]. These interactions rely heavily on the alpha-helical nature of these filaments and their exposed acidic residues which participate in electrostatic interactions with plakin PRDs[28]. Specific vimentin residues responsible for the interaction between vimentin at envoplakin's PRD-C have been studied and found to reside in the coil 1A region of vimentin[28]. However, the heptad repeat nature of vimentin creates a potential multivalency and variability in possible binding mechanisms, positions, and critical residues necessary for interaction with plakin proteins.

1.4 The Periplakin Linker Domain

An x-ray crystal structure of the periplakin linker domain has been elucidated by John Hunt's structural genomics group (Figure 1.8)[47]. The structure reveals β-hairpin and anti-parallel α-helix structures like those seen in the plakin repeats found in the PRDs of other plakin proteins. More specifically, 10 non-native residues, including a 6xHis-tag, form the N-terminus of the structure. This is followed by a non-canonical plakin repeat, identical to those seen in typical PRDs. A short β-strand then leads to a C-terminal non-canonical PR, structurally similar to those seen in PRD-A and PRD-B of desmoplakin. It is likely that the large β-strand conformation adopted by the linker domain is an artifact of the first 10 non-native residues. As such, this structure is in need of physiologically relevant elucidation. However, the two PRs which are visible is indicates that this structure is still valuable as a starting point to characterize the function of this domain.
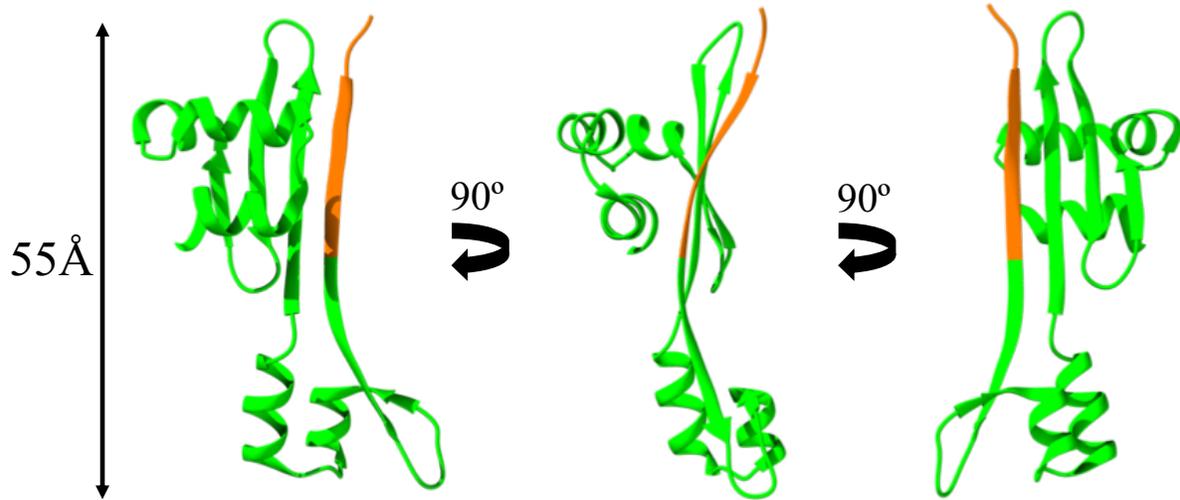
Figure 1.8 X-ray crystal structure of the periplakin linker domain.
The crystal structure of $PPL_{1655-1754}$ to 2.64Å reaching a maximal dimension of roughly 55Å. Highlighted in orange are 10 non-native residues, six of which are a His-tag used for purification purposes. These 10 non-native residues participate in part of an artifactual β-sheet. In green is the native sequence of linker domain. However, this crystal structure is incomplete as it is missing the first 10 native N-terminal residues along with the last two C-terminal residues. There are, however, striking similarities between this structure and that of other plakin PRDs. (PDB 4Q28)

The periplakin linker domain has also been identified as a target for phosphorylation at serine 1657[48]. Mass spectrometry has isolated phosphorylated periplakin linker domain in breast cancer tissue samples, lung cancer and in differentiating cells[49], though the exact significance of this particular phosphorylation remains unclear[50–52] in each of the identified cases. Although the actual kinase responsible for catalyzing this reaction has yet to be determined, likely candidates are PKA and PKB based on previously mapped interactions with periplakin, or MAPK since the phosphorylation site on periplakin (RR<u>S</u>IVV) shares marked similarity to consensus sequences targeted by MAPK as calculated by the Eukaryotic Linear Motif (ELM) resource[53]. Serine 1657 is located immediately before its first canonical PR and adjacent to a cluster of basic residues (Figure 1.9). Our collaborator, Fiyaz Mohammed, has run modeling analyses using the current PDB structures available for the periplakin linker domain and coil 1A and 1B regions of vimentin using

High Ambiguity Driven protein-protein DOCKing (HADDOCK) modeling[54,55]. These models predict the basic residues adjacent to S1657 to mediate interaction between the linker domain and vimentin intermediate filaments (Figure 1.10).
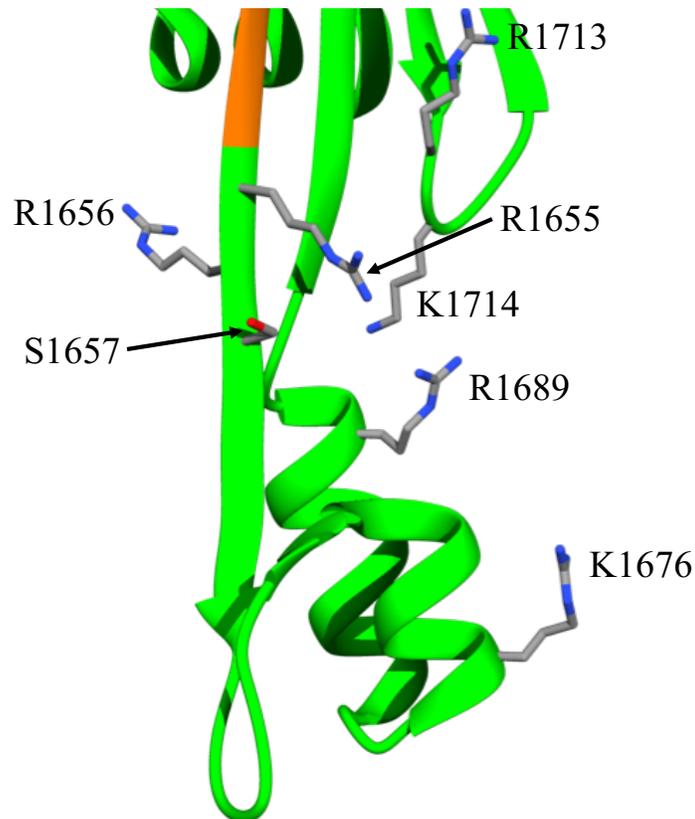
Figure 1.9 The phosphorylated serine of the periplakin linker domain.
The periplakin linker domain with critical residues highlighted. Serine 1657 has been found by mass spectrometry to have a high incidence of phosphorylation. This residue resides in close proximity to basic residues which may be involved in binding to intermediate filaments via electrostatic mechanisms. Phosphorylation at this serine would introduce a large negative charge which could shield the positive nature of the labeled lysine and arginine residues thereby affecting binding capabilities. (PDB 4Q28)
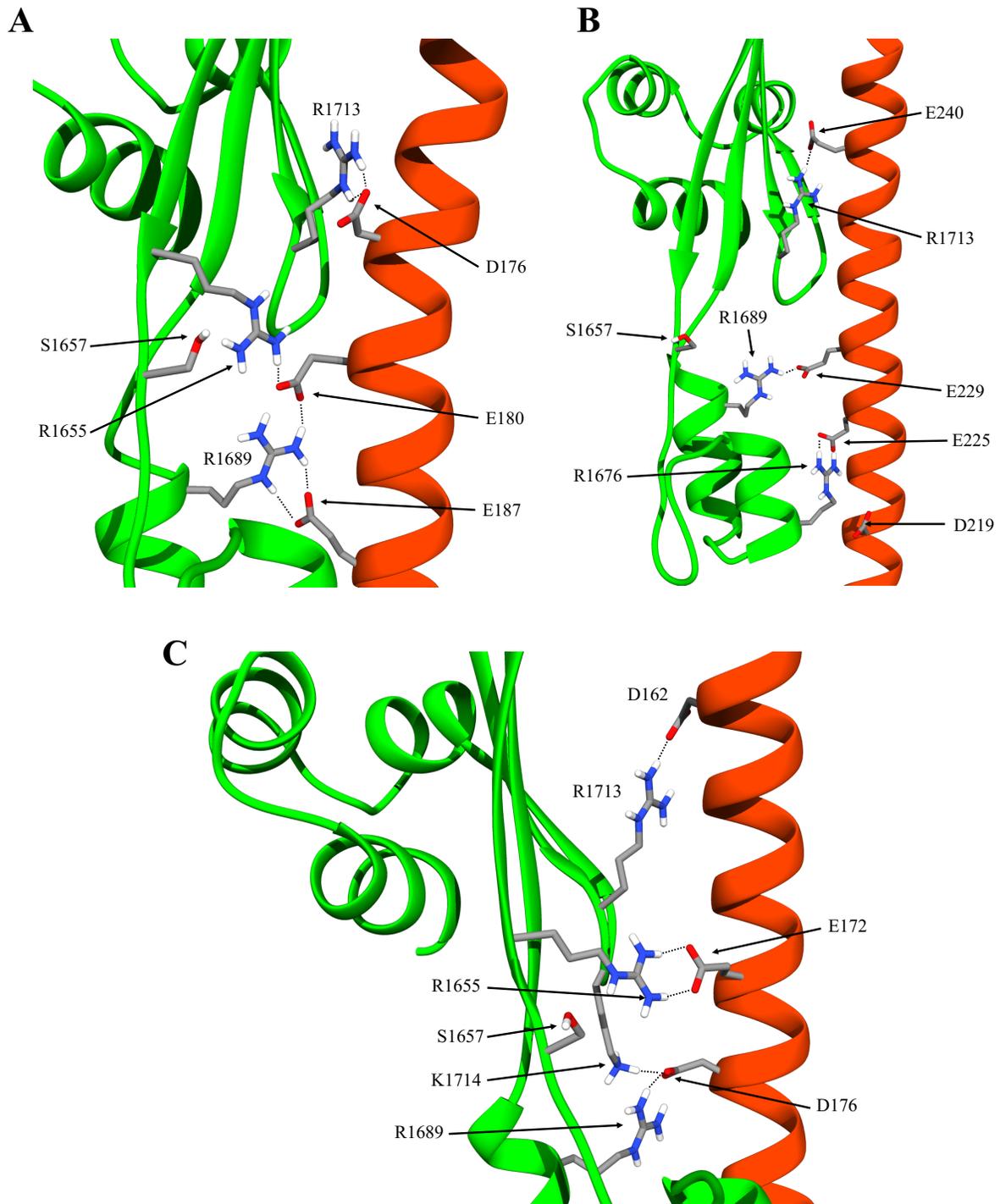
Figure 1.10 HADDOCK models of probable periplakin-vimentin binding interactions.
Three models for predicted binding interactions between vimentin (PDB 3UF1 and 3SWK) and the periplakin linker domain (PDB 4Q28) using the current available crystal structures. Basic residues on the linker domain make electrostatic contacts with acidic residues on vimentin. The highly phosphorylated serine 1657 is in close proximity to the predicted binding sites and would likely weaken the strength of these interactions. Models were constructed by Fiyaz Mohammed.

1.5 Thesis Aims and Hypotheses

The purpose of this work is to elucidate the structural and functional significance of periplakin's linker domain. Although these linkers are found to be ubiquitous, periplakin's is unique in the sense that it is the only member of the plakin family to have a C-terminus consisting solely of a linker domain. This implies some distinct and critical function for periplakin's linker domain over that of other plakin proteins. Study of the periplakin linker also provides a relatively simple system in which to study the role of phosphorylation in a desmosomal context.

Using the existing periplakin linker crystal structure as an initial reference point already provides some intriguing insights into its possible role in desmosomal dynamics. The presence of canonical and non-canonical PRs suggests that, like the PRDs of other plakin proteins, the linker domain of periplakin may also participate in binding interactions with intermediate filaments of the cytoskeleton. However, it is also clear that the structure of the linker domain requires refinement, given that the structure solved by John Hunt's group is missing the first 10 native residues and adopts a large β-sheet conformation unlike that seen in any related PRD structures. Given the current lack of knowledge in the area of the periplakin linker domain, the aims for this thesis are:

1. Establish of the function of the periplakin linker domain.

2. Determine the effect of phosphorylation on the function of the periplakin linker domain.

3. Use NMR to characterize the periplakin linker domain structure.

My hypotheses are:

1. The periplakin linker domain binds to intermediate filaments of the cytoskeleton.

2. Phosphorylation of S1657 will decrease the strength of interaction between the periplakin linker domain and intermediate filaments of the cytoskeleton.

CHAPTER 2

MATERIALS AND METHODS

Chapter Two: Materials and Methods

2.1 Molecular Biology and Gene Design

2.1.1 N-terminally extended periplakin linker domain (PPL$_{1626-1756}$)

Past attempts to characterize periplakin linker domain (K1646-K1756) structure and interactions have been only moderately successful, particularly in the case of NMR spectral quality and artifactual crystal contacts. For this reason, an N-terminally extended construct of the linker domain incorporating part of the adjacent rod domain was created to improve the folding behavior and structural integrity of the protein. This extension includes 20 amino acid residues of periplakin's coiled-coil rod domain and results in the construct S1626-K1756 (Figure 2.1A). These 20 amino acids were added to the linker domain using polymerase chain reaction (PCR)[56]. The PCR reaction was set up as outlined in Table 2.1. The forward primer encodes a BamHI restriction site followed by S1626-E1651 of periplakin. The reverse primer includes an EcoRI restriction site. The original linker domain construct (K1646-K1756) in a pGEX-6P-1 vector was used as the template. The gene is also flanked by BamHI and EcoRI restriction sites.

Table 2.1 Reaction mixture for PCR extension of PPL$_{1646-1756}$.

Reagent	Quantity
5X Phusion High Fidelity Buffer	40 μL
10 mM dNTPs	1 μL
10 μM forward periplakin N-terminal extension (Table 2.8)	10 μL
10 μM reverse periplakin N-terminal extension (Table 2.8)	10 μL
10 ng/μL template DNA	1 μL
Phusion Polymerase (2 U/μL)	2 μL
Nuclease free ddH$_2$O	130 μL
	200 μL total

The 200µL reaction mixture was mixed thoroughly by pipetting before being divided into three separate tubes to try three different annealing temperatures. The PCR protocol used is outlined in Table 2.2.

Table 2.2 PCR conditions for N-terminal extension of PPL$_{1646-1756}$.

Temperature (°C)	Time (min:sec)	Description
98	2:00	Initial denaturation
98	0:15	Denaturation
72, 74.3, 78	0:30	Annealing
72	0:30	Elongation
72	5:00	Final elongation
12	Infinite	Stable hold

The PCR products were then run on a 1% agarose gel at 150 V for 25 minutes. Bands were visualized using UV light, cut from the gel, and purified using the GenepHlow Gel/PCR kit (GeneAid) to final volume of 50 µL. The purified PPL$_{1626-1756}$ PCR product and template vector were then subjected to restriction digestion as outlined in Tables 2.3 and 2.4.

Table 2.3 Reaction mixture for restriction digestion of PPL$_{1626-1756}$ insert.

Reagent	Quantity
10X Fast Digest Green Buffer	6 µL
255 ng/µL PPL$_{1626-1756}$ purified PCR product	50 µL
FastDigest BamHI (Thermo Scientific)	3 µL
FastDigest EcoRI (Thermo Scientific)	3 µL
	62 µL total

Table 2.4 Reaction mixture for restriction digestion of pGEX-6P-1 vector.

Reagent	Quantity
10X FastDigest Green Buffer	4 µL
134 ng/µL pGEX-6P-1 vector	30 µL (4 µg)
FastDigest BamHI (Thermo Scientific)	3 µL
FastDigest EcoRI (Thermo Scientific)	3 µL
	40 µL total

Restriction digestions were performed at 37 ºC for 1.5 hours. Restriction enzymes were then deactivated by incubating the reaction mixtures at 80 ºC for 5 mins. Digested template was then isolated by running it on a 1% agarose gel at 150 V for 35 mins. The template band was then visualized under UV, cut from the gel, and purified using the GenepHlow Gel/PCR kit (GeneAid). The digested pGEX-6P-1 vector and $PPL_{1626-1756}$ were then ligated as outlined in Table 2.5 before being transfected into *E. coli* as described in section 2.2.

Table 2.5 Reaction mixture for ligation of $PPL_{1626-1756}$ into pGEX-6P-1.

Reagent	Quantity
5X ligase buffer	4 µL
$PPL_{1626-1756}$ PCR product	13 µL
Digested pGEX-6P-1 vector	2 µL
T4 Ligase (5 U/µL)	1 µL
	20 µL total

The mixture was incubated at room temperature for 30 mins before transformation into *E. coli*. A volume of 4 µL of ligated product was used for the transformation.

2.1.2 $PPL_{1626-1756}$ S1657E phosphomimic mutagenesis

As previously discussed in section 1.4, serine 1657 of periplakin's linker domain has been identified through mass spectrometry to have a high incidence of phosphorylation[48]. A common form of post translational modification, we hypothesize that this phosphorylation event has some modulating effect on the linker domain's ability to bind to intermediate filaments such as vimentin and keratin. Site-directed mutagenesis was performed using the Agilent QuikChange Lightning Site-Directed Mutagenesis Kit. Primers were ordered from Integrated DNA Technologies.

Table 2.6 Reaction mixture for phosphomimic mutagenesis.

Reagent	Quantity
10X reaction buffer	5 µL
dsDNA template (90 ng/µL)	1 µL
Forward primer S1657E (107 ng/µL) (Table	1.17 µL
Reverse primer S1657E (108 ng/µL) (Table 2.8)	1.16 µL
dNTP mix	1 µL
QuikSolution reagent	1.5 µL
ddH$_2$O	39.2 µL
QuikChange Lightning enzyme	1 µL
	51 µL total

Following the addition of 1 µL of QuikChange Lightning enzyme, the tube contents were then mixed by pipetting and split into three tubes of 17 µL each to try different annealing temperatures (55 ºC-75 ºC) for the primers during the PCR.

Table 2.7 PCR conditions for mutagenesis.

Temperature (ºC)	Time (min:sec)	Description
98	5:00	Initial denaturation
95	0:20	Denaturation
55, 62.5, 75	0:15	Annealing
68	2:42	Elongation
68	5:00	Final elongation
12	Infinite	Stable hold

The denaturation, annealing, and elongation steps were repeated 18 times. Upon completion of the PCR, methylated and hemi-methylated template DNA was digested by adding and mixing in 1 µL of DpnI restriction enzyme to each tube. The tubes were then briefly spun down before a 5 minute incubation at 37ºC for digestion of the methylated/hemi-methylated template DNA. Following template digestion, 2 µL of β-ME was added to 45 µL of thawed XL10-

Gold Ultracompetent Cells (Agilent), which were subsequently split into 3 aliquots of 15 µL. The digested DNA mixture (2 µL) was added to the aliquoted cells, swirled to mix, and incubated on ice for 30 minutes. The tubes were then heat shocked at 42 ºC in a water bath for 30 seconds and then incubated on ice again for 2 minutes. Then, 200 µL of 42 ºC preheated SOC media was added to each tube followed by incubation at 37 ºC for 1 hour. The mixture (150 µL) from each tube was then plated on agar plates containing 50 µg/mL ampicillin and incubated for 18 hours at 37 ºC. Cultures from the plates were then used for generation of DNA stocks as described in section 2.2. Purified DNA was then sent for Sanger Sequencing for confirmation of specific mutagenesis (sequencing primers are listed in Table 2.8).

2.1.3 Vimentin$_{99-249}$

The plasmid containing the gene for vimentin sequence T99-I249 was obtained from collaborators in Birmingham, UK whom had previously cloned the gene into the pET21b- vector, yielding a C-terminal 6xHis-tag (Figure 2.1). It should be noted that this vimentin construct does not contain any tryptophan residues, precluding detection via stain-free gel protocols and making concentration determination by UV absorbance unreliable. Both the pGEX-6P-1 and PET21b-vectors used contain the gene for ampicillin resistance (Amp^R).

Table 2.8 List of DNA primers.

Primer Sequence	Function
5' - A GAG GGA TCC TCC AAG GAC AAA GAC CTC GAG ATC GAC GAG CTG CAG AAG CGC CTG GGC TCC GTG GCC GTC AAG CGG GAG CAG CGG GAG - 3'	Forward periplakin N-terminal extension
5'- C CAT AGA CCC GTC TTC ATC ACT CTT AAG GAG AG -3'	Reverse periplakin N-terminal extension
5' - GAA CCA CCT GCG GCG CGA GAT CGT AGT CAT CCA CC - 3'	Forward S1657E mutagenesis
5' - GGT GGA TGA CTA CGA TCT CGC GCC GCA GGT GGT TC - 3'	Reverse S1657E mutagenesis
5'- GGC AAG CCA CGT TTG GTG - 3'	Forward pGEX-6P-1 sequencing primer
5' - GAG CTG CAT GTG TCA GAG G - 3'	Reverse pGEX-6P-1 sequencing primer
5' - TAA TAC GAC TCA CTA TAG GG - 3'	Forward T7 sequencing primer
5' - TAG TTA TTG CTC AGC GGT GG - 3'	Reverse T7 sequencing primer

Table 2.9 List of proteins, vectors and E. coli expression strains.

Protein Construct	Expression Vector	Expression Strain
PPL$_{1626-1754}$ (WT)	pGEX-6P-1	BL21(DE3)
PPL$_{1626-1754}$ (S1657E)	pGEX-6P-1	BL21(DE3)
Vimentin$_{99-249}$ (WT)	pET21b-	BL21(DE3)

A

PPL₁₆₂₆₋₁₇₅₆

$PPL_{1626-1756}$

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIR

YIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGD

HVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK

CS   1626   1645   1655   1676

SDLEVLFQGPLGSSSKDKDLEIDELQKRLGSVAVKREQRENHL**RRS**IVVIHPDTGRELSPEEAH**R**AGLIDWNMF – 1685

1689   1713 1714

1686 – VKL**RS**QECDWEEISVKGPNGESSVIHD**RK**SGKKFSIEEALQSGRLTPAQYDRYVNKDMSIQELAVLVSGQK

1756

B

Vimentin₉₉₋₂₄₉

$Vimentin_{99-249}$

99   162

M TRTNEKVELQELNDRFANYIDKVRFLEQQNKIL LAELEQLKGQGKSRLGDLYEE MRELRRQV**D**QLTNDKA – 169

172 176 180 187   225 229 240

170 – RVE VER**DN**LA**E**DIMRLR**E**KLQEEMLQREEAENTLQSFRQDVDNASLARLDLERK**V**ESLQEEIAFLKKLHEEEJ – 242

243 – QELQAQIHHHHHH

249

Figure 2.1 Primary structures of PPL₁₆₂₆₋₁₇₅₆ and vimentin₉₉₋₂₄₉ constructs.

A) The periplakin construct studied in this work as confirmed by Sanger sequencing. In purple is the GST affinity tag. The following green section is the 3C PreScission protease cleavage motif with the actual cleavage site (CS) denoted by the arrow. Greyed out residues correspond to non-native residues of periplakin. There is a total of 5 non-native residues in the final protein product. Boxed in red are 20 amino acids from the coiled coil central rod domain of periplakin. The linker domain begins at K1645. Bolded residues indicate those which have been predicted to bind vimentin in HADDOCK models. Serine 1657 is indicated as a phosphorylation site. This residue is mutated to glutamate to mimic phosphoserine. **B)** The vimentin construct used in this work. The blue box is coil 1A and the red box is coil 1B. Bolded residues are those which have been predicted to bind the periplakin linker domain based off of the current crystal structure. In green is the C-terminal 6xHis-tag used for purification and MST labeling.

2.2 Plasmid Transformation and Production

Transformation was performed by adding 1 µL of 80-90 ng/µL plasmid DNA to 50-60 µL of thawed *E. coli* DH5α competent cells (Invitrogen) for DNA stock production, or BL21(DE3) competent cells (New England Biolabs) for protein production. The tubes of cells were then gently flicked to mix, and then set to incubate for 30 minutes on ice. The cells were then heat shocked at 42 ºC for 30 seconds before being returned to ice for an additional minute. A volume of 300 µL of SOC media was then added to the cells and mixed by gently flicking the tube before incubation at 37 ºC for 45 minutes to allow expression of the Amp^R gene. Following this incubation, a 150 µL aliquot from the tubes was spread on agar plates containing 50 µg/mL ampicillin. The plates were then incubated for overnight at 37 ºC.

The generation of plasmid stocks was completed by picking single colonies from plates of transformed DH5α *E. coli* and transferring them to 100 mL of TB containing 100 µg/mL ampicillin. These cultures were allowed to grow at 37 ºC for 18-20 hours with shaking at 180 rpm. Plasmid DNA from these cultures was extracted and purified using the Geneaid™ Mini or Midi Plasmid Kits. Plasmid purity was checked using A_{260}/A_{280}, which gave values from 1.8-2.0, indicative of sufficiently pure DNA. The concentration of purified plasmid DNA was quantified using A_{260} measurements at 1:100 dilution in ddH_20. Plasmid yield ranged from 50 µL – 2 mL at 80-110 ng/µL.

All plasmid sequences were confirmed using Sanger Sequencing performed by The Applied Genomics Core (TAGC) at the University of Alberta. For constructs cloned into pGEX-6P-1, pGEX forward and reverse sequencing primers were used (Table 2.8). The vimentin construct was obtained from collaborator Dr. Martyn Chidgey (University of Birmingham, UK) and was sequenced using T7 promoter primers (Table. 2.8).

2.3 Protein Production

2.3.1 Production of unlabeled protein

Following transformation of BL21(DE3) *E. coli*, single colonies from agar plates were picked and transferred to 20 mL of LB+B_1 or TB+B_1 containing 100 µg/mL ampicillin. These cultures were allowed to incubate for 16-18 hours at 37 ºC with shaking at 180 rpm. The entirety of the 20mL culture was then transferred to 1L of LB+B_1 or TB+B_1 containing 100 µg/mL of ampicillin. For standard protein expression in LB+B_1, these 1L cultures were incubated at 37 ºC with shaking at 180 rpm until reaching an OD_{600} = 0.6-0.8. Once the desired OD was reached, the temperature was reduced to 18 ºC and IPTG was added to a final concentration of 1mM to induce gene expression via the Lac operon mechanism. The cultures were then incubated at 18 ºC for 18-20 hours with shaking at 180rpm.

2.3.2 Production of isotopically labeled protein

In the case of ^{15}N and ^{13}C labeled protein expression, a 20 mL TB+B_1 starter culture was added to 1 L of TB+B_1 plus 4.0 g of tryptone along with 5.0 g of NaCl containing 100 µg/mL ampicillin. The 1 L culture was incubated at 37 ºC with shaking at 180 rpm until OD_{600} = 1.6-1.9. The culture was then centrifuged at 8000xg for 15 minutes. The cell pellet was washed by resuspension in 40 mL of 1X PBS followed by another centrifugation at 8000xg for 15 minutes. The pellet was resuspended in M9 Minimal Media (Table 2.10) and transferred to 950 mL of M9 Minimal Media containing 100 µg/mL ampicillin. A 50 mL volume of Nutrient Mix (Tables 2.11 and 2.12) was then filter sterilized (0.22 µm) and added to the 950 mL of M9 Minimal Media for a final volume of 1 L. The culture was then normalized at 18 ºC for 30 minutes before adding

IPTG to a final concentration of 1 mM and incubating for 18-20 hours at 18 ºC with shaking at 180 rpm.

Following protein expression, both unlabeled and labeled cell cultures were harvested by centrifugation at 8000xg for 15 minutes. Cell pellets were then washed in 40 mL of 1X PBS, transferred to 50 mL tubes and centrifuged again at 8000xg for 15 minutes. Final cell pellets were either frozen at -20 ºC for later use or taken immediately for protein purification.

Table 2.10 M9 Minimal Media Recipe – 950mL

Reagent	Quantity
Sodium phosphate, dibasic (anhydrous)	6.0 g
Potassium phosphate, monobasic (anhydrous)	3.0 g
Sodium Chloride	0.5 g
ddH$_2$O	Up to 950 mL

Table 2.11 Nutrient Mix Recipe – 50mL

Reagent	Quantity
^{15}NH$_4$Cl	1.0 g
Glucose (or ^{13}C-glucose)	2.0 g
1 M MgSO$_4$	0.5 g
50 mM CaCl$_2$	4 mL
20 mg/mL Thiamine	1.0 mL
3 mM FeCl$_3$	400 μL
Metal Mix (Table 2.12)	500 μL
ddH$_2$O	Up to 50 mL

Table 2.12 Metal Mix Recipe – 500mL

Reagent	Quantity	Final conc.
$ZnSO_4$	323 mg	4 mM
$MnSO_4$	75.5 mg	1 mM
H_3BO_3	145 mg	4.7 mM
$CuSO_4$	55.9 mg	0.7 mM
ddH_2O	Up to 500 mL	

2.4 Protein Purification

2.4.1 Affinity chromatography of $PPL_{1626-1756}$ and $PPL_{1626-1756}$ S1657E

Pierce™ Protease Inhibitor mini tablets (Thermo Scientific™) were dissolved into 50 mL of 1X PBS. This solution was used to resuspend the cell pellet. The cells were then homogenized on ice and subsequently lysed on ice with sonication at 80% amplitude for 1 s on/1 s off for a total of 90 seconds. Sonication was performed two additional times with brief swirling of the mixture between runs. The lysed cells were then centrifuged at 75 000xg for 45 minutes at 4 ºC. The supernatant was filtered using a 0.45 µm syringe filter for the purposes of FPLC.

Isotope labeled and unlabeled wild-type and mutant constructs of the extended periplakin linker domain were expressed as N-terminal GST-fusion proteins. As such, affinity chromatography was performed using 5mL GSTrap HP or 20 mL GSTPrep FF 16/10 columns (GE Healthcare Life Sciences) on GE ÄKTA Pure or ÄKTA Start systems. All buffers were filtered (0.45 µm filter) and degassed before use.

The column was initially equilibrated with 1X PBS for 3 CV at 2 mL/min. The filtered supernatant was then applied to the column at 0.5 mL/min with typical volumes ranging from 24-30 mL. The column was then washed with 4 CV of 1X PBS at 1 mL/min. Elution of the GST-fusion protein was done using 10 mM glutathione, 20 mM HEPES, 150 mM NaCl, pH 8.0. This buffer was applied for 2.5CV at 0.5 mL/min. Elution fractions of 1.5 mL were collected. Final equilibration of the column was performed with an additional 3 CV of 1X PBS.

The peak in UV_{280} absorption of the elution chromatogram was used to inform pooling of appropriate elution fractions. Concentration of the pooled elution fractions was roughly determined using A_{280}. The blank used was 99 µL of elution buffer. A 1 µL aliquot of pooled elution was then added to the blank, mixed by pipetting, and gave A_{280} readings ranging from 0.096-0.213. Accounting for the 1:100 dilution of sample, this put the yield of GST-fusion protein at 9.6-21.3 mg/mL in 8-14 mL. The samples grown in M9 Minimal Media accounted for the yields at the lower end of the range. For GST tag cleavage, 1 mg/mL 3C PreScission protease was then added at 10 µL per milligram of protein. Tag cleavage occurred for 16-18 hours at 4 ºC. The cleaved GST tag and the N-terminally extended periplakin constructs were then isolated from each other using size exclusion chromatography (see section 2.4.3).

2.4.2 Immobilized Metal Affinity Chromatography (IMAC) of vimentin$_{99-249}$

Cell pellets of the vimentin construct were resuspended, lysed, and centrifuged in the same manner as the extended periplakin linker cell pellets. However, EDTA-Free Pierce™ Protease Inhibitor tablets were used instead. Following the centrifugation at 75 000xg for 45 minutes, the supernatant was filtered using 0.45 µm syringe filters.

The vimentin construct was designed with a C-terminal His$_6$ tag which allows for purification via Ni^{2+}-NTA affinity chromatography. A 5 mL GE Health Life Sciences HisTrap

High Performance (HP) column was equilibrated with 5 CV of 10 mM imidazole, 20 mM HEPES, 500 mM NaCl, pH 7.5 at 1 mL/min. Sample was then loaded onto the column at 0.5 mL/min before washing the column with 5CV of 50 mM imidazole, 20 mM HEPES, 500 mM NaCl, pH 7.5 at 1 mL/min. Up-flow elution was then performed with 3 CV of 350 mM imidazole, 20 mM HEPES, 500 mM NaCl, pH 7.5 at 0.5 mL/min. The elution was collected in 1.5 mL fractions. Fractions residing in the peak of the elution profile were pooled, spin concentrated, and run through the same size exclusion chromatography protocol used for the extended periplakin linker constructs.

2.4.3 Size exclusion chromatography of $PPL_{1624-1756}$, $PPL_{1624-1756}$ S1657E, and $vimentin_{99-249}$

This step serves to separate the GST tag from the cleaved $PPL_{1626-1754}$, and as a final step for the purification of $vimentin_{99-249}$. Furthermore, this second purification phase also provided a means to simultaneously buffer exchange proteins into optimal buffers for various assays as identified by Protein Thermal Shift (see section 3.3). Size exclusion chromatography (SEC) was carried out in 20 mM HEPES, 10 mM NaCl, pH 7.5 for proteins to be used in MicroScale Thermophoresis (MST), or 25 mM Tris-HCl, pH 6.3 for those to be used in NMR and SAXS. Protein samples were concentrated to 2mL using centrifugal ultrafiltration devices (MWCO 3kDa, 5 mL or 15 mL). A Sigma-Aldrich HiLoad® 16/600 Superdex® 75pg column was equilibrated with 1.5 CV of buffer at 1 mL/min. Concentrated sample was then applied to the column via capillary loop at 0.2 mL/min. Isocratic elution was then performed in 1.5 mL fractions for 1.0 CV at 0.2 mL/min. This was followed by a final column equilibration of 2.0 CV at 1 mL/min. Elution fractions were analyzed by SDS-PAGE and pure protein was pooled. Protein concentration was determined by A_{280} as well as Bradford assays. For all purifications, 5.0 µL fractions were taken periodically throughout for analysis by SDS-PAGE (Figures 3.1C and 3.3C).

33

2.4.4 Assessment of protein purity

Upon completion of the purification protocols, the degree of protein purity was assessed using densitometry. This was performed during the imaging of SDS-PAGE gels. First, Coomassie Blue stained gels or stain-free gels were loaded into a BioRad Gel Doc EZ Imager on the appropriate tray (White tray for Coomassie stained gels and the Stain-tree tray for stain-free gels). Stain free gels were then developed using a standard 5 minute gel activation protocol. The intensity of protein bands was then quantified in each lane using Image Lab software (BioRad). The intensities of the bands in each lane were then compared to each other to determine their relative abundance. Pure protein used for the experiments described in this thesis was designated to have >95% homogeneity i.e. the band for the desired protein constituted >95% of the total band intensity observed for the lane.

2.5 Western Blotting

Western blotting is a common method used in the identification and detection of specific proteins. Antibodies raised against particular epitopes can bind to proteins transferred from an SDS-PAGE gel to a polyvinylidene fluoride (PVDF) membrane. Following protein transfer, the membrane is blocked, typically with milk, to reduce the incidence of non-specific binding interactions. An incubation is then performed with an antibody which recognizes a target protein for several hours. A wash is then performed to remove any unbound antibody. Any antibody that remains bound is then often detected by a secondary antibody which is conjugated to a species which provides a detectable signal such as Horse Radish Peroxidase (HRP). The primary antibody may also be directly conjugated to HRP, removing the necessity for the secondary antibody. After adding the appropriate substrate for HRP, a chemiluminescent reaction occurs which can be

detected on an x-ray film. In this work, western blotting was used to identify the oligomeric species present over the course of vimentin$_{99-249}$ purification.

2.5.1 SDS-PAGE

Fractions taken over the course of the vimentin$_{99-249}$ purification were run on a precast 12% BioRad gel. Samples fractions included the cell lysate, HisTrap Flow-through, HisTrap Washes, and final HisTrap Elutions. The samples were prepared for polyacrylamide gel electrophoresis by diluting them with 5X Laemmli (Invitrogen) sample buffer containing β-ME (Thermofisher). Approximately 5 µg of protein on each well. Samples were denatured at 105 °C for 10 minutes, placed on ice for 2 minutes, then centrifuged at 3000xg for 2 minutes. Samples were then loaded in their respective lanes and the gel was run at 130 V for 1hr 15 minutes using a BioRad electrophoresis apparatus. 1X Tris/glycine buffer at pH 8.0 was used for gel electrophoresis.

2.5.2 Protein transfer to PVDF membrane

A BioRad western blot apparatus was used to transfer vimentin$_{99-249}$ to a polyvinylidene difluoride membrane (PVDF). Cold transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol) was used to presoak the transfer apparatus (filter papers, sponges, and transfer sandwich) for approximately 15 minutes. The PVDF membrane was activated with pure methanol for approximately 2 minutes and was then placed in the transfer buffer for presoaking.

Following completion of the initial SDS-PAGE, protein was transferred to the PVDF membrane. First, the transfer cassette was assembled using a layer of sponge, then filter paper, PVDF membrane, polyacrylamide gel, filter paper, and a final sponge. Protein transfer was done at a constant voltage, 110 V for 1hour 20 minutes at 4°C.

2.5.3 Protein visualization on PVDF membrane

Following the transfer, the PVDF membrane was stained with Ponceau S stain consisting of 0.1% of Ponceau S (Sigma) and 3% tri-chloroacetic acid (TCA) to visualize the proteins for approximately 45 seconds. The PVDF membrane was washed three times with ddH$_2$O and then washed with 0.05% PBST (phosphate buffered saline with 0.05% Tween-20).

2.5.4 Western blotting

PVDF membrane containing the transferred protein was then blocked with 5% non-fat dry milk diluted in 0.05% PBST for 2 hours at room temperature with constant shaking on a rocking table to block non-specific binding sites. Following blocking, the primary antibody, anti-6xHis-tag conjugated to Horse-radish peroxidase (HRP) was diluted 1:5000 with 3% non-fat dry milk and incubated with the PVDF membrane for 1 hour 30 minutes at room temperature with constant shaking on a rocking table. Following incubation, the PVDF membrane was washed 6 times with 0.1% PBST for 5 minutes each time. Proteins were then visualized by adding enhanced chemiluminescence reagent (ECL) on PVDF membrane and incubating for 1 minute, sealed in saran wrap and exposed to autoradiography film (Fujifilm) for 1 minute of exposure time.

2.6 Protein Thermal Shift

Protein thermal shift has a wide range of applicability. The assay itself consists of combining protein, buffer, and SYPRO Orange[57] (Figure 2.1) dye in various combinations in a 96 well plate. At low temperature, the protein of interest resides in its natively folded conformation. As the temperature of the system begins to increase, the added thermal energy causes a gradual unfolding of the protein. As the protein becomes sufficiently unfolded, the hydrophobic residues in the core of the protein become exposed. Interaction of hydrophobic residues with the SYPRO

Orange dye produces a fluorescent signal, so protein unfolding can be tracked as a function of temperature up to a maximum fluorescence where the protein is completely unfolded. The normalized fluorescence values can be analyzed to determine the T_m of the protein, or the temperature at which the inflection point occurs. In the context of this work, PTS was used to identify optimal buffers to implement in the structural and binding affinity studies for the periplakin linker domain constructs.
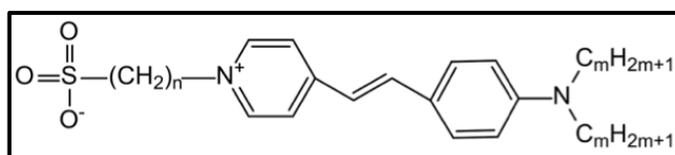
Figure 2.2 Structure of SYPRO Orange dye.
The chemical structure of the SYPRO Orange dye used in Protein Thermal Shift experiments where n=3 and m=5. Excitation wavelength = 300/372 nm. Emission wavelength = 570 nm. Taken from (Yarmoluk et al., 2011).

Assays were prepared in a 96-well plate by combining buffer conditions outlined in Table 2.13 with 1X SYPRO Orange dye and 0.08mg/mL of protein in a total volume of 25 µL. Fluorescence values were obtained using a ThermoFischer QuantStudio 3 system sampling every 0.2 °C. Triplicate values were obtained by assaying three separate wells of identical conditions (as seen in Table 2.13). Samples were compared to two types of control wells: buffer with protein but lacking dye and buffer with dye but lacking protein.

Table 2.13. 96-well plate set up for Protein Thermal Shift

Primary Buffer	1	2	3	4	5	6	7	8	9	10	11	12
25mM Tris pH 6.3	300mM NaCl	300mM NaCl	300mM NaCl	100mM NaCl	100mM NaCl	100mM NaCl	50mM NaCl	50mM NaCl	50mM NaCl	0mM NaCl	0mM NaCl	0mM NaCl
25mM Tris pH 7.0	300mM NaCl	300mM NaCl	300mM NaCl	100mM NaCl	100mM NaCl	100mM NaCl	50mM NaCl	50mM NaCl	50mM NaCl	0mM NaCl	0mM NaCl	0mM NaCl
25mM Tris pH 8.0	300mM NaCl	300mM NaCl	300mM NaCl	100mM NaCl	100mM NaCl	100mM NaCl	50mM NaCl	50mM NaCl	50mM NaCl	0mM NaCl	0mM NaCl	0mM NaCl
20mM HEPES pH 7.0	300mM NaCl	300mM NaCl	300mM NaCl	100mM NaCl	100mM NaCl	100mM NaCl	50mM NaCl	50mM NaCl	50mM NaCl	0mM NaCl	0mM NaCl	0mM NaCl
20mM HEPES pH 7.5	300mM NaCl	300mM NaCl	300mM NaCl	100mM NaCl	100mM NaCl	100mM NaCl	50mM NaCl	50mM NaCl	50mM NaCl	0mM NaCl	0mM NaCl	0mM NaCl
20mM HEPES pH 8.0	300mM NaCl	300mM NaCl	300mM NaCl	100mM NaCl	100mM NaCl	100mM NaCl	50mM NaCl	50mM NaCl	50mM NaCl	0mM NaCl	0mM NaCl	0mM NaCl
PBS	pH 7.0	pH 7.0	pH 7.0	pH 7.5	pH 7.5	pH 7.5	pH 8.0	pH 8.0	pH 8.0			
Controls	25mM Tris pH 6.3, 100mM NaCl	25mM Tris pH 7.0, 100mM NaCl	25mM Tris pH 8.0, 100mM NaCl	20mM HEPES pH 7.0, 100mM NaCl	20mM HEPES pH 7.5, 100mM NaCl	20mM HEPES pH 8.0, 100mM NaCl	PBS pH 7.0	PBS pH 7.5	PBS pH 8.0	25mM Tris pH 7.0, 100mM NaCl	20mM HEPES pH 7.5, 100mM NaCl	PBS pH 7.5
					Dye-free						Protein-free	

38

2.7 Small Angle X-ray Scattering (SAXS)

Small Angle X-ray Scattering is a method used to elucidate relatively low-resolution envelope structures of proteins or protein complexes. Synchrotrons are used to generate high energy x-rays which are passed through a protein sample. As the photons come into contact with the sample, they are scattered at characteristic angles dependent on the shape of the object on which the photon is incident. Because the components of the sample are tumbling randomly in solution, the scattering of the incident x-ray beam is isotropic. The resulting scattering pattern is detected and can be used to elucidate a low resolution envelope of the protein sample in solution after subtracting the background scattering produced by the buffer. Values obtained from this technique include the radius of gyration (Rg), and maximum dimension (Dmax) of the protein.

All protein samples sent for SAXS data collection were purified as described in section 2.4. Purified protein was concentrated to 10-12 mg/mL using centrifugal ultrafiltration devices. Data for all SAXS samples was collected by technicians at Diamond Light Source in Oxfordshire, UK. Concentrated protein sample was loaded onto a Superdex 200pg column and eluted using 25mM Tris, pH 6.3. Protein was eluted directly into the X-ray beam for data collection.

Buffer subtraction was performed with ScÅtter (BIOISIS). All subsequent analysis and processing was done using the ATSAS suite[58]. Buffer-subtracted data was loaded into PRIMUS[59], scaled and merged, then analyzed in a Guinier plot[60,61]. Once Guinier analysis confirmed the monodispersed, aggregate-free nature of the sample, data was analyzed in a Kratky plot to qualitatively assess the degree of structural rigidity and "folded-ness" of the sample[62]. The merged data set was then used to calculate the P(r) curve for the sample[63,64]. DAMMIN/F[65,66] was then used to generate 17 models which were averaged with DAMAVER[67] to generate a final *ab initio* bead model[68]. Models were visualized using Chimera[69].

2.8 Nuclear Magnetic Resonance Spectroscopy (NMR)

Nuclear magnetic resonance spectroscopy relies on two inherent properties of atoms: spin and magnetism. Molecules are placed in a strong magnetic field. The atoms are then pulsed with a relatively weak radio-frequency which stimulates a transition of the atoms to a higher energy spin state. As these higher energy states begin to relax, a signal is produced which can be detected.

Double and triple resonance NMR experiments were performed by Dr. Jitendra Kumar on a Bruker 800 MHz spectrometer at 25 ºC (298 K). Double resonance experiments were performed using protein at 400 µM. Triple resonance experiments were performed using protein at 800µM. The buffer selected for all NMR experiments was 25 mM Tris-HCl, pH 6.3 with 10% D_2O. Optimal buffer conditions were identified using PTS (see section 3.3).

2.9 MicroScale Thermophoresis (MST)

MicroScale Thermophoresis is a technique used to elucidate binding kinetics of a wide variety of chemical species[70]. Because MST relies on ubiquitous, inherent physical properties such as hydration shell, charge, and size, it can be used in a multitude of different systems from simple ion binding to large-scale protein-protein interactions. In this case, MST was used to quantify the binding affinities of $PPL_{1626-1756}$ and $PPL_{1626-1756}$ S1657E to $vimentin_{99-249}$.

2.9.1 Preparation of labeled vimentin$_{99-249}$

Purified $vimentin_{99-249}$ was diluted to 1 µM by adding MST buffer (20 mM HEPES, 10 mM NaCl, pH 7.5). Tween-20 (20%) was then added to a final concentration of 0.015%. A 2 µL aliquot of 5 µM RED-tris-NTA dye (Nanotemper technologies MO-L008) was resuspended to 500 nM using the same MST buffer and 20% Tween-20 was again added to a final concentration of

0.015%. A 20 µL volume of 1µM vimentin$_{99-249}$ was added to 20 µL of 500 nM RED-tris-NTA dye, yielding a 2:1 mixture of vimentin:dye and mixed by pipetting. The mixture was allowed to incubate at room temperature for 30 minutes. Following the incubation, the mixture was centrifuged at 15 000xg for 10 minutes at 4 ºC. The labeled vimentin was then kept on ice until it was used. Fully dyed vimentin was designated to have an effective starting concentration of 250 nM.

2.9.2 Preparation of PPL$_{1626-1756}$ (S1657E) ligand and dilution series

Purified PPL$_{1626-1756}$ and PPL$_{1626-1756}$ S1657E were dialyzed in 1 L of MST buffer using Pur-A-Lyzer Midi 3500 Dialysis Kits (Sigma). Dialyzed protein was then concentrated using centrifugal ultrafiltration units (3kDa, 500µL) until the protein reached ~1.1 mM as quantified by A$_{280}$ and their respective extinction coefficients. Tween-20 (20%) was then added to the concentrated proteins to a final concentration of 0.015%.

To the first tube, 54 µL of 1.1 mM protein was added. To tubes 2-16, 27 µL of MST buffer containing 0.015% Tween-20 was added. From tube 1, 27 µL was removed and transferred to tube 2. The contents of tube 2 were then thoroughly mixed by pipetting up and down, taking care to minimize the number of bubbles introduced. A volume of 27 µL was then taken from tube 2 and transferred to tube 3, which was then mixed as described previously. This series of dilutions was repeated to tube 16, where the final 27 µL removed was discarded, leaving 27 µL in tubes 1-16.

2.9.3 MST binding experiments

To the 27 µL in each tube of the dilution series, 3 µL of dyed vimentin$_{99\text{-}249}$ from section 2.7.1 was added, using a fresh pipette tip each time. The dyed vimentin$_{99\text{-}249}$ was then thoroughly mixed into each tube by pipetting without introducing bubbles. The final tubes contained 30 µL of sample consisting of 25 nM labeled vimentin with PPL$_{1626\text{-}1756}$ or PPL$_{1626\text{-}1756}$ S1657E content ranging from 1 mM-30 nM. Following thorough mixing, the tubes were allowed to incubate at room temperature for 10 minutes. The contents of each tube were then loaded (~10 µL) into Monolith NT.115 Capillaries (MO-K022) and placed in the capillary tray.

MST runs were performed on a Nanotemper Technologies Monolith NT.115 system at 25 ºC. Excitation power was set to 40% and MST power was set to medium. Following the initial run, new capillaries were loaded with the contents of the same tubes to obtain triplicate values. Data analysis was performed using MO.Affinity Analysis software v2.2.6. (Nanotemper Technologies).

CHAPTER 3

GENERAL RESULTS AND DISCUSSION

Chapter Three: General Results

3.1 Purification of PPL$_{1626-1756}$, PPL$_{1626-1756}$ S1657E, and Vimentin$_{99-249}$

3.1.1 PPL$_{1626-1756}$

The extended periplakin linker was purified as previously described in sections 2.4.1. and 2.4.3. The elution peak for the initial GST-affinity chromatography tracked by A$_{280}$ (Figure 3.1A) reveals a large quantity of pure GST-fusion protein purified from cell lysate, indicating efficient expression by the BL21(DE3) *E. coli* strain. This data is corroborated by SDS-PAGE where a thick band is visible at 42 kDa in Figure 3.1C. Typical yields of GST-fusion protein ranged from approximately 60 mg/mL-110 mg/mL, with the lower yields corresponding to labeled protein grown in minimal media and the higher yields corresponding to protein grow in in LB.

Cleavage of the GST tag with 3C PreScission protease was found to be sufficient for the separation of the tag from PPL$_{1626-1756}$. This is reflected both in the size exclusion chromatogram (Figure 3.1B) and the SDS-PAGE analysis (Figure 3.1C). The size exclusion chromatogram reveals two elution peaks; the first peak (58-65 mL) corresponds to isolated GST, which elutes earlier than PPL$_{1626-1756}$ since it is larger. The second peak (67 mL-75 mL) corresponds to isolated PPL$_{1626-1756}$ and has a substantially lower A$_{280}$ value than the first peak. This is a result of the differing extinction coefficients (ε) for GST and PPL$_{1626-1756}$ at 280 nm[71]. The majority of absorbance at 280nm is done by Trp and Tyr residues of which, the cleaved GST has 4 and 14, respectively, while PPL$_{1626-1756}$ has 2 of each. This difference in absorptive capability results in $\varepsilon_{GST} = 42860$ cm^{-1}M^{-1} and $\varepsilon_{PPL1626-1756} = 13890$ cm^{-1}M^{-1}. Ultimately, this leads to PPL$_{1626-1756}$ producing a much weaker absorbance signal for the same stoichiometric quantity of protein. The SDS-PAGE shows two intense bands corresponding to the cleavage products which were

successfully separated with SEC: GST at 26 kDa, and $PPL_{1626-1756}$ at 15 kDa. The protein was ultimately purified to >95% homogeneity as confirmed by densitometry.
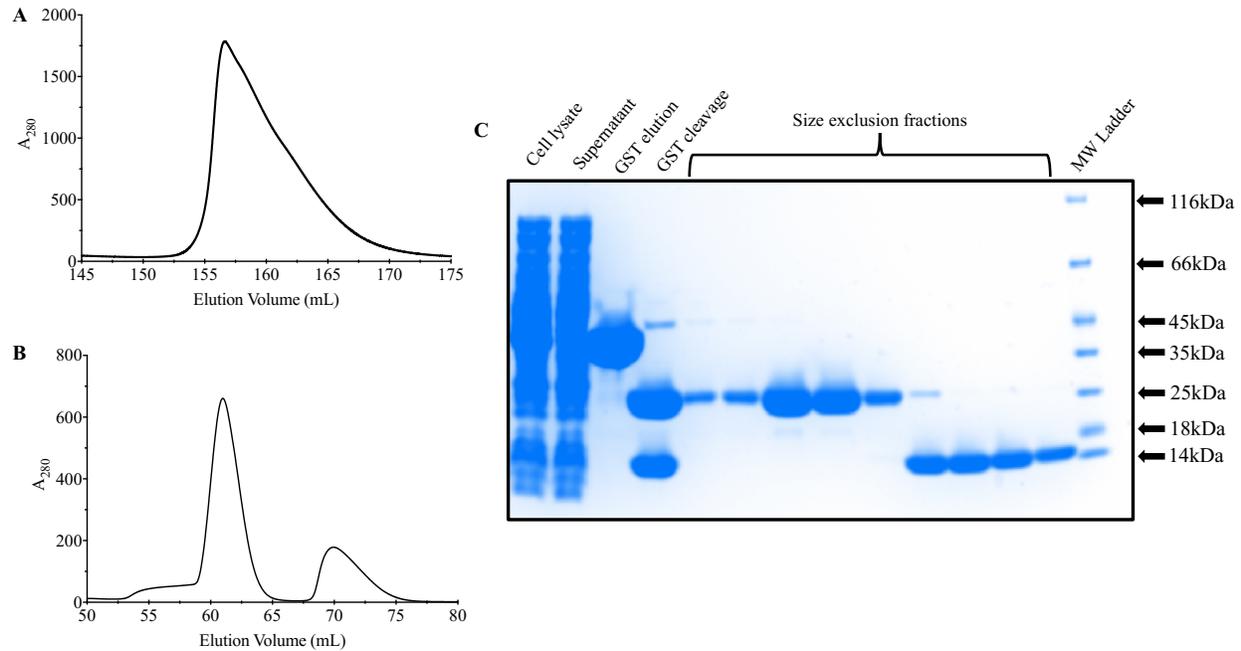
Figure 3.1 Purification of $PPL_{1626-1754}$.
A) Chromatogram of the GST-affinity chromatography elution as tracked by A_{280}. Fractions between 155 mL-165 mL were pooled and underwent 3C PreScission protease cleavage. **B)** Chromatogram of the SEC for separation of the GST tag and pure $PPL_{1626-1756}$. The first peak (58 mL-65 mL) corresponds to the GST tag. The second peak (67 mL-75 mL) corresponds to pure $PPL_{1626-1756}$. **C)** 4-20% Tris-glycine polyacrylamide gel (BioRad) showing SDS-PAGE analysis of fractions taken over the course of the purification of wild-type $PPL_{1626-1754}$. Fractions of 5 µL were taken for each sample. They were mixed with 5 µL of 2X Laemmli Sample Buffer, boiled at 98 ºC for 5 minutes, allowed to cool, then spun down briefly before loading. The gel was run at 200 V for 28 minutes, washed with ddH_2O for 15 minutes, then stained with Coomassie Blue for 45 minutes. The gel was destained overnight with ddH_2O. "GST elution" fractions produce a band at approximately 42 kDa, the expected MW for the GST-fusion protein. The "GST cleavage" yields a faint band at 42 kDa, indicative of uncleaved protein, and intense bands at 26 kDa and 15 kDa, which correspond to the expected MWs of GST and periplakin S1626-K1756. The size exclusion fractions containing only protein at 15 kDa were pooled and used as pure sample.

3.1.2 PPL$_{1626-1756}$ S1657E phosphomimic

The mutation of serine 1657 to glutamate creates a phosphomimic of the wild-type protein. A phosphorylation event at a Tyr, Thr, or Ser residue introduces a bulky group with a charge of approximately -1.5 at pH 7, which often results in change to protein tertiary or quaternary structure. These changes in structure can lead to an alteration of protein activity and functionality. Alternatively, phosphorylation of specific residues may leave protein folding and structure unaltered and instead can contribute to other peripheral functionality of the phosphorylated species such as signalling or recognition by other proteins. Because canonical phosphorylation relies on kinases, difficulty is introduced for *in vitro* study of phosphorylated protein species for several reasons: an ideal kinase will phosphorylate the entirety of a substrate population, it will do so at only one specific site, and it must be subsequently removed from solution. To circumvent these issues, we can do a point mutation of serine to glutamate instead. As reflected in Figure 3.2, this mutation introduces a charge of -1 where there was previously none and increases the size of the amino acid side chain, mimicking a phosphorylation event. This method also guarantees that each copy of the protein produced will be in the phosphomimic state. It also enables use of the identical purification protocol as the wild-type protein.

The wild-type and mutant constructs of the periplakin linker domain share identical purification protocols and their results are unsurprisingly similar. The phosphomimic protein is effectively expressed in *E. coli* BL21(DE3) cells and initially purified as a GST-fusion protein in sufficient quantities with high purity as indicated by Figure 3.3A and the 42 kDa band in the "GST elution lane" of Figure 3.3C. The cleavage of the GST tag with 3C PreScission protease is also highly efficient and reaches completion following the protocols described in section 2.4.1, seen in the "GST cleavage" lane of Figure 3.3C. Size exclusion chromatography is able to separate the
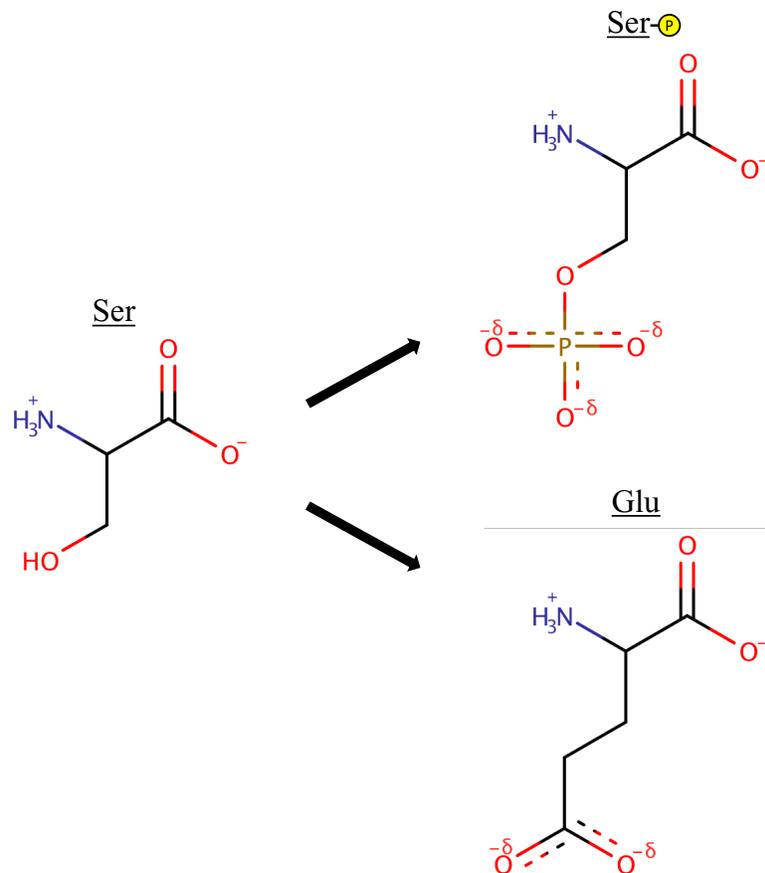
Figure 3.2 Illustration of phosphorylation versus phosphomimic mutagenesis.
The phosphorylation of serine results in the addition of a bulky, large negatively charge group which has consequences on protein structure and function. Mutation of a serine to a glutamate introduces similar characteristics at the same residue. This helps avoid issues with kinase activity and specificity and helps to streamline overall workflow since purification procedures are identical between wild-type and mutant protein.

GST tag from the phosphomimic protein effectively (Figure 3.3B). The first peak in the SEC chromatogram (58-64 mL) corresponds to isolated GST and appears in Figure 3.3C as bands at 26 kDa. The second peak (65-70 mL) corresponds to pure $PPL_{1626-1756}$ S1657E, and appears in Figure 3.3C as the bands present at 15 kDa. Unlike in the purification of wild-type $PPL_{1626-1756}$, the mutant purification shows the second elution peak with a more intense A_{280}. This is because the size exclusion step was done twice since the first round of SEC resulted in an appreciable amount of GST contamination in the pure $PPL_{1626-1756}$ S1657E fractions. The second round of SEC

removed the contaminating GST and consisted mostly of already pure $PPL_{1626-1756}$ S1657E. Pooled fractions of $PPL_{1626-1756}$ S1657E were ultimately purified to >95% homogeneity as confirmed by densitometry.
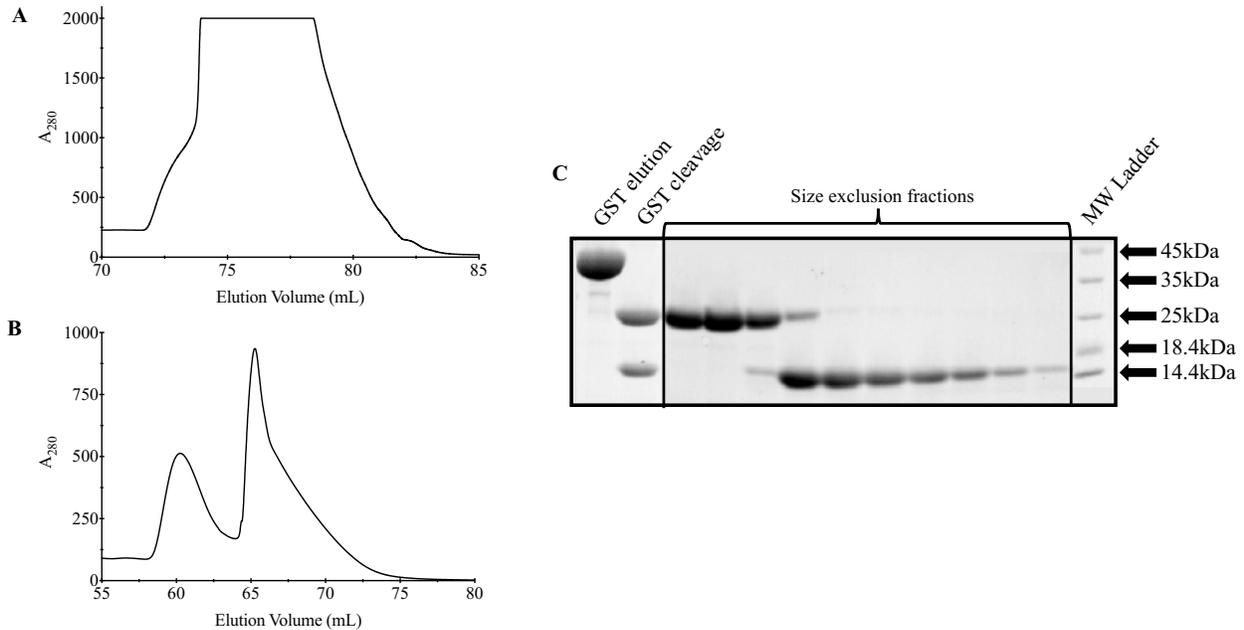
Figure 3.3 Purification of $PPL_{1626-1756}$ S1657E.
A) Chromatogram of the GST affinity chromatography elution from an AKTA Start. The A_{280} detector maximizes at 2000 milli-absorbance units (mAU). **B)** Chromatogram of the SEC performed on the cleaved mutant periplakin linker domain. The first peak (58-64 mL) corresponds to isolated GST while the second peak (65-70 mL) corresponds to the isolated mutant linker domain. **C)** Composite of two SDS-PAGE gels tracking the purification of mutant periplakin linker domain. Lanes containing only the linker domain were pooled and taken as pure final sample. The gel was prepared in an identical manner to that described in Figure 3.1C.

3.1.3 Vimentin$_{99-249}$

IMAC chromatography was able to purify vimentin$_{99-249}$ with very few contaminating proteins. Imidazole concentrations of 50 mM and 350 mM in the wash and elution steps, respectively, were found to be critical for efficient purification using this method. These conditions resulted in an intense, concentrated elution from a 5 mL HisTrap HP column as reflected in Figure

3.4A. In the case of vimentin$_{99-249}$, the SEC step of the purification was mainly used as a buffer exchange procedure since the imidazole present in the elution fractions needed to be removed from the purified protein for subsequent assays.
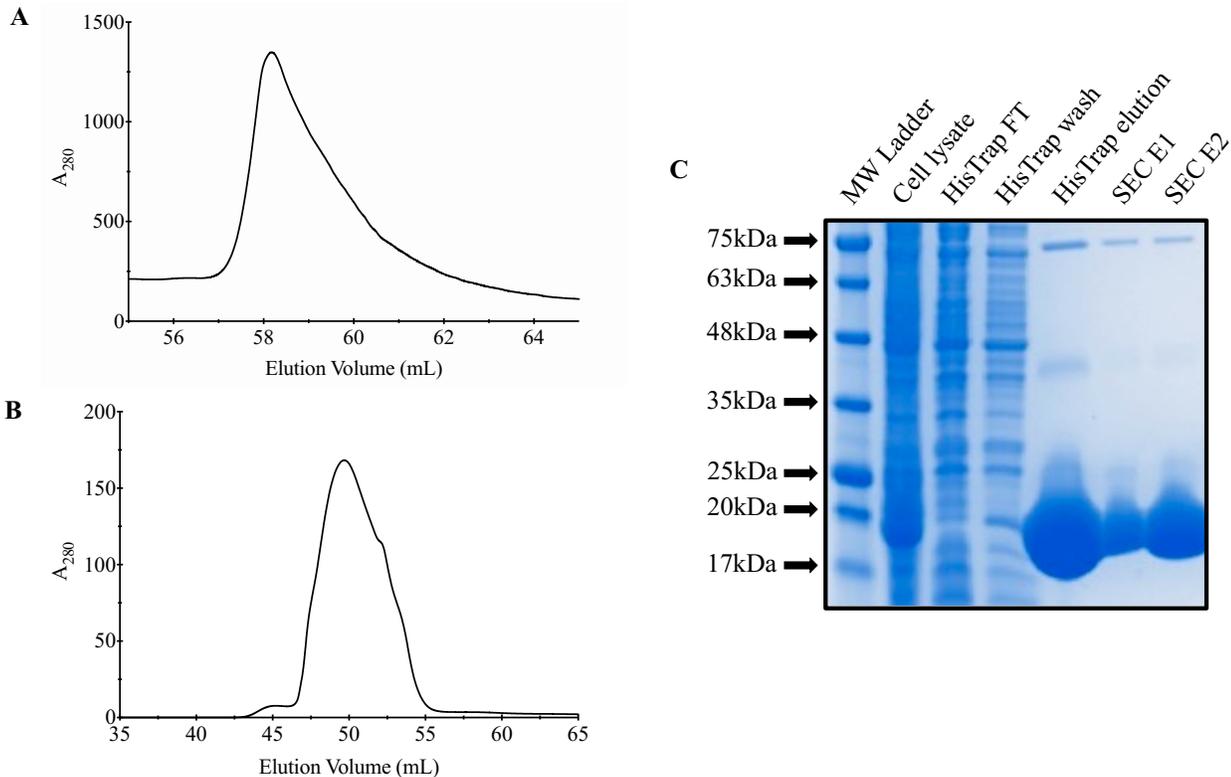
Figure 3.4 Purification of vimentin$_{99-249}$.
A) Chromatogram of the IMAC purification of vimentin showing the elution peak as tracked by A$_{280}$. Fractions falling within approximately 57 mL-61 mL of elution volume were pooled for subsequent SEC. **B)** Chromatogram of size exclusion chromatography of vimentin. A similar elution is peak is observed to that in A), however the magnitude of the A$_{280}$ is substantially lower. Fractions between approximately 46 mL-55 mL were pooled as pure sample. **C)** SDS-PAGE purification analysis of vimentin carried out as described in Figure 3.1C. "HisTrap elution" corresponds to the pooled fractions of A). "SEC E1" and "SEC E2" correspond to fractions falling within the range of B). Monomeric vimentin$_{99-249}$ corresponds to bands at 20 kDa. Dimeric and tetrameric vimentin is also visible at MW of ~ 40kDa and ~75 kDa, respectively as confirmed by Western Blot (see section 3.2, Figure 3.5).

The chromatogram for the SEC (Figure 3.4B) shows a weak signal for A$_{280}$ despite there being large quantities of protein in the fractions of the elution as seen in the "SEC E1" and "SEC

E2" lanes of Figure 3.4C. This is because the vimentin$_{99-249}$ construct contains no Tryptophan residues and only 2 Tyrosine residues, meaning it has a very low extinction coefficient at 280 nm ($\varepsilon = 2980$ cm^{-1}M^{-1}) and produces an unusually weak absorbance signal.

Figure 3.4C shows interesting results in the elution lanes as well: pure, monomeric vimentin can be seen as the large bands at ~ 20kDa. However, in the HisTrap and SEC elution lanes, there are additional bands visible at around 40 kDa and 75 kDa. These higher molecular weight bands are not detectable using Mini-PROTEAN® TGX Stain-Free gel (BioRad) imaging techniques, indicating the absence of tryptophan residues in those protein species. This result suggests that these higher molecular weight bands are dimeric and tetrameric species of vimentin. Corroborating this result is the SEC chromatogram: the species eluting from the column as a single peak centered around ~50 mL, corresponds to a species with a MW of around 75-80 kDa, not the 20 kDa expected for monomeric vimentin$_{99-249}$. The fact that this gel was run under denaturing conditions and in the presence of a reducing agent implies that the tetrameric complexes formed by this vimentin construct are extremely stable. This purification procedure typically yielded around 65mg of pure protein per litre of cell culture.

3.2 Vimentin$_{99-249}$ Forms Tetrameric Species

Western blot analysis of purified vimentin using Anti-6xHis HRP antibody (Figure 3.5) reveals non-specific binding for the lysate fraction as indicated by faint dark bands throughout the lane. No signal was detected in the "His FT" or "His wash" lanes, which contain less protein than the cell lysate fraction and is free from non-specific binding. Interestingly, both fractions taken from the HisTrap elution (His E1 and His E2) show four distinct band populations: a wide, faint band at 20 kDa, a wide, highly intense band at 40kDa, and thinner intense bands at 60 kDa and 80 kDa. These bands all correspond to monomeric, dimeric, trimeric, and tetrameric populations

of vimentin$_{99\text{-}249}$. This result corroborates the findings from the SEC chromatogram and SDS-PAGE of the vimentin purification which previously identified tetrameric vimentin$_{99\text{-}249}$.

The size exclusion fraction lanes show signal at 20 kDa of approximately equal intensity to those seen at 20 kDa in the His elution lanes. There is also a barely-visible signal present at 40 kDa and no signal visible at any higher molecular weights. These two band populations correspond to monomeric and dimeric vimentin$_{99\text{-}249}$. The likely reason for the lack of higher molecular weight bands is the difference in protein concentration between the SEC fractions and the HisTrap elution fractions; protein eluting from the HisTrap column is much more concentrated than protein eluted from SEC and high protein content shifts the equilibrium of oligomeric state towards the tetrameric population. This effect is even present in the difference between "His E1", which was taken at the peak of the elution (58.5 mL in Figure 3.4A) and "His E2" lanes, which was taken at 60 mL of elution volume. The band intensity visibly decreases for the trimeric and tetrameric populations in the "His E2" fraction. The protein was transferred to the PVDF membrane from a denaturing gel one which all four oligomeric states of vimentin are present. Based on these findings, it is clear that under the native conditions under which MST assays were performed, tetrameric vimentin was the functional species, although no functional assays were conducted to rigorously verify this.
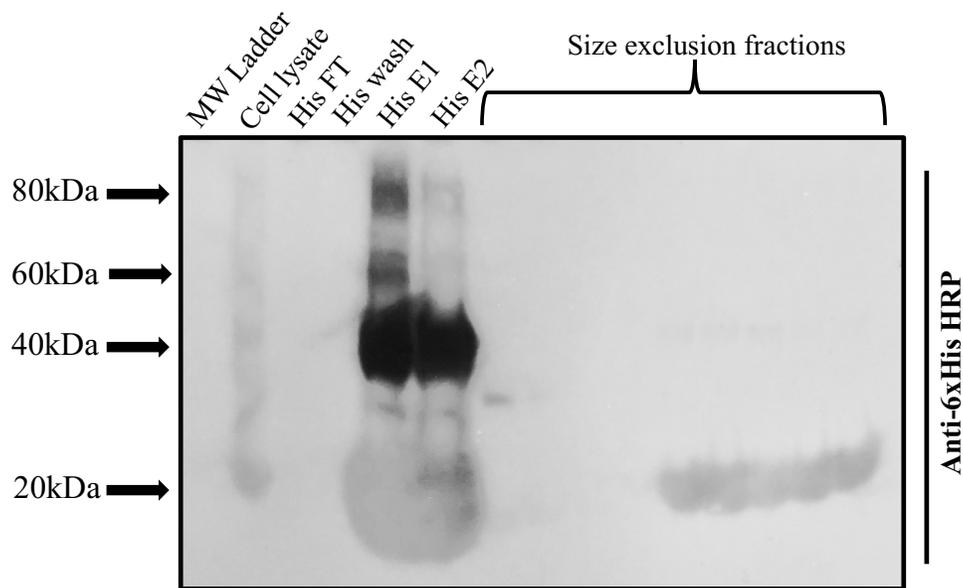
Figure 3.5 Western blot of vimentin₉₉₋₂₄₉ purification.
The antibody used was Anti-6xHis conjugated to Horse Radish Peroxidase (HRP). The film was exposed for 1 minute. Non-specific binding produces signal in the cell lysate fraction. Strong signal is seen for fractions in the HisTrap column elutions (His E1 and His E2) at 20, 40, 60, and 80 kDa, which correspond to monomeric, dimeric, trimeric, and tetrameric vimentin. Size exclusion fractions show monomeric vimentin and faint signal for dimeric vimentin.

3.3 PPL₁₆₂₆₋₁₇₅₆ is Stabilized Under Low Salt Conditions

Many of the experiments in this work, particularly MST and 3D NMR, required that protein be stable at room temperature for prolonged periods of time. For this reason, it was crucial to find buffer conditions which provided a sustained, stable environment while supplying conditions appropriate for a given assay. The Protein Thermal Shift assay provides a high-throughput means of identifying such buffers. MST, used for quantification of binding affinities, was predicted to be sensitive to the salt concentration of buffer since the proposed interaction between PPL₁₆₂₆₋₁₇₅₆ and vimentin₉₉₋₂₄₉ relies on an electrostatic interaction of unknown strength. As such, PTS was used determine the effect of salt concentration on the stability of PPL₁₆₂₆₋₁₇₅₆ at varying pHs. The buffer used for MST was 20 mM HEPES since HEPES provides adequate buffering capacity around

physical pH and because previous analysis of envoplakin's PRD-C to the same vimentin[99-249]

construct found success in HEPES[28]. The chosen starting buffer for 2D and 3D NMR experiments

was the widely used 25 mM Tris-HCl, which has also been previously found to be useful for the

NMR characterization of plakin PRDs[28]

3.3.1 MST Buffer: 20 mM HEPES

As shown in the top three panels of Figure 3.6, 20 mM HEPES was tested at varying pH

conditions. At pH 7.0, 300 mM NaCl yielded the least stable condition (T_m = 49.5±2.7 °C) while

0 mM NaCl yielded the highest stability (T_m = 52.3±0.8 °C). Intermediate stability was found for

100 mM and 50mM NaCl conditions, which had T_m values of 50.1±0.6 °C and 51.7±1.4 °C,

respectively. When the pH is increased to 7.5, the T_m values of the 300 mM, 100 mM, and 50 mM

conditions cluster together at 50.7±2.2 °C, 51.6±0.5 °C, and 51.7±1.6 °C, respectively, while the

T_m of the protein in 0 mM NaCl stands alone at 53.1±1.0 °C. Further increasing the pH to 8.0

causes the T_m values to range from 50.0±0.4 °C to 53.3±0.7 °C with these values corresponding to

the   50 mM and 0 mM NaCl conditions, respectively.

3.3.2 NMR/SAXS buffer: 25 mM Tris-HCl

The bottom three panels of Figure 3.6 show the PTS results for PPL[1626-1756] stability in

25 mM Tris-HCl NMR/SAXS buffer with varying salt and pH conditions. At pH 6.3, proton

signals are expected to be relatively strong in NMR HSQC assays. This pH yields increasing T_m

values with decreasing salt concentration: 46.1±2.9 °C, 50.1±1.7 °C, 51.3±1.7 °C, and 51.6±1.3

°C for 300 mM, 100 mM, 50 mM, and 0 mM NaCl, respectively. Increasing the pH to 7.0 changes

this trend in that the T_m values change to 48.4±0.4 °C, 50.3±1.2 °C, 50.9±1.7 °C, and 51.2±3.2 °C

for 300 mM, 0 mM, 50 mM, and 100 mM NaCl, respectively. The same trend is seen in the pH

8.0 condition as in the pH 6.3 condition with T_m values increasing as the salt concentration is lowered: melting temperatures of 48.8±0.5 °C, 50.2±2.3 °C, 51.6±1.1 °C, and 52.5±1.0 °C were observed for 300 mM, 100 mM, 50 mM, and 0 mM NaCl, respectively.

3.3.3 Discussion

Overall, there is a trend for $PPL_{1626-1756}$ to favor low salt conditions, although there does seem to be some pH dependency on the optimal salt concentration. A separate PTS analysis was conducted using PBS at pH 7.0, 7.5, and 8.0 found that $PPL_{1626-1756}$ was preferentially stabilized at pH 7.5 (data not shown). Although the highest T_m for $PPL_{1626-1756}$ in 20 mM HEPES was found at pH 8.0 and 0 mM NaCl, the buffer that was ultimately selected for MST was 20 mM HEPES, 10 mM NaCl, pH 7.5. This is because we wanted to select a buffer that balanced protein stability with physiologically representative conditions for *in vitro* assay. A pH of 7.5 more closely mimics physiological conditions and according to the PTS data, the T_m at pH 7.5 is only 0.2 °C lower than that for pH 8.0. An issue arises with the salt concentration, however, since 0 mM salt does not accurately reflect physiological conditions. To compromise, 10 mM NaCl was the chosen salt condition for MST assays to balance the maintenance of protein stability while providing robust data in a more physiologically accurate environment.

NMR was used primarily to characterize the structural parameters of the periplakin linker domain. Because these experiments did not require conditions capable of maintaining or mimicking potential physiologically relevant electrostatic interactions, the optimal buffer of 25 mM Tris-HCl, 0 mM NaCl, pH 6.3 was selected for all experiments since lower pH conditions yield stronger proton signal. This buffer also provided only 1 °C less thermal stability than its pH 8.0 counterpart, which was deemed a practical sacrifice for the increase in NMR signal strength.

Initially, PTS was used as a primary method to eliminate early issues encountered with the precipitation of purified protein. This assay yielded important data which informed many of the solution conditions for the purification of proteins in this work. In particular, lowering the salt concentration from 150 mM to 10 mM for the size exclusion chromatography was found to prevent precipitation through several freeze-thaw cycles from -80 °C. Furthermore, PTS was found to be a useful, valuable method for the improvement of buffer conditions for analysis of the periplakin linker domain by MST and NMR, as previously demonstrated with other proteins[72,73].
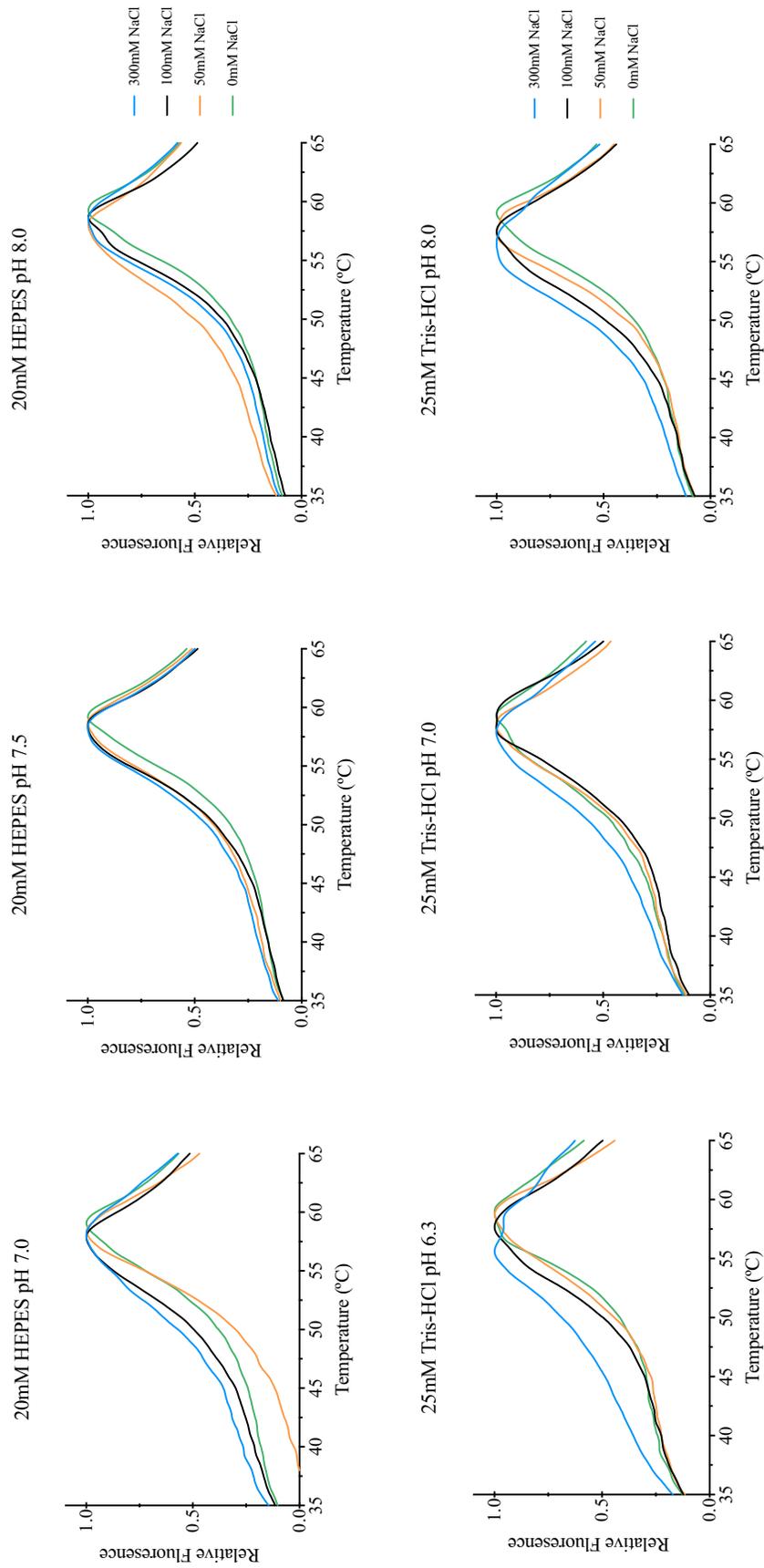
Figure 3.6 Protein thermal shift curves for PPL$_{1626-1754}$.
The protein thermal shift curves for determination of optimal buffer conditions. Analyses were performed on a ThermoFischer QuantStudio 3 with 0.08 mg/mL protein, 1X SYPRO Orange dye, and various buffers. Data was collected in triplicate with readings at every 0.2 °C. In nearly every case, low salt conditions yield the highest T$_m$ for this protein.

3.4 Heteronuclear Single Quantum Coherence (HSQC) Spectrum of PPL$_{1626-1756}$

The HSQC spectrum collected shows peaks from the backbone amide protons of 400 μM PPL$_{1626-1756}$ (Figure 3.7). These peaks range from approximately 6.2-9.8 ppm in the proton dimension and 106-131 ppm in the ^{15}N dimension. Automatic peak picking places the peak count at approximately 150. The cleaved PPL$_{1626-1756}$ has 136 residues, 11 of which are asparagine or glutamine and 5 of which are prolines. This place the peak count of 150 in the acceptable range, since proline residues will not produce a peak in the HSQC fingerprint region due to their lack of amide protons. Asparagine and glutamine residues will produce mirror peaks due to their side chains containing amide groups. Overall, a peak count of 150 would predict a protein of 134 residues, given the number of asparagine, glutamine, and proline residues present.

The spectrum is well resolved, and the peaks are not collapsed around 8.0ppm in the proton dimension, which indicates that the protein is properly folded; each amino acid is residing in a unique chemical environment and consequently display unique chemical shifts. The quality of the HSQC spectra collected justified the acquisition of triple resonance spectra (not shown) to obtain the backbone assignments for this protein to use in the rigorous study of intermediate filament binding mechanics. The 3D spectra gathered in these experiments will be used for determination of the backbone assignments and the solution structure of PPL$_{1626-1756}$, which will improve upon the current crystal structure solved by John Hunt's group.
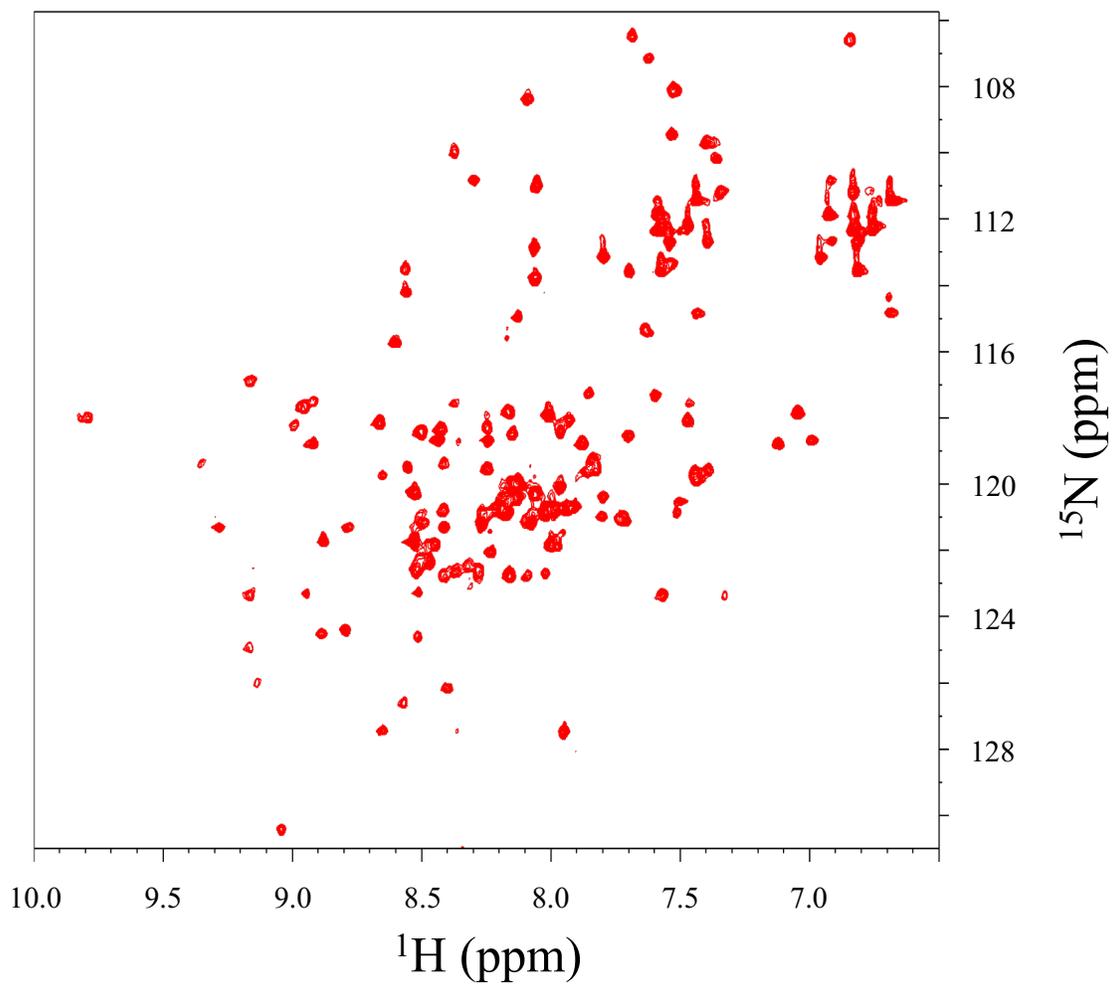
Figure 3.7 HSQC spectrum of PPL$_{1626-1756}$.
The HSQC spectra of 400 µM PPL$_{1626-1756}$ in 25 mM Tris-HCl, pH 6.3 collected on a Bruker 800 MHz spectrometer. The spectrum shows the appropriate number of peaks, nicely resolved in the fingerprint region of the spectrum.

CHAPTER 4

THE FUNCTIONAL ROLE OF THE PERIPLAKIN LINKER DOMAIN

Chapter Four: The Structure and Function of the Periplakin Linker Domain

4.1 The Small Angle X-ray Scattering (SAXS) Structure of PPL$_{1626-1756}$

The SAXS data obtained from DIAMOND Light Source was found to be of adequate quality for subsequent processing as compared to data collected for BSA standards, which were run immediately before PPL$_{1626-1756}$ samples.

4.1.1 Single SAXS envelope for PPL$_{1626-1756}$

The SAXS data collected was processed as described in section 2.5. The results of processing are reflected in Figure 4.1. The 1D scatter plot (Figure 4.1A) shows the experimental scattering values produced by PPL$_{1626-1756}$. These values were used to calculate a P(r) curve (Figure 4.1B). The P(r) curve was used to calculate a radius of gyration (Rg) value of 26.98 Å and a maximum dimension (Dmax) of 93 Å for this sample. The P(r) curve was then used in DAMMIN/F to produce several *ab initio* bead models consistent with the collected data, one of which is shown in Figure 4.1C. The 1D scatter pattern for the generated envelope calculated is then back-calculated, shown as the blue line in Figure 4.1A. This provides a means of quality control in that we can visually inspect the agreement between the back-calculated scatter pattern (blue) and the experimental scatter pattern (red). In this case, the values are highly consistent throughout the plot, providing confidence in the quality of the 3D envelope generated from this data.

A

B

C

$D_{max} = 93Å$

90° 90°
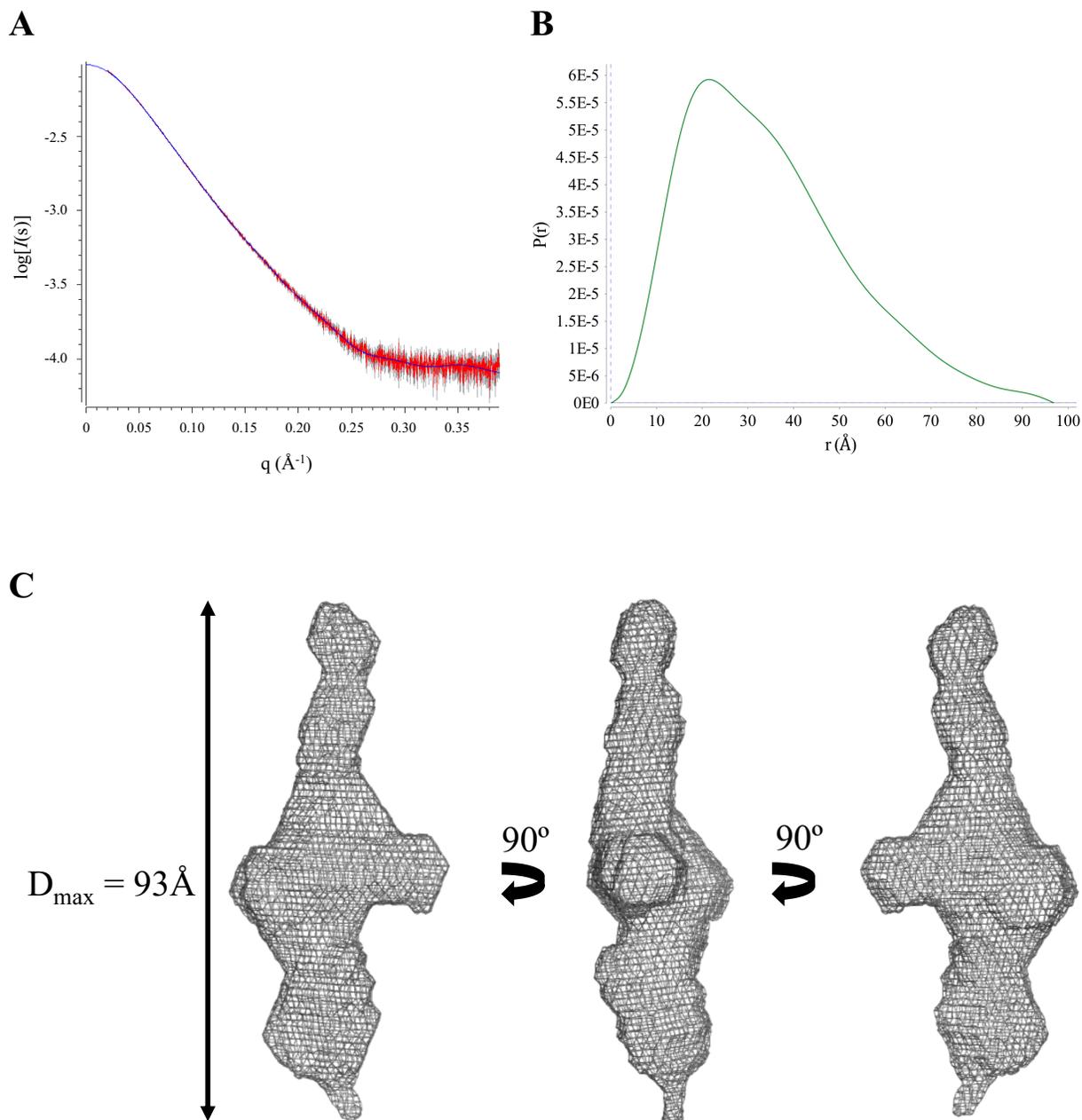
Figure 4.1 Single SAXS envelope of PPL$_{1626-1756}$.
Data was collected at the Diamond Light Source synchrotron. HPLC SAXS was performed on protein at a concentration of 10 mg/mL in 25 mM Tris, pH 6.3. Buffer subtraction, data merging, and Radius of Gyration analysis was done using ScÅtter (BioIsis) and the ATSAS suite. **A)** 1D experimental SAXS scatter pattern (red) and its comparison to back calculation for the DAMMIN/F envelope fit (blue). **B)** The P(r) reciprocal space plot used to generate the DAMMIN/F envelopes. **C)** A single envelope generated by DAMMIN/F for PPL$_{1626-1756}$. Rg = 26.98 Å, D$_{max}$ = 93 Å.

4.1.2 Comparison of averaged PPL$_{1626-1756}$ SAXS envelopes to the existing crystal structure

Initial models generated using the P(r) curve can then be averaged using DAMAVER. Averaging several envelopes produces a final envelope which is more representative of the protein sample population since some components of the protein may be highly dynamic and could produce a relatively wide range of scattering patterns. Figure 4.2 shows the averaged envelope generated from 15 initial envelopes (gray) overlaid with the space-filling model of the PPL$_{1655-1754}$ crystal structure (green) discussed in section 1.4. This envelope reflects the same Rg and Dmax values of each initial envelope: 26.98 Å and 93 Å, respectively.

What becomes apparent in the overlay of these two structures is that the existing crystal structure fits snugly within the averaged envelope. The N-terminus of the crystal structure starts right at the beginning of the upwardly extending unoccupied envelope space; this is likely accounted for given that protein used to generate the envelope has 20 additional N-terminal residues compared to the crystal structure which would extend upwards in the direction of the unoccupied envelope space. The 20 additional residues in the envelope construct form the last stretch of periplakin's central coiled-coil rod domain. Secondary structure prediction places each of these 20 amino acids as part of a helix or coil[74–76]. Although no structure has been solved for this region of periplakin, sequence alignment with a partially solved structure of the coiled-coil rod domain of plectin[77] (PDB: 4GDO) of 55% similarity would indicate that these 20 residues should form a helix as predicted. However, a coil of 20 amino acids would likely not have the length sufficient to occupy the entire volume of the upward extending envelope. This may be because the 20 residues are a relatively poorly behaved helix and could undergo spontaneous and dynamic coiling and uncoiling. The dynamic loss of certain stretches of the helix could extend the length of the coil to occupy more of the space seen in the envelope.

The small bulge of unoccupied envelope space seen in the middle right of the right-most panel of Figure 4.2 may be accounted for given the inaccuracy of the crystal structure introduced by its first 10 non-native residues which participate in an artifactual β-sheet. The 10 native residues present in the envelope construct structure may adopt a different conformation which fills the space unoccupied by the crystal structure, although concrete evidence to support this depends on the completion of the NMR 3D solution structure.

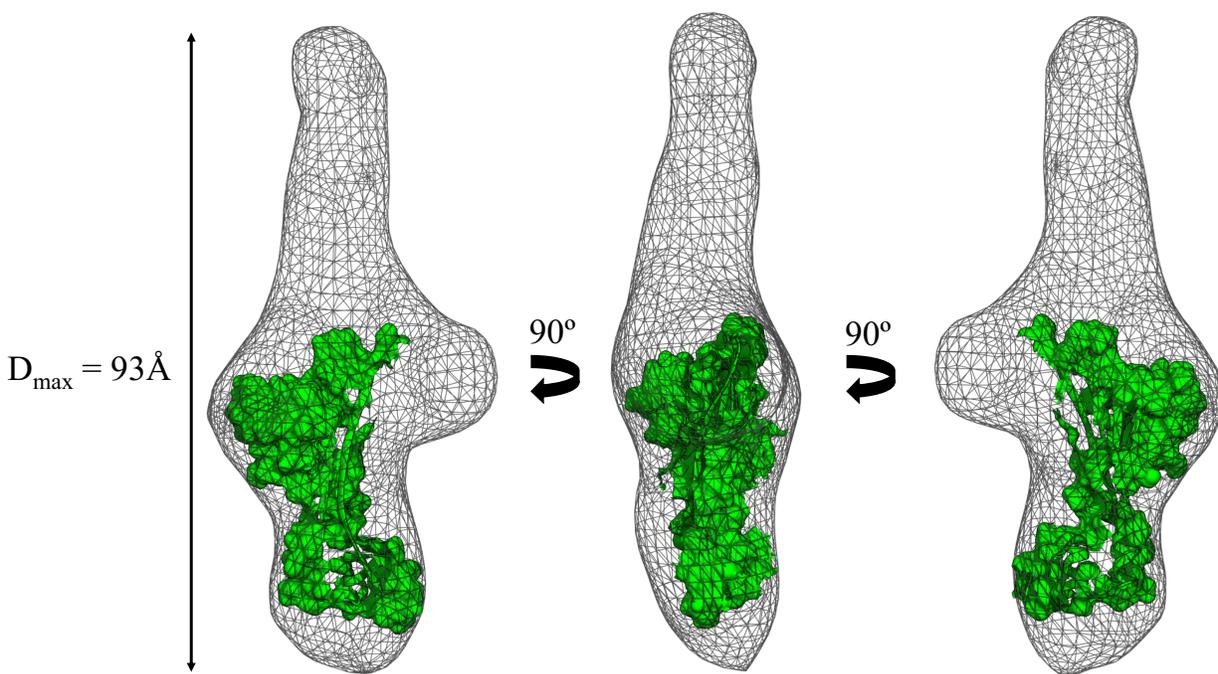**Figure 4.2 Averaged PPL$_{1626-1756}$ SAXS envelope with PPL$_{1655-1754}$ crystal structure overlay.** The averaged SAXS envelope (gray) generated from averaging 17 envelopes using DAMAVER. The existing crystal structure (PDB 4Q28) is overlaid in a space filling model (green).

4.2 Elucidating the Function of PPL$_{1626-1756}$ Using MST

MicroScale Thermophoresis experiments were used to quantify the binding affinity of vimentin$_{99-249}$ for both wild-type and phosphomimic mutant (S1657E) PPL$_{1626-1756}$. All data points were run in triplicate and reproduced in two separate trials. The negative control run for MST consisted of a binding test between BSA and labeled vimentin$_{99-249}$ (not shown). This control showed no quantifiable binding. The positive control was performed by Dr. Catharine Trieber and consisted of a full MST titration of envoplakin's PRD-C with labeled vimentin$_{99-249}$ (not shown). The details of this interaction had been previously quantified by Surface Plasmon Resonance (SPR)[28] and revealed very similar affinity values between the two systems (26±2 µM with MST versus 19 ± 1 µM by SPR.

4.2.1 Wild-type PPL$_{1626-1756}$ binds to vimentin intermediate filaments

Wild-type PPL$_{1626-1756}$ exhibited a K$_d$ of 25 ± 7 µM (the blue curve of Figure 4.3), which is of comparable strength to the interaction of envoplakin's PRD-C to the same vimentin construct as quantified by both SPR and MST. Full saturation of PPL$_{1626-1756}$ occurred between 250 µM - 1 mM with the vimentin$_{99-249}$ concentration kept constant at 25 nM. 50% saturation occurred near 30 µM, with almost no detectable binding occurring around 100 nM of PPL$_{1626-1756}$. The two right-most data points (1 mM and 500 µM) display somewhat odd behavior in that they appear to decrease in their saturation relative to the third data point (250 µM). This is indicative of some ligand-induced fluorescence change at these higher concentrations, possibly due to the viscosity increase resulting from the high protein concentration. The options available to mitigate this effect are quite limited although the effect was highly reproducible for the wild-type PPL$_{1626-1756}$.

4.2.2 Phosphomimic PPL$_{1626-1756}$ S1657E decreases IF binding affinity

The phosphomimic mutant of PPL$_{1626-1756}$ exhibited a K$_d$ of 90 ± 10 µM as shown in Figure 4.3. Full saturation of the protein occurs at the 500 µM – 1 mM concentration and does not appear to display any of the ligand-induced fluorescence change seen in the wild-type case at these higher concentrations. Like the wild-type protein, 50% saturation occurs at approximately 30 µM and binding is undetectable around the 100 nM range. The data reproducibility also appears to be noticeably better over the wild-type protein as indicated by the substantially narrower error bars.
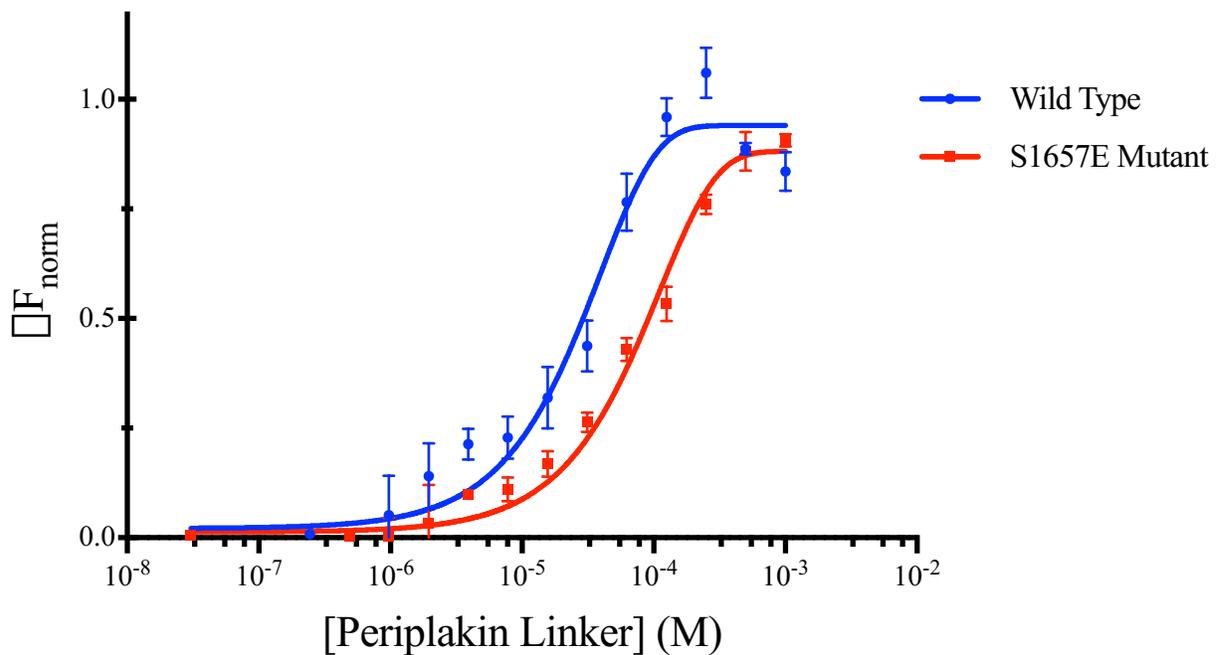
Figure 4.3 Binding curves for WT and phosphomimic PPL$_{1626-1756}$ with vimentin$_{99-249}$.
The normalized MST curves of WT (blue) and phosphomimic (red) PPL$_{1626-1756}$ binding to vimentin$_{99-249}$. Binding quantification analysis revealed WT PPL$_{1626-1756}$ K$_d$ = 25 ± 7 µM while the S1657E mutant K$_d$ = 90 ± 10 µM. PPL constructs were serially diluted from ~1mM down to ~30 nM, while the concentration of labeled vimentin$_{99-249}$ was kept constant at 25nM. (n=3)

4.2.3 Discussion

The results from these MST titrations provide novel insight into the function of the periplakin linker domain; the wild-type protein supports the original hypothesis that the linker

domain does in fact bind to intermediate filaments of the cytoskeleton. Although this has been qualitatively observed[29], it is surprising that the strength of this binding interaction appears to be of comparable strength to the canonical sites for intermediate filament attachment (the plakin PRDs). This finding introduces the linker domain of periplakin at least as an important player when considering desmosomal dynamics, multivalency, and avidity, and opens the road for the linker domains of other plakins to follow. Unpublished results from our lab identified residues critical for maintenance of this interaction based on charge reversal point mutations. As expected, mutation of residues R1655, R1656, K1687, and R1689 resulted in drastic decreases in binding affinity as quantified by MST, with over a 100-fold decrease in affinity for a R1655/56EE double mutant. These results are consistent with the predicted reliance on the basic residues for this interaction as seen in the HADDOCK binding models constructed by our collaborator Dr. Fiyaz Mohammed (see Figure 1.10).

Perhaps more interesting is the effect of the phosphomimic mutation: the steric effects and added negative charge introduced by the point mutation of serine to glutamate result in a clear and quantifiable difference in the affinity of the protein for vimentin filaments. The three-fold decrease in binding strength presented by this construct is consistent with the hypothesis that a phosphorylated state would interfere with the electrostatic contacts between the linker domain and the vimentin filaments. One would expect an even more pronounced reduction in affinity in the case of actual phosphorylation of the serine 1657 residue given its larger steric and electronegative nature. This is corroborated by the MST results from the aforementioned R1655/56EE double mutant, which would place a greater deal of negative charge in the modeled binding site much like an actual phosphate would. The highly reported incidence of phosphorylation at this residue combined with its clear impact on IF binding functionality supports the postulation that

66

phosphorylation of S1657 of the periplakin linker domain may have important implications in the dynamic disassembly of desmosomal complexes.

This becomes more interesting when taken in the context of the alignment of plectin, desmoplakin, envoplakin, and periplakin linker domains[78] (Figure 4.4). It is interesting to note that the phosphorylation site on the periplakin linker domain has a unique sequence: it is preceded by two basic residues (RR, which begin the native sequence of the existing crystal structure), and is immediately followed by four aliphatic residues (IVVI) which begin the first canonical PR of the linker domain. Comparatively, plectin and desmoplakin have an even larger number of basic residues (RKRR) followed immediately by a string of aliphatic residues (VVIV for both), similar to periplakin. However, unlike in periplakin, both the plectin and desmoplakin linkers lack separation of these basic and aliphatic regions by a serine i.e., there is no potential phosphorylation site. The same region of envoplakin appears to be unique as well: the nature of the initial residues are less basic (only a single K) and the following aliphatic region has a serine in the place of an otherwise conserved valine at the second position. To date, this area of envoplakin does not have any corresponding reports of phosphorylation despite the presence of additional adjacent threonine and serine residues instead of basic ones.

Overall, the alignment of the other plakin linker domains with periplakin's provides interesting clues into their possible function. The sequence of the first PR (Figure 4.2, boxed in red) appears to be quite highly conserved, including the glycine which immediately follows the first helix in canonical PRs. The second, non-canonical PR of periplakin (boxed in green) appears to be less well conserved; the critical glycine following the first helix, which is present in the periplakin crystal structure, is only also found in desmoplakin. Another interesting observation of this alignment is the native sequence of the linker domain which participates in the artifactual β-

sheet seen in the crystal structure (boxed in blue). This region appears to be quite highly conserved, particularly at the beginning. The highly conserved nature of this region of the linker domains emphasizes the need for improvement on the structure of this domain. This stretch of residues may impart some unique structural feature, which may account for some of the unoccupied space seen in the SAXS envelope (Figure 4.2), or some unique function which remains to be characterized.
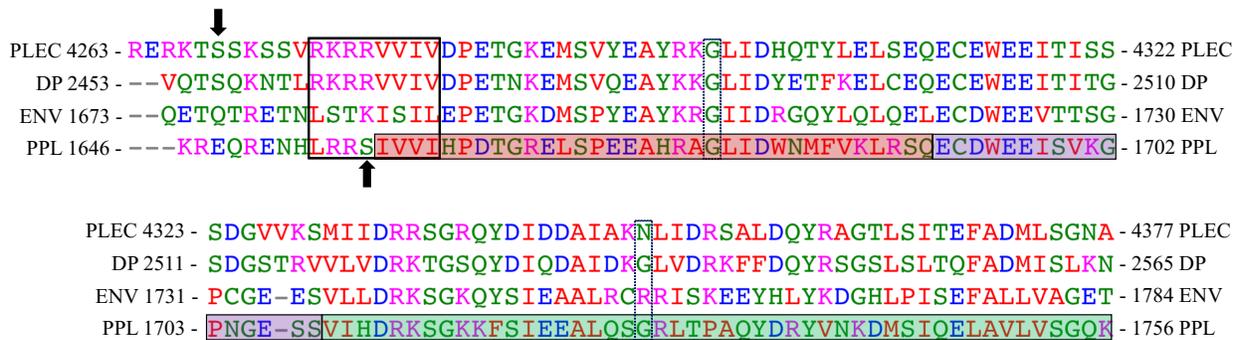


```
PLEC 4263 - RERKTSSKSSVRKRRVVIVDPETGKEMSVYEAYRKGLIDHQTYLELSEQECEWEEITISS - 4322 PLEC
  DP 2453 - --VQTSQKNTLRKRRVVIVDPETNKEMSVQEAYKKGLIDYETFKELCEQECEWEEITITG - 2510 DP
 ENV 1673 - --QETQTRETNLSTKISILEPETGKDMSPYEAYKRGIIDRGQYLQLQELECDWEEVTTSG - 1730 ENV
 PPL 1646 - ---KREQRENHLRRSIVVIHPDTGRELSPEEAHRAGLIDWNMFVKLRSQECDWEEISVKG - 1702 PPL

PLEC 4323 - SDGVVKSMIIDRRSGRQYDIDDAIAKNLIDRSALDQYRAGTLSITEFADMLSGNA - 4377 PLEC
  DP 2511 - SDGSTRVVLVDRKTGSQYDIQDAIDKGLVDRKFFDQYRSGSLSLTQFADMISLKN - 2565 DP
 ENV 1731 - PCGE-ESVLLDRKSGKQYSIEAALRCRRISKEEYHLYKDGHLPISEFALLVAGET - 1784 ENV
 PPL 1703 - PNGE-SSVIHDRKSGKKFSIEEALQSGRLTPAQYDRYVNKDMSIQELAVLVSGQK - 1756 PPL
```

Figure 4.4 Alignment of plakin protein linker domains.
The alignment of plectin (PLEC), desmoplakin (DP), envoplakin (ENV), and periplakin (PPL) linker domains. The red box indicates the first PR seen in the periplakin linker crystal structure. The blue box indicates the region of the periplakin linker which adopts an artifactual β-sheet. The green box indicates the non-canonical PR of the periplakin linker domain. Conserved PR glycine residues are denoted by the dashed boxes. Black arrows indicate the phosphorylation sites of plectin and periplakin.

However, S1657 presents a potentially unique functionality for the linker domain of periplakin compared to the other plakins, given the results of the phosphomimic MST experiments. Phosphorylation of S1657 may be a fundamental step in desmosomal disassembly and regulation for tissues expressing higher levels of periplakin over other plakins. Furthermore, these results provide some insight into the potential implications of the phosphorylation of this residue identified by mass spectrometry in cancers and cell differentiation. Upon phosphorylation, the MST data presented here would support the postulation that desmosomes would favor a disassembled state at the level of cytoskeletal attachment. This may be a critical component of cell differentiation, since differentiated cells may rely on the loss of adhesion to their progenitor.

In plectin, desmoplakin, and envoplakin, phosphorylation has been reported on the PRDs themselves[48] which may have similar effects on their IF binding capabilities, although this has yet to be determined. Furthermore, the linker domain of plectin does have a frequently observed phosphoserine at position 4248[48], the function of which requires characterization. It is difficult to say whether or not this residue would lie in close enough proximity to the clustered basic residues to have the same possible effect on IF binding capability seen with periplakin. Desmoplakin's linker domain shares the same serine residue seen phosphorylated in plectin, though it does not have any reports of phosphorylation to date[48].

Alignment of the periplakin linker domain through phylogenetic space reveals a high degree of conservation (Figure 4.5). In particular, basic residues which have been identified as critical for vimentin binding through modeling and MST studies are highly conserved in the vast majority of cases. This pattern likely would confer the same IF binding capabilities to other species as it does to human periplakin linker domain as presented in this work. Furthermore, the critical, conserved glycine present in the first PR is conserved entirely throughout the variety of species. The second critical glycine which constitutes part of the second, non-canonical PR is also highly conserved in all species except for *Xenopus tropicalis*. These glycine residues indicate the periplakin linker domains of these species would adopt PRs, which have been shown to be the functional units which confer IF binding capabilities as seen in the PRDs of the other plakins. Similarly, the phosphorylated S1657 seen in human periplakin linker domains is conserved in all cases presented here. The presence of this serine residue in a vast array of species, combined with the finding that a mimicked phosphorylated state decreases the linker domain's affinity for vimentin, suggests that the functionality introduced by this serine is evolutionarily important for desmosomal dynamics and assembly. Although further study is required to verify this notion, it

does seem likely to be the case considering the lack of this phosphorylation site in any other plakin protein linker domains.

Figure 4.5 Phylogenetic alignment of the periplakin linker domain.
Alignment of the periplakin linker domains of *Homo sapiens* (Hs), *Calypte anna* (Ca), *Gallus gallus* (Ga), *Alligator sinensis* (Al), *Anolis carolinensis* (Ac), *Xenopus tropicalis* (Xt), *Oreochromis niloticus* (On), *Tetraodon nigroviridis* (Tn), *Danio rerio* (Dr), and *Oryzias latipes* (Ol). The red boxed area indicates the first PR seen in the human periplakin linker domain crystal structure. The green box indicates the non-canonical PR seen in the human periplakin linker domain crystal structure. The conserved glycine residues found in the PRs are indicated by the dashed boxes. The S1657 which is found phosphorylated in *Homo sapiens* is denoted by the black arrow. Figure created and adapted from Michael Overduin.

CHAPTER 5

FUTURE AIMS AND DIRECTIONS

Chapter Five: Conclusions and Future Directions

5.1 Major Conclusions

The work in this thesis focused on elucidating the functional role of the periplakin linker domain. Although much remains unclear, several interesting findings were made regarding this protein: periplakin's linker domain does bind to the intermediate filament vimentin, and mimicked phosphorylation of the serine residue adjacent to modeled binding sites decrease the strength of this interaction, as predicted. These results support the original hypotheses presented in this thesis. The SAXS-generated envelope structure revealed a sound agreement with the existing crystal structure and the artifactual β-sheet in the crystal structure likely accounts for some of the observable discrepancies between the two structures. However, the binding models used to inform these hypotheses were based on the existing crystal structure, which indicates that the crystal structure is still useful and informative.

5.2 Future Aims and Directions

5.2.1 Characterization of vimentin

The periplakin residues involved in the interaction of the periplakin linker and vimentin$_{99-249}$ have been characterized through work in this thesis and unpublished lab data. However, the question remains as to which residues facilitate this interaction on vimentin since the vimentin construct used in this work encompasses coils 1A and 1B. Based on the original models in Figure 1.10, small, 12 residue peptides have been ordered which represent the three unique binding sites on vimentin. Characterization of the interactions of these peptides with the periplakin linker will give insight into the preferred region and mechanism used to facilitate this interaction. Equally

possible is the discovery that there are multiple sites on vimentin capable of this binding event, which would add further still to the complexity of desmosomal dynamics.

5.2.2 NMR backbone assignments and solution structure determination

Although the 3D NMR data for $PPL_{1626-1756}$ has been collected, however, the iterative process of assigning the backbone amide groups has proven challenging and remains incomplete. Thus, this information was not able to be used for the purposes of NMR titrations between $PPL_{1626-1756}$ and $vimentin_{99-249}$. Following completion of these assignments, these titrations would prove extremely useful in pinpointing molecular contacts between the two proteins beyond those identified through previously discussed charge reversal mutants. The 3D spectra can further be used to determine the 3D solution structure of the $PPL_{1626-1756}$. Solving this structure will provide a substantial improvement in accuracy over the current crystal structure which will undoubtedly lead to additional insights about the linker domain.

5.2.3 The impact of true phosphorylation

This thesis presented the results of phosphomimic mutation on the function of the periplakin linker domain. However, a glutamate is no replacement for canonical phosphorylation. Although we can speculate fairly accurately what effect true phosphorylation would have, determination of the actual consequences of phosphorylation remain to be seen. Furthermore, identification of the kinase(s) and phosphatase(s) responsible for this phosphorylation event is required.

5.2.4 Characterization of other plakin linker domains

Given the results presented in this work regarding the periplakin linker domain, a logical next step is to determine whether or not the linker domains of other plakin proteins exhibit the same structure or function. As discussed in section 4.2.3, to date the phosphorylation of the linker domain appears to be unique to periplakin. The function of the other linker domains may also be unique in their own way such as for signalling processes or trafficking.

BIBLIOGRAPHY

Bibliography:

1. Wegener, J. Cell Junctions. in *eLS* (American Cancer Society, 2011). doi:10.1002/9780470015902.a0001275.pub2

2. Evans, W. H. & Martin, P. E. M. Gap junctions: structure and function (Review). *Mol. Membr. Biol.* **19,** 121–136 (2002).

3. Niessen, C. M. Tight Junctions/Adherens Junctions: Basic Structure and Function. *J. Invest. Dermatol.* **127,** 2525–2532 (2007).

4. Zihni, C., Mills, C., Matter, K. & Balda, M. S. Tight junctions: from simple barriers to multifunctional molecular gates. *Nat. Rev. Mol. Cell Biol.* **17,** 564–580 (2016).

5. Walko, G., Castañón, M. J. & Wiche, G. Molecular architecture and function of the hemidesmosome. *Cell Tissue Res.* **360,** 363–378 (2015).

6. Garrod, D. R., Berika, M. Y., Bardsley, W. F., Holmes, D. & Tabernero, L. Hyper-adhesion in desmosomes: its regulation in wound healing and possible relationship to cadherin crystal structure. *J. Cell Sci.* **118,** 5743–5754 (2005).

7. Windoffer, R., Borchert-Stuhlträger, M. & Leube, R. E. Desmosomes: interconnected calcium-dependent structures of remarkable stability with significant integral membrane protein turnover. *J. Cell Sci.* **115,** 1717–1732 (2002).

8. Brooke, M. A., Nitoiu, D. & Kelsell, D. P. Cell–cell connectivity: desmosomes and disease. *J. Pathol.* **226,** 158–171 (2012).

9. Cirillo, N. & Prime, S. S. Desmosomal interactome in keratinocytes: a systems biology approach leading to an understanding of the pathogenesis of skin disease. *Cell. Mol. Life Sci.* **66,** 3517–3533 (2009).

10. Celentano, A., Mignogna, M. D., McCullough, M. & Cirillo, N. Pathophysiology of the Desmo-Adhesome. *J. Cell. Physiol.* **232,** 496–505 (2017).

11. Syed, S.-H. *et al.* Molecular interactions between desmosomal cadherins. *Biochem. J.* **362,** 317–327 (2002).

12. Al-Amoudi, A. & Frangakis, A. S. Structural studies on desmosomes. *Biochem. Soc. Trans.* **36,** 181–187 (2008).

13. Holthöfer, B., Windoffer, R., Troyanovsky, S. & Leube, R. E. Structure and Function of Desmosomes. in *International Review of Cytology* **264,** 65–163 (Academic Press, 2007).

14. Troyanovsky, S. M. *et al.* Identification of the plakoglobin-binding domain in desmoglein and its role in plaque assembly and intermediate filament anchorage. *J. Cell Biol.* **127,** 151–160 (1994).

15. Garrod, D. R., Merritt, A. J. & Nie, Z. Desmosomal cadherins. *Curr. Opin. Cell Biol.* **14,** 537–545 (2002).

16. Stokes, D. L. Desmosomes from a structural perspective. *Curr. Opin. Cell Biol.* **19,** 565–571 (2007).

17. Choi, H.-J., Gross, J. C., Pokutta, S. & Weis, W. I. Interactions of Plakoglobin and β-Catenin with Desmosomal Cadherins BASIS OF SELECTIVE EXCLUSION OF α- AND β-CATENIN FROM DESMOSOMES. *J. Biol. Chem.* **284,** 31776–31788 (2009).

18. Künzli, K., Favre, B., Chofflon, M. & Borradori, L. One gene but different proteins and diseases: the complexity of dystonin and bullous pemphigoid antigen 1. *Exp. Dermatol.* **25,** 10–16 (2016).

19. Yan, Y. *et al.* Crystal structure of the repetitive segments of spectrin. *Science* **262,** 2027–2030 (1993).

20. Kowalczyk, A. P. *et al.* The Amino-terminal Domain of Desmoplakin Binds to Plakoglobin and Clusters Desmosomal Cadherin–Plakoglobin Complexes. *J. Cell Biol.* **139,** 773–784 (1997).

21. Choi, H.-J. & Weis, W. I. Structure of the Armadillo Repeat Domain of Plakophilin 1. *J. Mol. Biol.* **346,** 367–376 (2005).

22. Nollet, F., Kools, P. & van Roy, F. Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members11Edited by M. Yaniv. *J. Mol. Biol.* **299,** 551–572 (2000).

23. Ruhrberg, C., Hajibagheri, M. A. N., Parry, D. A. D. & Watt, F. M. Periplakin, a Novel Component of Cornified Envelopes and Desmosomes That Belongs to the Plakin Family and Forms Complexes with Envoplakin. *J. Cell Biol.* **139,** 1835–1849 (1997).

24. Stahley, S. N., Bartle, E. I., Atkinson, C. E., Kowalczyk, A. P. & Mattheyses, A. L. Molecular organization of the desmosome as revealed by direct stochastic optical reconstruction microscopy. *J Cell Sci* **129,** 2897–2904 (2016).

25. Fontao, L. *et al.* Interaction of the Bullous Pemphigoid Antigen 1 (BP230) and Desmoplakin with Intermediate Filaments Is Mediated by Distinct Sequences within Their COOH Terminus. *Mol. Biol. Cell* **14,** 1978–1992 (2003).

26. Choi, H.-J., Park-Snyder, S., Pascoe, L. T., Green, K. J. & Weis, W. I. Structures of two intermediate filament-binding fragments of desmoplakin reveal a unique repeat motif structure. *Nat. Struct. Mol. Biol.* **9,** 612–620 (2002).

27. Choi, H.-J. & Weis, W. I. Chapter Eleven - Purification and Structural Analysis of Desmoplakin. in *Methods in Enzymology* (eds. Wilson, K. L. & Sonnenberg, A.) **569,** 197–213 (Academic Press, 2016).

28. Fogl, C. *et al.* Mechanism of intermediate filament recognition by plakin repeat domains revealed by envoplakin targeting of vimentin. *Nat. Commun.* **7,** (2016).

29. Kazerounian, S., Uitto, J. & Aho, S. Unique role for the periplakin tail in intermediate filament association: specific binding to keratin 8 and vimentin. *Exp. Dermatol.* **11,** 428–438 (2002).

30. Stappenbeck, T. S., Lamb, J. A., Corcoran, C. M. & Green, K. J. Phosphorylation of the desmoplakin COOH terminus negatively regulates its interaction with keratin intermediate filament networks. *J. Biol. Chem.* **269,** 29351–29354 (1994).

31. Amar, L. S., Shabana, A. H., Oboeuf, M., Martin, N. & Forest, N. Involvement of desmoplakin phosphorylation in the regulation of desmosomes by protein kinase C, in HeLa cells. *Cell Adhes. Commun.* **7,** 125–138 (1999).

32. Amar, L. S., Shabana al-HM, null, Oboeuf, M., Martin, N. & Forest, N. Desmosomes are regulated by protein kinase C in primary rat epithelial cells. *Cell Adhes. Commun.* **5,** 1–12 (1998).

33. Kang, H., Weiss, T. M., Bang, I., Weis, W. I. & Choi, H.-J. Structure of the Intermediate Filament-Binding Region of Desmoplakin. *PLOS ONE* **11,** e0147641 (2016).

34. Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumors and cultured cells. *Cell* **31,** 11–24 (1982).

35. Moll, R., Divo, M. & Langbein, L. The human keratins: biology and pathology. *Histochem. Cell Biol.* **129,** 705–733 (2008).

36. Schweizer, J. *et al.* New consensus nomenclature for mammalian keratins. *J Cell Biol* **174,** 169–174 (2006).

37. Bunick, C. G. & Milstone, L. M. The X-Ray Crystal Structure of the Keratin 1-Keratin 10 Helix 2B Heterodimer Reveals Molecular Surface Properties and Biochemical Insights into Human Skin Disease. *J. Invest. Dermatol.* **137,** 142–150 (2017).

38. Aziz, A. *et al.* The Structure of Vimentin Linker 1 and Rod 1B Domains Characterized by Site-directed Spin-labeling Electron Paramagnetic Resonance (SDSL-EPR) and X-ray Crystallography. *J. Biol. Chem.* **287,** 28349–28361 (2012).

39. Chernyatina, A. A. & Strelkov, S. V. Stabilization of vimentin coil2 fragment via an engineered disulfide. *J. Struct. Biol.* **177,** 46–53 (2012).

40. Chernyatina, A. A., Nicolet, S., Aebi, U., Herrmann, H. & Strelkov, S. V. Atomic structure of the vimentin central α-helical domain and its implications for intermediate filament assembly. *Proc. Natl. Acad. Sci.* **109,** 13620–13625 (2012).

41. Meier, M. *et al.* Vimentin Coil 1A—A Molecular Switch Involved in the Initiation of Filament Elongation. *J. Mol. Biol.* **390,** 245–261 (2009).

42. Strelkov, S. V. *et al.* Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. *EMBO J.* **21,** 1255–1266 (2002).

43. Eriksson, J. E. *et al.* Specific in vivo phosphorylation sites determine the assembly dynamics of vimentin intermediate filaments. *J. Cell Sci.* **117,** 919–932 (2004).

44. Guo, M. *et al.* The Role of Vimentin Intermediate Filaments in Cortical and Cytoplasmic Mechanics. *Biophys. J.* **105,** 1562–1568 (2013).

45. Inagaki, M., Nishi, Y., Nishizawa, K., Matsuyama, M. & Sato, C. Site-specific phosphorylation induces disassembly of vimentin filaments in vitro. *Nature* **328,** 649–652 (1987).

46. Kartenbeck, J., Schwechheimer, K., Moll, R. & Franke, W. W. Attachment of vimentin filaments to desmosomal plaques in human meningiomal cells and arachnoidal tissue. *J. Cell Biol.* **98,** 1072–1081 (1984).

47. Vorobiev, S. *et al.* Crystal Structure of the Plectin 1 and 2 Repeats of Human Periplakin. *BE Publ.* (2014). doi:10.2210/pdb4q28/pdb

48. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43,** D512–D520 (2015).

49. Rigbolt, K. T. G. *et al.* System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.* **4,** rs3 (2011).

50. Carrier, M., Joint, M., Lutzing, R., Page, A. & Rochette-Egly, C. Phosphoproteome and Transcriptome of RA-Responsive and RA-Resistant Breast Cancer Cell Lines. *PLoS ONE* **11,** (2016).

51. Klammer, M. *et al.* Phosphosignature Predicts Dasatinib Response in Non-small Cell Lung Cancer. *Mol. Cell. Proteomics MCP* **11,** 651–668 (2012).

52. Mertins, P. *et al.* Ischemia in Tumors Induces Early and Sustained Phosphorylation Changes in Stress Kinase Pathways but Does Not Affect Global Protein Levels. *Mol. Cell. Proteomics MCP* **13,** 1690–1704 (2014).

53. Dinkel, H. *et al.* ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.* **44,** D294–D300 (2016).

54. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **125,** 1731–1737 (2003).

55. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative

    Modeling of Biomolecular Complexes. *J. Mol. Biol.* **428,** 720–725 (2016).

56. Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable

    DNA polymerase. *Science* **239,** 487–491 (1988).

57. Yarmoluk, S. M., Kovalska, V. B. & Volkova, K. D. Optimized Dyes for Protein and

    Nucleic Acid Detection. in *Advanced Fluorescence Reporters in Chemistry and Biology III*

    161–199 (Springer, Berlin, Heidelberg, 2011). doi:10.1007/978-3-642-18035-4_5

58. Franke, D. *et al.* ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering

    from macromolecular solutions. *J. Appl. Crystallogr.* **50,** 1212–1225 (2017).

59. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a

    Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36,**

    1277–1282 (2003).

60. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-

    angle scattering. *Nature* **496,** 477–481 (2013).

61. Tuukkanen, A. T. & Svergun, D. I. Weak protein–ligand interactions studied by small-angle

    X-ray scattering. *FEBS J.* **281,** 1974–1987 (2014).

62. Koch, M. H., Vachette, P. & Svergun, D. I. Small-angle scattering: a view on the properties,

    structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.*

    36, 147–227 (2003).

63. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and

    applications in biomedical sciences. *Biophys. Rev.* **8,** 409–427 (2016).

64. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40,** 191–285 (2007).

65. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42,** 342–346 (2009).

66. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879–2886 (1999).

67. Volkov, V. V., Svergun, D. I. & IUCr. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* (2003). doi:10.1107/S0021889803000268

68. Blanchet, C. E. & Svergun, D. I. Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. *Annu. Rev. Phys. Chem.* **64,** 37–54 (2013).

69. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

70. Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1,** 100 (2010).

71. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci. Publ. Protein Soc.* **4,** 2411–2423 (1995).

72. Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357,** 289–298 (2006).

73. Kozak, S. *et al.* Optimization of protein samples for NMR using thermal shift assays. *J. Biomol. Nmr* **64,** 281–289 (2016).

74. T, A. K. CFSSP: Chou and Fasman Secondary Structure Prediction server. *Zenodo* (2013). doi:10.5281/zenodo.50733

75. Chou, P. Y. & Fasman, G. D. Prediction of protein conformation. *Biochemistry (Mosc.)* **13,** 222–245 (1974).

76. Chou, P. Y. & Fasman, G. D. Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry (Mosc.)* **13,** 211–222 (1974).

77. Sammito, M. *et al.* Exploiting tertiary structure through local folds for crystallographic phasing. *Nat. Methods* **10,** 1099–1101 (2013).

78. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539–539 (2011).